

WIESŁAW GALUS

## ARCHITEKTURA ŚWIADOMOŚCI CZĘŚĆ I: LOGIKA I MORFOLOGIA SIECI NEURONOWEJ

### 1. WSTĘP

Ostatnie dziesięciolecie przyniosły niesłychany rozwój metod obrazowania procesów neuronowych w mózгах żywych zwierząt i ludzi. Pozwoliło to odkryć wiele właściwości ludzkiej i zwierzęcej pamięci oraz jej plastyczności, umożliwiającej uczenie. Wszystkie te sukcesy nie doprowadziły jednak do wyjaśnienia, jak w mózgu rodzi się świadomość, świadome odczuwanie, świadomość przestrzeni, w której znajduje się myślący umysł, rozumienie własnego położenia w otaczającym środowisku, do objaśnienia istoty odczuwania jakości, a tym bardziej – jak powstaje bogata psychika, zdolna do refleksji nad własnym stanem umysłu, czyli świadomość refleksyjna, według terminologii Velmansa [VELMANS 2009].

W okresie ostatnich 10 lat opublikowano przełomowe rezultaty prac badawczych i analiz funkcjonowania sieci neuronowych, zarówno naturalnych jak i sztucznych, które pozwalają na zbudowanie modelu mózgu zdolnego do tworzenia umysłu. Rezultaty tych prac, z zakresu badań nad inteligencją obliczeniową, z zakresu psychologii, neuromorfologii i neurofizjologii, fizyki zjawisk na membranach synaptycznych oraz filozofii umysłu, zebrano w niniejszej pracy, łącząc je w ciąg hipotez i wyników badań ukazujący sposób przetwarzania percepowanych sygnałów od wstępnych pól percepcyjnych, aż do poziomu wytwarzania świadomości i samoświadomości. Trzeba tu zaznaczyć, że wiele tych najnowszych rezultatów nie uzyskało jeszcze powszechnego uznania i nie stały się standardem w objaśnianiu procesów psychicz-

nych w mózgach. Wszystkie one jednak są dobrze udokumentowane eksperymentalnie i potwierdzone w wielu niezależnych, czołowych ośrodkach naukowych zajmujących się tymi specjalistycznymi dziedzinami wiedzy. Oparcie się na tych rezultatach uzasadnione jest także olbrzymią liczbą cytoowań wyselekcjonowanych tutaj prac, która wzrasta z każdym miesiącem. Praca niniejsza jest próbą zbudowania kompleksowego modelu umysłu w oparciu o te nowe odkrycia, jak również w oparciu o własne koncepcje i hipotezy dotyczące zasad jego funkcjonowania. Ze względu na obszerność multidyscyplinarnej tematyki koniecznej dla objaśnienia modelu umysłu świadomego pracę podzielono na trzy części, które będą za prezentowane w kolejnych numerach. Są one wzajemnie powiązane i wymagają zapoznania się z całością dla prześledzenia wzajemnych zależności między poszczególnymi funkcjami mózgu i umysłu.

Zbudowanie satysfakcjonującego modelu umysłu jest równoważne zrozumieniu fenomenu, jakim jest świadomość. Model „satysfakcjonujący”, oznacza model możliwy do zrozumienia, unikając paradoksu Boniniego. Potwierdzeniem prawidłowego rozumienia powinien być walor predykcyjny modelu, co oznacza zdolność wyjaśnienia wielu fenomenów rozpoznawanych w umyśle przez psychologię i neurologię, a także przez psychiatrię, w odniesieniu do mózgów i umysłów patologicznych. W szczególności model wytłumaczyć musi biofizyczne podłoże pamięci trwałej, jak i krótkotrwałej, roboczej, deklaratywnej, semantycznej i epizodycznej oraz jej niezwykle właściwości występujące u człowieka. Należy do nich zdolność do trwałego zapamiętania epizodów w jednorazowym akcie zapamiętywania, przy zachowaniu plastyczności pamięci, gwarantującej zdolność do dalszego uczenia się; zdolność do zapamiętywania krótkotrwałego informacji aktualnie przetwarzanych i sposób integracji pamięci krótkotrwałej z pamięcią długoterminową; sposób adresowania zawartości pamięci epizodycznej tak, aby możliwe było szybkie przypominanie zapamiętanych sekwencji w dowolnej konfiguracji czasoprzestrzennej. Model wskazywać musi na algorytmy przetwarzania postrzeżeń dostarczanych przez wejściowe pola percepcyjne na wrażenia i odczucia oraz na rozpoznawalne i uświadomione percepty podlegające dalszej kategoryzacji, asocjacji i generalizacji. Model musi także wyjaśniać, w jaki sposób morfologia i reguły funkcjonowania modelu kreują emocje i motywacje obserwowane w realnym umyśle.

Biorąc pod uwagę tak sformułowane zadanie, w niniejszej pracy przedstawiono w trzech częściach zamieszczonych w kolejnych numerach „Roczników Filozoficznych” podstawy biofizyczne i epistemologiczne, na których

wspiera się proponowany model. W części pierwszej w rozdziale 1 zasygnalizowano, jak wykorzystywana w dalszej części pracy idea torowała sobie drogę w koncepcjach dwudziestowiecznych badaczy, kontynuujących tradycje Arystotelesowskie. Rozdział 2 przypomina właściwości sieci neuronowych stosowanych do rozpoznawania obrazów i postrzeżeń zmysłowych oraz tworzenia perceptów. Zdolności te wykorzystywane zostały w dalszych torach uświadamiania perceptów, tworzonych na niższych poziomach przetwarzania. Przedstawiono także trudności, na jakie natrafiły dotychczasowe próby modelowania tych procesów w formie tak zwanej klasycznej sztucznej inteligencji (AI). W rozdziale 3 pokazano, jak wsparcie się na epistemologicznych podstawach rozumienia pozwala matematycznie modelować ten proces, a w rozdziale 4 podkreślono skuteczność algorytmów logiki dynamicznej, pozwalającej nie tylko na rozpoznanie obiektów i zrozumienie otoczenia, ale także kreującej motywację do działań poznawczych. Rozdział 5 wskazuje na ograniczenia czysto matematycznego modelowania procesu poznawczego. W rozdziale 6 zademonstrowano, w jaki sposób usunąć można te ograniczenia i jak algorytmy rozpoznawania realizowane być mogą przez sprzężenia synaptyczne realnych struktur neuronowych mózgów naturalnych. Rozdział 7 wskazuje na konieczność znalezienia biofizycznego podłoża procesów neuronowych a szczególnie synaptycznych. W części drugiej w rozdziale 8 wykazano, że bez wniknięcia w molekularną strukturę przekazywanych sygnałów nie sposób jest wykorzystać struktury morfologicznej mózgu do wyjaśnienia obserwowanych w naturze procesów mózgowych. Dopiero zastosowanie praw Neuro-Elektro-Dynamiki pozwala wyjaśnić niezwykle cechy ludzkiej pamięci. Mechanizmy pamięci, sposoby kolekcjonowania wiedzy zapełniającej poszczególne rodzaje pamięci i znaczenie mechanizmów koncentracji uwagi, które selekcjonują zawartość pamięci roboczej, pozwalają zrozumieć, jak następuje proces uświadamiania sobie qualii. Procesy te opisano bardziej szczegółowo w rozdziale 9. Przedstawiono tu także koncepcję funkcjonowania pamięci epizodycznej zdolnej do przetwarzania relacji czasowych w konfiguracje przestrzenne typowe dla pól synaptycznych w korze mózgowej. W części trzeciej rozdział 10 pomieścił koncepcje motywacji do działań inteligentnych i świadomych. W nim też przedstawiono znaczenie bólu i mechanizmów nagrody, odczuwanych jako qualia przez mózgi naturalne i z tego tytułu traktowane jako dodatkowy rodzaj zmysłów. W rozdziale 11. omówiono fundamentalne znaczenie ciekawości i potrzeby rozumienia jako motywacji procesów poznawczych. Rozdział 12 łączy omawiane we wcześniejszych rozdziałach elementy w kom-

pleksowy model umysłu świadomego. Omówiono tu zróżnicowane stopnie świadomości. Objąsniiono istotę procesu „rozumienia”. Rozdział 13 wykorystuje cechy kompleksowego modelu umysłu do głębszego wniknięcia w istotę stanu świadomości i analizy, jak umysł określać może autonomicznie cel i sens życia. W rozdziale 14 rozważono, czy umysł według proponowanego modelu odznaczał się będzie wolną wolą i samoświadomością. Rozdział 15 przedstawia relacje stanu świadomości generowanej przez model do potocznie rozumianej duszy ludzkiej oraz sposób usunięcia dualizmu między przyczynowym charakterem redukcjonistycznego modelu mózgu a teleologicznym charakterem umysłu tworzonego przez ten mózg. Rozdział 16 stanowi podsumowanie i przedstawia wnioski wynikające z przedstawionego modelu i objaśnienia, czym jest świadomość.

Bez omówienia wszystkich tych procesów i wyjaśnienia wszystkich wymienionych zjawisk umysłowych model pozbawiony będzie waloru przewidywania cech odnajdywanych w umyśle ludzkim i zwierzęcym. Oznaczać to będzie, że nie osiągnęliśmy pełnego zrozumienia procesów w nim zachodzących i że model nie działa. Z tego powodu podjęto próbę zebrania głównych założeń modelu w jednym artykule, umożliwiającym całościowe objęcie wzajemnych powiązań procesów mózgowych i zależności objaśniających, w jaki sposób tak złożona całość może jednocześnie wykazywać tak specyficzne dla umysłu zachowania.

## 1. ŹRÓDŁA WIEDZY W UMYŚLE

Procesy poznawcze, nazywane ogólnie myśleniem, są funkcją sieci neuronowej, która jest dominującą strukturą ludzkiego mózgu odpowiedzialną za rozumienie, świadomość i wszelkie inne wyższe funkcje psychiczne. Na procesy te składa się postrzeganie obiektów, ich rozpoznanie, kategoryzacja i generalizacja. Obiektem może być wszelki obiekt materialny, jak i proces postrzegany przez nasze zmysły. Obiekty te, oddziałując na nasze zmysły, tworzą scenę reprezentującą otaczające nas środowisko i wszelką rzeczywistość, aż po złożoność wszechświata. Rozpoznane obiekty sceny tworzą mentalną reprezentację otaczającej nas rzeczywistości, którą nazywać będziemy modelem rzeczywistości. Modele dotyczyć też mogą fragmentów rzeczywistości znajdujących się w zasięgu naszych zmysłów (np. w polu widzenia). Poszczególne zmysły obsługiwane są przez specyficzne pola mózgu tworzące tzw. modalności.

Właściwością sieci neuronowej, odbierającej pobudzenia neuronowe od receptorów zmysłowych, jest zdolność tworzenia tych modeli w toku indywidualnych doznań i przeżyć, których doświadczaliśmy w przeszłości. Odpowiada to procesowi uczenia. Umysł rozumie świat poprzez te modele, odnosząc je do postrzeganych obiektów. Logiczną konsekwencją tego faktu jest konieczność porównywania mentalnych reprezentacji modeli rezydujących w naszej pamięci z reprezentacjami mentalnymi postrzeganych obiektów.

Nowoczesne pojmowanie procesów poznawczych prowadzących do świadomości miało wielu prekursorów, poczynając od Arystotelesa i Platona. Koncepcje Arystotelesa dopuszczały uczenie się (poprzez dynamiczne oddziaływanie formy i materii) oraz wskazywały na motywację do poznania (w eudajmonii, poprzez czysty akt poznania w odróżnieniu od hedonii, bezpośredniego odczuwania przyjemności bytu).

Istotną koncepcję, z punktu widzenia modelu prezentowanego w niniejszej pracy, przedstawił Marvin Minsky. Postulował on struktury pamięci nazywane „K-line”, łączące aktualny stan umysłu ze stanami uprzednio utrwalonymi w naszej pamięci [MINSKY 1980]. W ten sposób pamięć indukuje stan, poprzez który widzimy bieżącą rzeczywistość, jako przykład zapamiętanego zdarzenia.

Prace nad sztuczną inteligencją zdominowały analizy procesów poznawczych, zachodzących w ludzkim umyśle. Pojawiały się kolejne koncepcje, próbujące odkryć algorytmy, które muszą być wykonywane w systemach naturalnych lub sztucznych, aby system mógł wykazywać inteligencję. Stephen Grossberg wprowadził pojęcie rezonansu między sygnałami górą-dół/dół-góra, co odpowiada konfrontacji ogólnych modeli/wzorców pamięciowych, ze szczegółowymi perceptami odbieranymi przez zmysły. Każde świadome postrzeżenie wymaga takiego rezonansu i gdy go brakuje, postrzeżenie nie trafia do świadomości [GROSSBERG 1982, 2000a, 2000b]. Model relacyjny Taylora odwoływał się w podobny sposób do relacji między sygnałami sensorycznymi i zapamiętanymi epizodami. Przy tym pamięć robocza miała być pierwszą, gdzie porównywane są te sygnały i gdzie emergentnie wyłania się poczucie świadomości [TAYLOR 1992, 1998a, 1998b].

W odróżnieniu od tego nurtu redukcjoniści brali pod uwagę jedynie aspekt funkcjonalny świadomości. Idea wywodząca się z behawioryzmu przybrała skrajną postać prób modelowania poszczególnych funkcji mózgu z przekonaniem, że jeśli takie modele będą dokonywały tych samych operacji co mózg ludzki, wówczas fenomen świadomości będzie musiał im towarzyszyć. Przy tym niekoniecznie musi się on wiązać z właściwościami

wewnętrzny i stanami mentalnymi mózgu. Odzwierciedla tę ideę hipoteza „globalnej przestrzeni roboczej” (*Global Workspace*), zaproponowana przez Baarsa i innych [BAARS 1997, 1998; DEHAENE & NACCACHE 2001], oraz idea inteligencji ucieleśnionej (*Embodied Intelligence*), sformułowana przez Hurley i innych [HURLEY 1998; NOË 2005, 2009].

Dalsze badania sieci neuronowych umożliwiły głębokie wniknięcie w charakter procesów mózgowych. Zdolność przetwarzania informacji dostarczanych przez nasze zmysły, tworzenia ich reprezentacji mentalnych i ich porównywania z reprezentacjami mentalnymi przechowywanymi w naszej pamięci, zawdzięczamy niezwykłym właściwościom struktury, którą tworzą zwierzęce neurony, a które nazywamy siecią neuronową. Jest to równoważne przedstawionym wyżej ideom porównywania modeli i perceptów. Stwierdzono, że neurony mają wrodzoną strukturę połączeń charakterystyczną gatunkowo, jednakże bardzo zbliżoną dla wszystkich zwierząt, posiadających centralny układ nerwowy [MARKRAM 2011]. Szczególnie dobrze poznano procesy rejestracji bodźców zmysłowych i ich przetwarzania w pierwotnych polach kory mózgowej.

Procesy postrzegania i rozpoznawania obiektów, a także procedury uczenia zostały w XX stuleciu dość dobrze poznane i opisane [BISHOP 1996]. Już w latach 80. XX wieku opisano dość szczegółowo i kompleksowo funkcjonowanie układu wzrokowego [INGLE 1982; Milner i inni 2008] i sposób postrzegania i rozpoznawania obrazów [ULLMAN 1996]. Opis oglądu i rozpoznawania sceny przez mózg przedstawił Francis Crick w swojej popularnej książce *Zdumiewająca hipoteza* [Crick 1997], a opis tworzenia reprezentacji neuronowych i dynamikę kształtowania się sieci neuronowej, przedstawił Gerald M. Edelman w książce *Przenikliwe powietrze, jasny ogień* [EDELMAN 1999]. Klasyczny model systemu wzrokowego opisuje proces tworzenia i rozpoznawania obrazu poprzez kaskadę pobudzeń skierowaną z dołu do góry (*bottom-up*) od pobudzenia wejściowych neuronów i pól percepcyjnych (V1, V2), do wyższych pól wzrokowych, w których dokonywana jest analiza i rozpoznawanie obrazów (V4 i wyższe). Wielu badaczy zwraca jednak uwagę na znaczenie procesu porównania neuronowych korelatów rejestrowanych przez zmysły obrazów, tworzonych w niższych warstwach sieci, z konfiguracjami pobudzeń neuronowych, korelatami modeli, płynących z „górných” pól neuronowych (leżących wyżej w hierarchii przetwarzania), czyli z góry w dół (*top-down*). Te korelaty modeli wytwarzane są w wyższych polach mózgowych w procesie wcześniejszego uczenia.

## 2. ARCHITEKTURA SIECI NEURONOWEJ

Dalsze badania pozwoliły rozpoznać złożoną warstwową strukturę kory mózgowej, co umożliwiło jej matematyczne i elektroniczne modelowanie. Starano się skopiować niezwykle właściwości sieci neuronowych, tworząc struktury, które miały działać tak jak naturalny mózg i zwierzęce neurony. W ostatnich latach dokonano w tej dziedzinie znaczących postępów. Zbudowano elektroniczne modele neuronowych sieci kognitywnych, naśladujące procesy przebiegające w umysłach istot żywych. Projekty te doprowadzić miały do zbudowania sztucznej inteligencji (*Artificial Intelligence* – AI). Architektura tych sieci w dużej mierze przypomina, a lepiej powiedzieć – intencjonalnie naśladuje, budowę mózgu. Większość tych sieci składa się ze stosunkowo prostych elementów – komórek, połączonych wzajemnie w sposób umożliwiający przesyłanie sygnałów do dużej liczby podobnych komórek odpowiadających neuronom. Własności przetwarzania sygnałów determinowane są strukturą sieci połączeń neuronów. Połączenia te mogą mieć zmienną charakterystykę transmisji poprzez elementy łączące, zwane – przez analogię z mózgiem – synapsami. Sieci posiadają budowę warstwową – od niższych poziomów tworzących pola percepcyjne, do poziomów wyższych. Posiadają one możliwość przekazywania bodźców reprezentujących stany mentalne w górę hierarchii od pól sprzężonych z receptorami do wyższych poziomów przetwarzania (*bottom-up*). Stany mentalne to konfiguracje stanów pobudzeń neuronów docierające do komórek wyższej warstwy. Są one wyznaczone przez matryce określające stany blokowania lub transmisji synaps, które zapamiętywane są w pamięci trwałej (*Long Term Memory* – LTM). Przekazywane w ten sposób do wyższych warstw informacje o odbieranych wrażeniach tworzą drzewiastą strukturę odpowiadającą hierarchicznej organizacji wzorców mentalnych.

Te ogólne zasady modyfikowane są w zależności od architektury sieci, rodzaju stosowanych sposobów zapamiętywania danych i parametrów struktury oraz sposobów uczenia. Sieci kognitywne posiadać mogą architektury różniące się sposobem przekształcania bodźców i relacji w język symboliczny, wykorzystywany do rozwiązywania problemów. W zależności od rodzaju architektury preferowane są różne techniki uczenia, a także specyficzne sposoby transformacji zapamiętanej wiedzy oraz jej wykorzystywania [DUCH 2010]. W niniejszej pracy staramy się dowieść, że funkcje te wypełniać może architektura koneksjonistyczna, wielowarstwowa, gdzie warstwy niższe tworzą reprezentacje niższych rzędów, a w warstwach wyższych

przetwarzana może być wiedza symboliczna w wyniku procesów jednorodnych, tego samego typu, w całym obszarze kory mózgowej.

Budowane modele matematyczne i elektroniczne nie mogły przekroczyć progu, wskazującego, że procesy myślenia właściwie rozumiemy, nie mówiąc już o szansach na to, żeby zdolność rozumienia mogły zademonstrować budowane modele. Nikt nie podejrzewa, że maszyny te rozumieją, co tak sprawnie i szybko wykonują. Co gorsza, nie rozumieliśmy dotychczas, jak to się dzieje, że rozumiemy otaczający nas świat. Co to znaczy rozumieć w kategoriach procesów neurologicznych i fizykochemicznych zachodzących w mózgu? Czy rozumienie ma związek z fenomenem świadomości? Dalej bez odpowiedzi pozostaje pytanie: co mózg właściwie robi, że tworzy świadomy umysł? [ABBOTT 2008].

### 3. MODEL OBLICZENIOWY PROCESU ROZPOZNAWANIA

Każdy fragment kory naszego mózgu zawiera sieć neuronową. Jest on nieustannie nastawiony na przyjmowanie pobudzeń i ich dalsze przekazywanie po porównaniu z innymi przychodzącymi równolegle sygnałami. Ponadto fragmenty kory o pewnych charakterystycznych rozmiarach (nazywanych niekiedy, w dalszej części pracy, polami mózgowymi) mają właściwość zapamiętywania stanów pobudzeń wywoływanych przez pobudzenia docierające z pól sąsiednich. Możemy się spodziewać, że to właśnie w tych polach odnajdziemy zapamiętane wzorce pobudzeń, konfiguracje stanów fizycznych neuronów i zbiorów synaps, reprezentujące modele obiektów z którymi zetknął się umysł za pośrednictwem zmysłów w toku życiowych doświadczeń odpowiadających uczeniu się.

Jak zauważają psychologowie, istotna jest własność percepcji wzrokowej, polegająca na rozmytej naturze reprezentacji wewnętrznej tych zagnieżdżonych w pamięci wzorców. Uzmysłowienie sobie rozmytego charakteru modeli obiektów ma zasadnicze znaczenie dla zrozumienia procesów „rozumienia” perceptów przez sieć neuronową, a zatem i przez umysł, niezależnie, czy jest to umysł naturalny, ludzki lub zwierzęcy, czy też sztuczny. Sygnały te są niejednoznacznie określone, zawierają niewiele informacji w porównaniu z bogactwem sygnałów docierających od naszych zmysłów, ponieważ reprezentują jedynie ogólne modele pojęć, które zawiera nasza pamięć. Zobrazujmy to przykładem modelu-pojęcia, jakim w pamięci większości ludzi jest pojęcie piłki. Przecież wydaje się, że wiemy, co to jest



piłka. Spróbujmy przywołać jej obraz z naszej pamięci. Zamknijmy uprzednio oczy. Jakże blade, nieostre, rozmyte jest to wyobrażenie piłki. Jakże odbiega od wyrazistego wyglądu piłki w słoneczny dzień, na plaży lub boisku. Model pojęcia piłki w naszej pamięci nie zawiera tych wszystkich licznych szczegółów, specyficznego rozmiaru, tekstury materiału, kolorów, śladów szycia, plamek, rys, zabrudzeń i mnóstwa innych elementów, które posiada konkretna piłka, kiedy jej się przyglądamy. To taki właśnie model-pojęcie miał na myśli Platon, kiedy pisał o ideach, to prawdopodobnie jest Arystotelesowska forma, którą trzeba połączyć z obrazem materialnym, aby rozpoznać konkretną piłkę, gdy jej doświadczamy naszymi zmysłami. Odpowiada to także stanom rezonansu sygnałów *top-down* i *bottom-up* Grossberga. Ogólność modeli-pojęć i możliwość ich uszczegółowienia poprzez porównanie z sygnałami z dolnych warstw neuronowych (*bottom-up*, od pól sensorycznych, najniższych w hierarchii przetwarzania informacji w mózgu) została wykorzystana przez Leonida Perlovsky'ego. Opracowana przez niego hipoteza neuronowych pól modelujących NPM (*Neural Modeling Fields*), przedstawiona i rozwijana w kolejnych pracach [PERLOVSKY 1997, 2001, 2002, 2006a, 2006b, 2007], najpełniej objaśnia możliwość funkcjonowania mózgu tak, aby mógł on nie tylko rozpoznawać obiekty, ale także rozumieć ich znacznie w obrazie otoczenia, w scenie i ich rolę w modelu świata, który go otacza. Uwzględnia także emocje towarzyszące poznaniu, odczucia estetyczne i ostatecznie kształtowanie świadomości.

Perlovsky zwrócił uwagę na fakt, że postrzeganie to porównanie rozmytego obrazu modelu z ostrym obrazem obiektów postrzeganych przez nasze zmysły. W kolejnych warstwach neuronowych mózgu mamy do czynienia z procesem uszczegóławiania ogólnych modeli, czyli z procesem „od rozmycia do ostrości” (rozmycie  $\rightarrow$  ostrość; RO). Na niższym poziomie proces RO ma zasadnicze znaczenie dla szybkiej percepcji. Perlovski przedstawił skuteczny matematyczny model takiego procesu i pokazał, jak stany logiczne wyłaniają się z ze stanów rozmytych w ciągłym procesie porównywania strumieni informacji z góry-na dół (*top-down*) i z dołu-do góry (*bottom-up*).

W tym celu wprowadził pojęcie neuronowych pól modelujących (NPM) do opisu funkcjonalnego percepcji biologicznej. Nazwa „pola modelujące” wiąże się z ich zdolnością do budowania modeli. Pola te opisują nie pojedyncze neurony, ale agregaty odpowiadające mikrokolumnom i kolumnom kory mózgowej [JONES (2000)]. Być może na wyższym poziomie integracji odpowiadają one mikropolom lub fragmentom pól mózgowych. Tworzą one system semihierarchiczny, wielopoziomowy. W poszczególnych warst-

wach neuronowych występuje wiele sprzężeń zwrotnych umożliwiających przesyłanie sygnałów w dół, ku pierwotnym polom percepcyjnym. Dlatego strukturę tę nazywamy semihierarchiczną. W trakcie percepcji następuje asocjacja podzbioru sygnałów odnoszących się do obiektów i pochodzących od zmysłów (*bottom-up*) z reprezentacją ogólnego modelu-pojęcia odnoszącego się do tych obiektów przywoływaną z pamięci (*top-down*). Dopasowanie tych reprezentacji skutkuje rozpoznaniem obiektu. Mamy więc proces asocjacja-rozpoznanie-rozumienie, który można opisać matematycznie, co przedstawione zostało wyczerpująco w cytowanych pracach.

W celu opisanego agregatu umożliwiającego wskazaną wyżej asocjację Perlovsky wyodrębnił grupę  $N$  neuronów  $n$  ponumerowanych  $n=1, \dots, N$ . Sygnały *bottom-up* oznaczamy  $\mathbf{X}(n)$ .  $\mathbf{X}(n)$  jest konfiguracją pobudzeń synaps *bottom-up*, przychodzącą z neuronów z niższego poziomu, przy czym poziom najniższy odpowiada sygnałom sensorycznym, niezależnie od modalności. Każdy neuron ma pewną liczbę synaps. Pobudzający sygnał *top-down* przysyłany jest od neuronów z wyższego poziomu jako reprezentacja mentalna modelu lub pojęcia  $\mathbf{M}_m(\mathbf{S}_m, n)$  gdzie  $m=1, \dots, M$ , to oznaczenie modelu lub reprezentacji. Każdy model  $\mathbf{m}$  opisywany jest przez zespół parametrów  $\mathbf{S}_m$ . W neuronowej strukturze mózgu są one kodowane przez właściwości transmisyjne kanałów połączeń, a w klasycznym modelu przez siłę połączeń synaptycznych, którą matematycznie możemy oznaczyć jako zestaw liczb:  $\mathbf{S}_m = \{S_{am}, a = 1 \dots A\}$ .

Reprezentacje są porównywane z sygnałami, co oznacza, że neuron  $n$  jest aktywowany przez oba sygnały – przychodzący z wejść dolnego poziomu (*bottom-up*)  $\mathbf{X}(n)$  i sygnał pobudzający  $\mathbf{M}_m$  reprezentujący model-pojęcie, tj. ich reprezentację mentalną z wyższych poziomów pamięci. Różne modele-pojęcia  $\mathbf{M}$  konkurują o porównanie z sygnałem *bottom-up*, dostosowując swe parametry w celu uzyskania lepszego dopasowania do wejściowego sygnału. NPM zakłada, że te same reguły dopasowania stosowane są do opisu procesu percepcji i rozpoznania obiektów, jak również do uświadomienia sobie złożonych pojęć abstrakcyjnych na wyższych poziomach. Ta sama dynamiczna interakcja sygnałów odnosi się do dopasowania modelu do obiektu w przypadku percepcji oraz reprezentacji do sytuacji i relacji między obiektami w przypadku procesu „rozumienia”. Według NPM sygnały *bottom-up* to dane nieustrukturyzowane  $\{\mathbf{X}(n)\}$ , a sygnały wyjściowe są rozpoznane i formują pojęcie  $\{m\}$ . Niepewność i rozmycie modelu odpowiada rozmyciu zasocjowanych zmiennych  $f(m|n)$ . Sygnały wyjściowe stają się sygnałami *bottom-up* dla kolejnej, wyższej warstwy.

Percepcja i rozumienie wymaga uczenia. Modele, którymi dysponuje pamięć tworzą się w procedurze uczenia. Potrzeba wiedzy, przejawia się jako dążenie do dopasowania sygnałów *top-down* i *bottom-up* na każdym poziomie struktury hierarchicznej. Matematycznie jest to realizowane jako poszukiwanie miary podobieństwa między zestawem modeli i sygnałów wejściowych:  $L(\{X\}\{M\})$ .

Tą miarą podobieństwa jest matematyczna funkcja podobieństwa. Funkcja podobieństwa jest iloczynem wszystkich pobudzeń  $\{n\}$  synaps neuronów *bottom-up* pobudzonych przez sygnał zewnętrzny  $\{X(n)\}$  oraz wszystkich stanów synaps neuronów reprezentujących model  $\{m\}$ .

Znane algorytmy dopasowywania modeli do ich parametrów i danych rzeczywistych prowadzą do kombinatorycznej mnogości obliczeń dla algorytmów zupełnych klasy N i NP, o której pisaliśmy wyżej. Także logika rozmyta jest tu nieprzydatna, ponieważ mechanizmy samouczenia, polegające na takiej logice w ich treningu lub procedurze uczenia i samouczenia, zmuszają do traktowania każdego przykładu lub cyklu treningowego jako odrębnego zdania logicznego. Prowadzi to do tak zwanego przekleństwa mnogości kombinatorycznej (*CombinatorialComplexityConundrum* – CCC) [PERLOVSKY 1998], uniemożliwiającej efektywne rozpoznanie złożonych obiektów, ponieważ dopasowanie musi polegać na operacjach logicznych w celu ustalenia stopnia rozmycia. W złożonych obiektach nie sposób jest przewidzieć mnogości wymaganych reguł, co także prowadzi do wewnętrznej niespójności. Jak widać, w ogólnym przypadku rozpoznawania nieznanymi, złożonych obiektów zasady Arystotelesowskiej logiki formalnej nie są przydatne.

Perlovski ominął te trudności, zauważając, że w procesie poznania rozmyty, ogólny model będzie się zmieniał w czasie i dostosowywał do szczegółów sygnałów *bottom-up*. To założenie pozwoliło mu na rozwiązanie problemu poprzez użycie logiki dynamicznej (LD) i procedury RO. Zależność parametrów modelu od czasu, pozwala na zróżniczkowanie funkcji podobieństwa po czasie. Uzyskuje się w ten sposób równanie różniczkowe pierwszego rzędu, którego rozwiązanie uzyskać można za pomocą standardowego solvera. Jest to istota logiki dynamicznej (LD), nazywanej tak przez Perlovskiego ze względu na dynamikę zmian modelu w czasie. W procedurze tej następuje także dopasowanie rozmycia i miary podobieństwa do stopnia „niepewności” modelu. Początkowo, w pierwszych iteracjach numerycznego wyznaczania rozwiązań wynikających z różniczkowania funkcji podobieństwa, gdy wartości parametrów są przypadkowe, wartość funkcji podobień-

stwa jest niska, mamy do czynienia z rozmyciem rozpoznawanego obiektu, wartości parametrów nie są znane i stopień „niepewności” jest wysoki, a rozpoznanie obiektu niemożliwe. Dotyczy to także rozmycia miary podobieństwa. Wówczas każdy punkt danych rejestrowanych wykazuje pewne podobieństwo do dowolnego modelu. Jednakże w procesie samouczenia, po kilku iteracjach, wartości parametrów poprawiają się. Model staje się coraz dokładniejszy i ostrzejszy (lepiej opisuje obiekt) – wartość miary podobieństwa wzrasta, zasocjowane zmienne  $f(m|n)$  dążą do zer lub jedynek, co oznacza wyostrenie, ujednoznacznienie asocjacji. Tylko jeden model wygrywa konkurencję o podzbiór generowany przez sygnał wejściowy  $x(n)$  (zasada WTA – zwycięzca bierze wszystko). Rozmyty *a priori* model-pojęcie staje się pojęciem ostrym (RO). Następuje rozpoznanie obiektu, a gdy proces dotyczy wielu warstw na wyższych poziomach abstrakcji, następuje ulokowanie obiektu na tle szerszej sceny, co równoznaczne jest ze zrozumieniem jego roli w modelu świata. Zbieżność tego procesu udowodnił Perlovski już w 2001 r. [PERLOVSKY 2001]. Na olbrzymi potencjał algorytmów logiki dynamicznej zastosowanej do sieci neuronowych w zdaniach wykrywania wielu niezdefiniowanych uprzednio obiektów wskazują też osiągnięcia sztucznych systemów wykorzystujących neuronowe pola modulujące, opisane w pracach Perlovsky’ego i Deminga [DEMING i inni 2009; PERLOVSKY i DEMING 2010; PERLOVSKY 2009]. W opisanych w cytowanych pracach systemach militarnych wykorzystujących algorytmy kognitywne uzyskiwano kilkudziesięciudecybelową poprawę wykrywalności systemów detekcyjnych.

Zrozumienie sceny i obiektów w niej występujących jest równoznaczne z dopasowaniem postrzeganych obiektów do modeli uprzednio rezydujących w naszej pamięci. Modele te mogą być wycinkowe, dotyczące rozpoznania obiektu, przebiegu zjawiska, gdy obiekt jest procesem, lub też dotyczyć mogą ogólnego modelu świata, w którym mózg funkcjonuje. Model mentalny jako podstawa poznania ma uznaną pozycję w psychologii [LAIRD 1993], lingwistyce (model sytuacyjny) [ZWAAN & RADVANSKY 1998] i robotyce [HAIKONEN 2012]. Po raz pierwszy jednak proces rozumienia został zdefiniowany w języku matematyki i odnosi się do całościowego opisu świadomego umysłu.

Z biologicznego punktu widzenia wzrost podobieństwa zaspokaja postulowany przez Perlovky’ego instynkt wiedzy (*Knowledge Instinct* – KI), który subiektywnie może być odczuwany jako pozytywna emocja estetyczna. NPM-DL, czyli neuronowe pola modelujące, działające z wykorzystaniem logiki dynamicznej, „odczuwają” satysfakcję.

#### 4. DOBROCZYNNE SKUTKI INSTYNKTU WIEDZY

Perlovski twierdzi, że udowodnił w ten sposób znaczenie emocji w procesie poznania, na co epistemologia i psychologia wskazywały od dawna. Według Perlovsky'ego instynkt wiedzy KI stanowi nowy rodzaj motywacji do działania. Realizowany jest przez osiągnięcie stanu wykrycia podobieństwa obserwowanych obiektów do modeli-konceptów istniejących uprzednio w pamięci. Postrzeżenia zmysłowe rozpoznawane są w ten sposób tworząc percepty. Jeśli mózg odbiera jednocześnie bodźce pochodzące od wielu zmysłów, to na percept składa się odpowiednio wiele modalności. Sieć neuronowa ma zdolność do asocjowania bodźców, które synchronicznie do niej docierają. Prezentując systemowi wzrokowemu wielokrotnie ten sam obiekt, wraz z towarzyszącym mu sygnałem dźwiękowym, nauczymy umysł kojarzyć percept z tym sygnałem. Umysł rozpoznawać może obiekt zarówno dostrzegając go jak i słysząc jego nazwę. Obiekty w formie procesów, kojarzone w ten sposób z ich nazwami, nauczyć mogą umysł rozpoznawania relacji czasowych. Tworzy się w ten sposób semantyka i syntaktyka języka. [PERLOVSKY & ILIN 2010; PERLOVSKY 2010; HICKOK & POEPEL 2007].

Jeśli obserwujemy scenę zawierającą wiele obiektów, to poszczególne pola modelujące rozpoznają poszczególne obiekty w procesie równoległego przetwarzania, tworząc jednocześnie neuronowe reprezentacje wszystkich tych obiektów. Te mentalne reprezentacje, przekazywane do wyższych warstw w semihierarchicznej strukturze przetwarzania, tworzą korelaty mentalne sceny, a ogólnie – otaczającej rzeczywistości, która może być rozpoznana i w ten sposób uświadomiona przez te wyższe pola. Postrzegane obiekty o złożonej strukturze oraz dynamicznie zmieniające się sceny rozpoznawane w wyższych warstwach neuronowych, mają coraz bardziej abstrakcyjny i symboliczny charakter. Dysponując językiem, umysł ma możliwość dokonywania operacji logicznych na pozyskanych w procesie uczenia kategoriach symbolicznych. Może to być również język matematyki, pozwalający osiągać niebywałą precyzję myślenia, według Kanta, niedostępną dla istot posługujących się tylko nieostrymi pojęciami-konceptami. W ten sposób w procesie RO umysł przechodzi od mglistych, rozmytych pojęć, niepoddających się arystotelesowskiej logice formalnej, do precyzyjnie zdefiniowanych, ostrych pojęć, tkwiących głęboko w tej logice. Te ostatnie, podobnie jak rezonanse Grossberga, rejestrowane są przez świadomość, podczas gdy pierwotne, mgliste modele-pojęcia pozostają ledwo uświadomione.

Kolejne doświadczenia życiowe będą uzupełniały wytworzony wcześniej światopogląd. Uczenie zwiększa precyzję naszego rozumienia rzeczywistości, a także poszerza dostępne modele o nowe szczegóły i w ten sposób pogłębia to rozumienie. Bogactwo i złożoność tworzącego się w ten sposób modelu świata określa stopień naszej świadomości. Emocja estetyczna, która temu towarzyszy, jest podstawą do dostrzegania harmonii i symetrii, co daje nam zdolność odbioru muzyki i wszelkiej sztuki. Zdolność do tworzenia modelu świata, jego opisu i rozumienia własnej roli w tym świecie jest istotą świadomości i samoświadomości.

Zaletą przedstawionej w zarysie teorii neuronalnych pól modelujących (NPM) jest głębokie jej osadzenie w historycznie rozwijanych kategoriach teorii poznania. Jest to bezpośrednia kontynuacja odwiecznej myśli filozoficznej, co budzi zaufanie do fundamentalnych założeń tej teorii. Co więcej, jej precyzyjne sformułowanie pozwala na opracowanie modelu i opisu matematycznego tej teorii. Umożliwiło to zbudowanie modeli elektronicznych, które praktycznie zademonstrowały skuteczność algorytmów NPM, przewyższającą wszelkie dotychczasowe usiłowania w tej dziedzinie [DEMING 2009; PERLOVSKY 2010]. Teoria ta pokazała także, jak emocje, czyli satysfakcja z zaspokojenia instynktu wiedzy, motywują sieć neuronową, a zatem i umysł, do procesu kognitywnego. Tęsknota Platońskiej duszy za wiedzą ze świata idei została w ten sposób matematycznie uzasadniona i objaśniona.

NPM wskazuje na jednolity charakter funkcji wykonywanych przez poszczególne pola modelujące, co oznacza jednorodną strukturę kolumn i mikrokolumn neuronowych w korze mózgowej. A przecież jest to zgodne z morfologicznym obrazem wszystkich obszarów i warstw korowych, co potwierdza hipotezę koneksjonistycznego charakteru sieci neuronowej tworzącej mózgi naturalne. Teoria NPM wyjaśnia także, jak umysł może dokonywać kompresji informacji dostarczanych przez zmysły. Pokazuje, że możemy zbudować symboliczny opis otaczającej nas rzeczywistości, omijając barierę złożoności tego świata CCC. Przechodzimy w ten sposób od szczegółowych obrazów sceny do pojęć symbolicznych, którymi możemy te sceny opisać. Ta wiedza symboliczna podlega arystotelesowskiej logice formalnej i jest w pełni uświadomiona. Dokonuje się w ten sposób gigantyczna kompresja informacji, która ma fundamentalne znaczenie dla skuteczności działania w otaczającym nas środowisku. Strumień informacji dostarczany przez zmysły sięgać może 10Mb/s dla bodźców wzrokowych, ok 1 Mb/s dla sygnałów

dźwiękowych, podczas gdy do skutecznego działania w naturalnym środowisku wystarczy przetwarzanie strumienia informacji na poziomie 10 b/s [DEECKE 2012]. Zredukowanie ilości przetwarzanej informacji ma duże znaczenie dla szybkości jej przetwarzania. Przede wszystkim umożliwia szybką reakcję świadomego systemu na dynamiczne zmiany środowiska. Co więcej, umożliwia także skompresowane przekazywanie wiedzy w procesie uczenia pod nadzorem, a przez to umożliwia istotom społecznym tworzenie kultury i budowanie cywilizacji. [PERLOVSKY 2005, 2007, 2008, 2009; PERLOVSKY i inni 2012]. Być może ta szczególna właściwość NPM, którą wykazują sieci neuronowe o odpowiedniej organizacji, była ewolucyjną przyczyną pojawienia się świadomości w organizmach żywych. Zdolność do szybkiej i adekwatnej reakcji daje przecież znaczącą przewagę ewolucyjną.

Koncepcja Neuronowych Pól Modelujących NPM i towarzyszący jej Instynkt Wiedzy KI ma rewolucyjne konsekwencje w interpretacji dualizmu sfery duchowej i cielesnej. Zauważmy, że wykorzystywana przez NPM logika dynamiczna LD ma charakter ściśle przyczynowy. Działa ona jak równania matematyczne opisujące przyczynowo-skutkowe relacje procesów fizycznych. Jednocześnie towarzyszy jej instynkt wiedzy KI, którego działanie jest absolutnie celowe i nakierowane na pozyskiwanie wiedzy o świecie. Tworzenie struktur umysłu ma dynamikę przyczynową i równocześnie teleologiczną, usuwając fundamentalną dychotomię ujawnioną przez Arystotelesowskie pojmowanie relacji formy do materii. Dualizm nie jest już potrzebny. Działanie fizyczne, przyczynowe jest w tym przypadku matematycznie równoważne działaniu kierowanemu celowo, teleologicznie. Nie ma sensu wybierać między duchem i materią, ponieważ niematerialna świadomość wyłania się emergentnie jako konsekwencja instynktu wiedzy KI generowanego logiką dynamiczną DL zastosowaną do neuronowych pól modelujących NPM.

Czy to oznacza, że teoria neuronowych pól modelujących NPM wyjaśnia całkowicie, jak pojawia się ludzka świadomość? Czy wyjaśnia wszystkie aspekty funkcjonowania mózgu ludzkiego lub zwierzęcego? Czy pozwala matematycznie przewidzieć nasze reakcje? Czy pozwoli skopiować mózg ludzki i zbudować maszyny rozumiejące to, co robią, na wzór ludzi? A może wykreuje istoty o stopniu inteligencji i świadomości przewyższającym ludzkie możliwości?

## 5. NIEDOSTATKI MODELU NPM

Model NPM nie aspiruje do wyjaśnienia wszystkich funkcji mózgu i nie odpowiada na wszystkie postawione wyżej pytania. Spekulatywna hipoteza Perlovsky'ego odnosi się do uproszczonego modelu struktur mózgowych w postaci zbioru neuronów przypominającego sztuczne sieci neuronowe i stara się objaśnić wyłącznie funkcje kognitywne takiej sieci, choć to właśnie te są dla nas najbardziej interesujące. Niestety NPM to tylko zgrubny model, model bardzo przekonujący, jednakże obarczony wieloma ograniczeniami. Najbardziej istotne to fakt, że w rzeczywistości naturalna sieć neuronowa nie przypomina sztucznej i nie wykonuje żadnych zaprogramowanych, złożonych obliczeń. W szczególności nie oblicza pochodnych funkcji podobieństwa ani nie rozwiązuje równań różniczkowych choćby pierwszego rzędu. Koncepcja NPM nie podaje także sposobów biologicznych umieszczania modeli w pamięci tak, aby były one gotowe do kolejnych prezentacji napływającym nowym postrzeżeniami. Brak jest mechanizmu powiązania sygnałów sensorycznych różnych modalności (*binding problem*), co stawia pod znakiem zapytania możliwość uczenia korelacyjnego, które jest z kolei podstawą samouczenia, nabywania doświadczeń i tworzenia wokół nich modelu środowiska i świata.

Ponadto NPM, choć operuje logiką dynamiczną (LD) w odniesieniu do zmieniającego się w kolejnych iteracjach modelu, to dotyczy obiektów stacjonarnych. Ruch sceny, skutkujący zmianami obserwowanych obiektów w czasie, znacznie by skomplikował przytoczone przez Perlovsky'ego równania. Oczywiście można założyć, że proces kognitywny przebiegał będzie tak szybko, że w infimezmalnym czasie percepcji obiekt pozostanie nieruchomy. Równania LD nie uwzględniają jednak przypadku ogólnego. Do tego celu konieczne byłoby wyróżnienie pamięci epizodycznej. Lecz postulat ten jest bez sensu, jeśli model NPM do sposobu zapamiętywania modeli w ogóle się nie odnosi. Co gorsza, brak neurologicznego mechanizmu asocjacji multimodalnych oraz brak mechanizmu pamięci epizodycznej nie wskazuje praktycznie, jak pojawiać się mogą zdolności lingwistyczne. Problem w tym, że zarówno generacja, jak i zapamiętywanie mowy jest równoważne zapamiętywaniu epizodów rozciągniętych w czasie, które muszą wykorzystywać pamięć epizodyczną. Brak zaś języka uniemożliwia przekazywanie i uczenie się złożonych pojęć ogólnych o abstrakcyjnym charakterze, a więc tworzenie wiedzy symbolicznej. Uczenie i intencjonalne przekazywanie wiedzy symbolicznej o charakterze abstrakcyjnym jest możliwe tylko za po-



mocą złożonego języka symbolicznego kontrolowanego świadomie. Wynika to z faktu, że pojęcia abstrakcyjne, w odróżnieniu od obserwowanych obiektów, nie mogą być bezpośrednio, bez pośrednictwa języka, zaprezentowane naszym zmysłom. Z tego powodu nie mogą one być przyswojone jako modele zakorzenione w świadomości i ugruntowane w pamięci semantycznej lub epizodycznej, jak to się dzieje w przypadku qualiów.

Model NPM abstrahuje od konkretnych struktur neuronowych. Nie wskazuje żadnego mechanizmu neuronowego, który mógłby realizować postulowane funkcje. Brak jest także mechanizmów generacji sygnałów *top-down*. Perlovsky wskazuje, że sieć neuronowa o architekturze NPM posiada motywację emocjonalną w postaci Instynktu Wiedzy (KI). Sugeruje, że to KI powoduje intencjonalnym zachowaniem sieci. Ten jednak instynkt w NPM jest określony behawioralnie rezultatem. Instynkt to przymus biologiczny działania w określony sposób. Sieć neuronowa nie posiada jednak software'u narzucającego algorytm działania prowadzący do KI. Z jej koneksjonistycznego charakteru wynika, że powinna istnieć przyczyna morfologiczna i fizjologiczna wymuszająca pojawienie się takiego instynktu. Perlovsky jednak na nią nie wskazuje. Charakter tego instynktu ma głęboki związek z motywacjami do funkcjonowania, o czym w dalszej części będziemy mówić. Warto przypomnieć, że kognitywistka na pierwszym miejscu wymienia inne motywacje autonomicznych systemów inteligentnych (jakimi są także ludzie i zwierzęta) do racjonalnego działania. Jeśli zatem NPM porównują sygnały *top-down* i *bottom-up*, to KI w rozumieniu Perlovskiego będzie się manifestował. Jeśli jednak mechanizm takiego instynktu nie jest nam znany, to praktyczne działanie NPM może być kwestionowane.

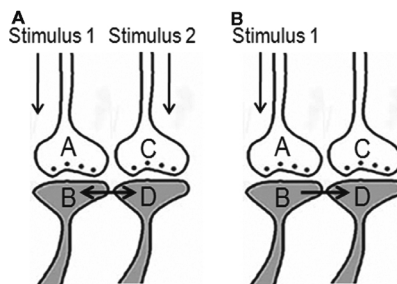
Czy wobec tego powinniśmy odrzucić tę teorię? Niekoniecznie. Teoria ta jest użyteczna w modelowaniu procesów kognitywnych AI, a jej algorytmy skutecznie pokonują barierę złożoności kombinatorycznej CCC. Jeszcze bardziej istotne są wnioski, które z niej wypływają, a istnieje wiele wskazań, że kora mózgow naturalnych działa w sposób wielce podobny do algorytmów NPM – LD.

## 6. BIOLOGICZNE MECHANIZMY NPM

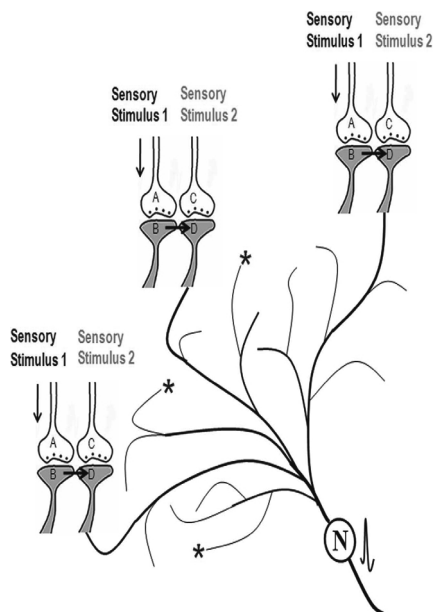
W sukurs NPM przychodzi neurologiczna koncepcja sprzężenia międzysynaptycznego rozwijana przez Kunjumona I. Vadakkana [VADAKKAN 2010,

2011, 2012]. Według tej koncepcji mózg ma architekturę koneksjonistyczną, tworzoną przez sieć neuronów o wielowarstwowej strukturze hierarchicznej, której węzły stanowią klasyczne neurony wyposażone w dużą liczbę synaps. Wiedza nabywana w trakcie uczenia utrwalana jest w postaci wzmocnienia synaptycznego LTP. Vadakkan zwraca uwagę, że dominującym mechanizmem samouczenia jest uczenie korelacyjne, względnie asocjacyjne, podczas którego następuje synchroniczne pobudzenia pól rozróżniających specyficzne, wzajemnie skorelowane cechy obiektu lub też pochodzące od różnych zmysłów (modalności). Postrzegane obiekty indukują wzorce pobudzeń neuronalnych w różnych polach lub modalnościach, które ze względu na synchronizację w czasie asocjują, tworząc nowy wzorec będący perceptem, umożliwiającym identyfikację obiektu, jeśli podobne pobudzenia pojawią się w przyszłości. Następuje wzajemne powiązanie tych wzorców pobudzeń (*binding*) [MALSBURG 1981]. Vadakkan postuluje, że ta asocjacja przebiega dzięki funkcjonalnemu sprzężeniu synaps znajdujących się na szlakach aferentnych jak przedstawiono to na rys. 1.

Rys. 1. Powstawanie sprzężenia między synapsami podczas uczenia (według K. Vadakkana [2012], za zezwoleniem autora).



- (A) Formowanie Funkcjonalnego Sprzężenia Synaptycznego FSS pomiędzy dwiema postsynapsami B i D podczas uczenia, czyli jednoczesnego pobudzenia bodźcami Stimulus 1 (S1) i Stimulus 2 (S2). A i C to odpowiadające im presynapsy.
- (B) Podczas przypominania, w obecności bodźca Stimuli 1, reaktywowane jest sprzężenie funkcjonalne FSS, wskutek czego postsynaptyczna membrana D jest także aktywowana. Aktywacja postsynapsy D indukuje impresjon aktywności, tak jak gdyby stymulowana była również presynapsa C. Tworzy to „iluzję” występowania bodźca S2.



Schemat aktywacji postsynapsy Dn poprzez Funkcjonalne Sprzężenie Synaptyczne FSS prowadzi do generacji potencjału czynnościowego na wierzchołku neuronu N (przyjmując, że przestrzenne i/lub czasowe sumowanie wystarczającego potencjału czynnościowego przed- i po-synaptycznego jest możliwe). W przypadku pobudzenia presynaps An tylko jednym bodźcem pobudzającym S1, synaptyczny impresjon indukowany jest w każdej synapsie D. Dendrytyczne odgałęzienia neuronu N zaznaczone gwiazdkami to postsynapsy, gdzie dochodzi pobudzający potencjał postsynaptyczny PPP, pochodzący z normalnej aktywności oscylacyjnej niższych warstw neuronowych. Potencjały PPP nie są wystarczające do pobudzenia neuronu N. Dopiero zsumowane z pobudzeniem S1 mogą wygenerować potencjał czynnościowy neuronu N.

Powstające w ten sposób powiązanie postsynaps nazywać będziemy Funkcjonalnym Sprzężeniem Synaptycznym (FSS). FSS pojawia się między synapsami B i D na rys. 1 w trakcie równoczesnego postrzegania dwóch aspektów sceny, co odpowiada procesowi samouczenia korelacyjnego (RL) w kategoriach AI. Po wytworzeniu takiego połączenia, pobudzenie wstępujące (*bottom-up*) jednej z synaps (B), indukowało będzie także pobudzenie synapsy sąsiedniej (D), co rejestrowane będzie jako pojawienie się wrażenia lub wyobrażenia (impresji) o postrzeganiu także drugiego aspektu sceny (lub pobudzenia drugiej modalności), który zwykle temu pierwszemu towarzyszy. Takie wrażenie/wyobrażenie indukowane w sprzężonej synapsie, gdy jedna z nich jest aktywowana, nazywać będziemy „impresją synaptyczną”. Postsynapsa D doświadcza wówczas komórkowej halucynacji, że aktywowany jest szlak pobudzeń neuronalnych, który normalnie aktywuje się doprowadzając zwykle pobudzenie do presynapsy C. Indukowany w postsynapsie D pobudzający potencjał postsynaptyczny, przekazywany w drzewiastej strukturze sieci komórkowej do wyższej neuronowej warstwy przetwarzającej, podlega sumowaniu przestrzennemu i czasowemu wraz z innymi pobudzeniami. Neurony wyższej warstwy traktują te pobudzenia identycznie jak pobudzenia płynące z licznego podzbioru pobudzeń sensorycznych odpowiadającemu matrycy/wzorcowi cechy postrzeganego obiektu, tworząc jego wirtualną halucynację. Hierarchiczna drzewiasta struktura sieci neuronowej tworzy wstępujące drzewo pobudzeń o licznych wejściach sensorycznych N i zmniejsz-

szającej się liczbie węzłów w wyższych warstwach sieci, aż do pojedynczego neuronu Z i należącej do niego presynapsy C, jak to przedstawiono na rys. 2.

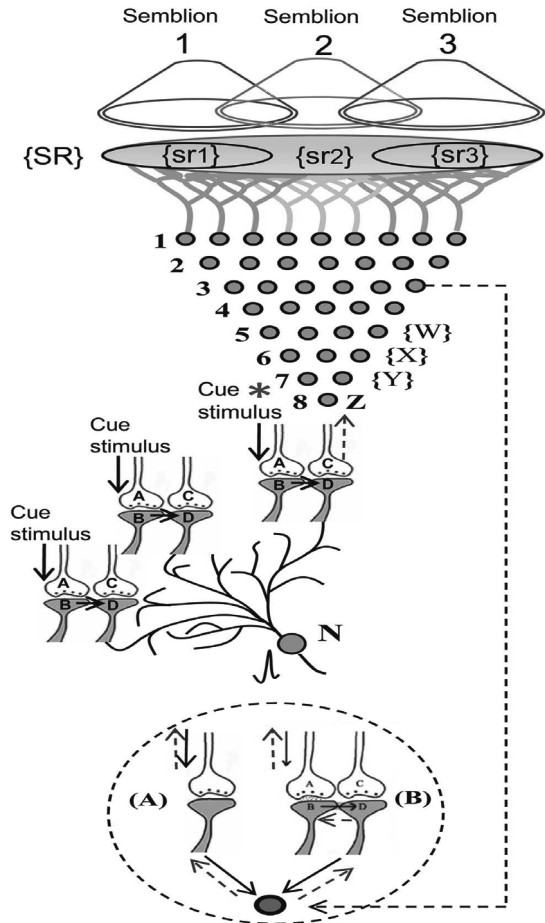
Rys. 2. Schemat tworzenia impresjonów sensorycznych i ich indukowania, gdy następuje pobudzenie sprzężonego neuronu lub synaps (według K. Vadakkana [2012], za zezwoleniem autora).

Szare kółeczka symbolizują neurony kolejnych warstw 1-8. {SR} – receptory sensoryczne {sr1} – {sr3}. Podczas przypominania sygnał stymulujący (oznaczony gwiazdką) dociera do wejścia presynaptycznego A i depolaryzuje membranę postsynaptyczną B, wzbudzając w niej postsynaptyczny potencjał czynnościowy PPP. Poprzez FSS aktywowana jest także postsynapsa D (jak na rys.1).

Pobudzenie membrany D generuje komórkową iluzję identyczną, jak gdyby to presynapsa C była pobudzona (synapsa C należy do neuronu Z). Pobudzenie Z jest równoważne pobudzeniu całej hierarchicznej struktury kolejnych warstw {W} → {X} → {Y} → Z. Neurony projekcyjne oraz równoległe zwrotne, mogą aktywować niższe pola neuronowe (na tym rysunku wyższe, bliższe polom sensorycznym), (pobudzenia zwrotne, nie są pokazane na rysunku). Stymulacja każdego z podzespołów receptorów sensorów {sr1}, {sr2}, {sr3} lub ich dowolnej kombinacji i sumy, może aktywować neuron Z. Te zestawy receptorów zdolne do pobudzenia Z nazywamy impresjonami odpowiednio 1,2,3 (Semblion 1,2,3).

Impuls pobudzający powoduje, że postsynapsy D doznają iluzji pobudzenia impresjonem 1,2 lub 3 lub dowolną kombinacją tych impresjonów. Sposób scałania i porównywania impresjonów, tak, aby pasowały do wewnętrznego wrażenia indukowanego przez sygnał pobudzający, jest poszukiwaną heurystyką neuronu i kolumny neuronowej, zbliżoną do algorytmu Perlovsky'ego.

Takie powiązane, drzewiaste struktury, reprezentujące konfigurację pobudzeń sensorycznych, odpowiadające modalnościom lub cechom obiektów proponuję nazywać impresjonami, przez analogie do Vadakkanowskich semblionów (od angielskiego *semblance*).



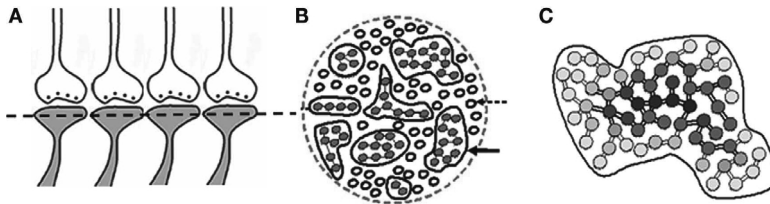
Vadakkan twierdzi, że indukowane przez FSS impresje synaptyczne wyższych poziomów, czyli impresjony, mogą być dopasowywane do perceptów i pozwalać nam w ten sposób identyfikować obiekty i ich cechy. Dotyczy to obiektów o cechach już nam znanych, których impresjony utrwalone zostały w naszej pamięci na podstawie wcześniejszych doświadczeń. Natomiast postrzeżenia związane z nowymi obiektami lub ich nowymi cechami, indukowane będą zestawy impresjonów znanych, uzupełnione przez nowe sprzężenia FSS, przez co formowane będą nowe impresjony, o bogatszej strukturze, co stanie się czynnikiem wzbogacającym naszą świadomość.

Tworzenie impresjonów może zachodzić na wielu poziomach wielowarstwowej struktury sieci. Czas opóźnienia odpalenia piku depolaryzacyjnego potencjału czynnościowego neuronu (actionpotential AP), wyzwolonego zsynchronizowanym pobudzeniem jego synaps dendrytycznych, wynosi ok. 1-2 milisekundy. Ponieważ typowy czas konsolidacji wrażeń (np. wrażenie ciągłości obrazu poklatkowego filmu) wynosi ok.  $\frac{1}{16}$  sekundy, można antycypować, że przetwarzanie sygnału wejściowego z pól sensorycznych może zachodzić nawet w 30-50 warstwach. Około 40 pobudzeń synaptycznych zsumowanych przestrzennie (lub nieco mniejsza liczba pobudzeń sumowana w czasie) może wyzwolić impuls depolaryzacyjny neuronu Z (spośród kilku do kilkudziesięciu tysięcy synaps występujących w każdym neuronie), co oznacza identyfikację specyficznego stanu warstwy neuronów, rejestrowanego przez wybrane, pobudzane synapsy niższego poziomu. Liczba synaps przekazujących swoje stany pobudzenia do synaps wyższych warstw rośnie wykładniczo. Oznacza to, że liczba impresjonów, czyli dostępnych reprezentacji postrzeganych przez nasze zmysły obiektów, wyznaczona kombinatorycznie dla reprezentacji przestrzennych lub poprzez permutacje impulsów pobudzających drzewiastą strukturę synaptyczną, rozłożonych w czasie, jest praktycznie nieskończona. Dotyczy to nawet mózgów dużo prostszych niż ludzki.

Warto zauważyć, iż – jak zaobserwowano eksperymentalnie – stymulacja różnych obszarów mózgu powoduje halucynacje. Można stąd wnioskować, że kiedy pośredni węzeł sieci podlegający neuronalnej aktywności oscylacyjnej będzie stymulowany, system postrzega i rozpoznaje wewnętrzne pobudzenie jako swoistą cechę lub właściwość [ANASTASSIOU i in. 2010, 2011]. Poprzeczne zatem pobudzenia pochodzące od sygnałów stymulujących, w wyodrębnionym obszarze ścieżki neuronowej, w której wystąpiły zmiany indukowane uczeniem asocjacyjnym, powinny wykazywać skuteczny mecha-

nizm komórkowy indukowania wewnętrznych wrażeń podczas procesu przypominania. Takie zjawisko może być wyjaśnione mechanizmem FSS i traktowane jako cecha strukturalna, odwracalna i stabilizująca mechanizmy asocjacyjne i pamięć. Kombinatoryczne sumowanie impresjonów różnego neuronalnego poziomu, kreuje zróżnicowane wyższe funkcje umysłowe [MILSTEIN & KOCH 2008]. Kontynuacja uczenia asocjacyjnego/korelacyjnego prowadzi do uformowania FSS między postsynapsami, które uprzednio utworzyły FSS z innymi postsynapsami. Te zgrupowania wzajemnie powiązanych synaps, które w przekroju ścieżki transmisji bodźców neuronowych tworzą wysepki synaptyczne. Mogą one być postrzegane jako rdzeń wiązki połączeń przewodzących, np. kolumny lub mikrokolumny. Być może przywraca to kolumnom i mikrokolumnom zrozumiałe funkcje, kwestionowane w pracy Hortona i Adamsa [HORTON & ADAMS 2005]. Zgrupowania wysepki synaptycznych w warstwie neuronowej uznawać można hipotetycznie za neuronowe pola modulujące NPM (rys.3).

Rys. 3. Wyspy postsynaptyczne powiązane FSS\_(według K. Vadakkana [2012], za zezwoleniem autora).



- (A) Cztery synapsy powiązane FSS. Linia przerywana symbolizuje płaszczyznę przekroju tych i innych synaps na sąsiednich rysunkach.
- (B) Przekrój przez większą liczbę synaps. Podczas uczenia formuje się FSS pomiędzy równocześnie pobudzonymi postsynapsami. Dalsze uczenie powiększa liczbę synaps powiązanych FSS. Formują one wysepki synaptyczne zaznaczone strzałką.
- (C) Schemat pojedynczej dużej wysepki synaptycznej. Nowo przyłączone synapsy pozostają na peryferiach wysepki. Często używane połączenia synaptyczne FSS znajdują się w centrum wysepki.

Założenie dotyczące przekazywania pobudzeń pomiędzy synapsami na tym samym poziomie hierarchii rozpoznawania wzorców były przyjmowane w modelach obliczeniowych przez Favorova i Kursuna [Favorov & Kursun 2011]. Badacze Ci udowodnili, że używając reguł Hebb'a do wyznaczania wag połączeń aferentnych dokonali konwersji konwencjonalnej sieci neuronowej do sieci RBF (*RadialBasisFunction*), która odznacza się zdolnością

do aproksymacji szerokiej klasy funkcji wejściowych [LOWE 2003; FAVOROV & KURSUN 2011]. Aczkolwiek statystyczne metody aproksymacji zmiennych liczbowych nie są poszukiwanymi przez nas algorytmami działania neuronowej sieci kortykałnej, jednakże zdolność do linearyzacji sygnałów wejściowych świadczy o sensowności przyjętego założenia dotyczącego horyzontalnego oddziaływania sąsiadujących neuronów. Co więcej założenie to, dotyczące przesyłania pobudzeń w kierunku poprzecznym do komórek sąsiednich, zostało dobrze udokumentowane badaniami neurologicznymi [PETERSEN & SAKMANN 2000; FELDMEYER i in. 1999; THOMSON i in. 2002; BINZEGGER i in. 2004].

Mimo potwierdzenia występowania sprzężeń międzysynaptycznych poprzecznych, horyzontalnych dotychczas prowadzone badania fizjologiczne nie doprowadziły do wykrycia biofizycznych lub biochemicznych zmian w sąsiadujących ze sobą synapsach w procesie uczenia korelacyjnego. Nie jest znany mechanizm takiego oddziaływania. Ten postulat FSS wynika jedynie z logiki działania sieci neuronowej tak, aby przy maksymalnej prostocie i jednorodności mogła ona realizować oczekiwane funkcje porównywania wzorców i jest próbą wyjaśnienia obserwowanych asocjacji horyzontalnych.

Jakie mechanizmy mogłyby zatem, choć teoretycznie, odpowiadać za tego rodzaju sprzężenie?

Vadakkan wymienia następujące zjawiska na poziomie synaptycznym, neuronowym, systemowym i behawioralnym, które mogą być odpowiedzialne za efekt formowania FSS i powstawanie impresjonów:

- poziom synaptyczny: a) indukcyjna depolaryzacja jednej postsynapsy przez sąsiednią, b) sprzężenie efatyczne, podczas którego przestrzeń międzysynaptyczna zaburzona jest efatycznymi zmianami potencjału na membranie drugiej postsynapsy [ANASTASSIOU 2010], c) informacja może być przenoszona przez proteiny czułe na potencjały czynnościowe.
- poziom neuronowy: d) otwieranie dodatkowego kanału jonowego [Tye i inni 2008] i rozpraszanie neurotransmitera AMPA do sąsiednich synaps [COGGAN i in. 2005, MAKINO & MALINOW 2009].
- poziom systemowy: e) fale i oscylacje mózgowe, szczególnie oscylacje gamma  $\gamma$ , powodujące sprzężenia podczas przypominania, możliwe do wizualizacji za pomocą współczesnych metod zobrazowania. Synchronizacji oscylacji  $\gamma$  przypisuje się dwie główne funkcje: sprzężenie neuronów reprezentujących różne cechy danego bodźca, umożliwiające

jego odróżnienie od „tła” [FRIES i in. 2001; Fell i inni 2003] oraz udział w formowaniu pamięci krótkotrwałej [LISMAN & IDIART 1995] i deklaratywnej [FELL i in. 2002; GRUBER i in. 2005]. Uczenie korelacyjne i asocjacyjne następuje poprzez korelacje czasową sygnałów z dokładnością milisekundową. Efekt może być wzmocniony przez oscylacyjną działalność mózgu, ponieważ pobudzenie nawet powyżej 40 synaps (co uznawane jest za minimalną liczbę spośród od średnio 24 do 80 tysięcy synaps dla jednego neuronu, powodującą generację piku pobudzenia) może nie spowodować wyzwolenia piku potencjału czynnościowego AP. Fale mózgowie mogą jednak synchronicznie zbliżyć potencjał na postsynapsach do wartości podprogowej. Wówczas nawet słabsze sygnały mogą synchronicznie przekraczać próg pobudzenia tworząc asocjację.

Występowanie FSS kreuje impresjony, jako drzewiaste struktury wielopoziomowe, które pobudzane sygnałami o wektorach horyzontalnych, w wielopoziomowej strukturze sieci, lub spontanicznie, np. na skutek oscylacyjnej aktywności mózgu, generować mogą wspomnienia i wyobrażenia (podobnie jak wspomniane wcześniej „halucynacje” pojedynczego neuronu) [VADAKKAN 2010]. Ta funkcja kreowania wyobrażeń może mieć fundamentalne znaczenie dla mechanizmów planowania i zdolności kreatywnych umysłu, ponieważ pozwala na wyobrażanie sobie skutków zaplanowanego działania. Utworzenie drzewiastej, hierarchicznej struktury impresjonów bardzo przypomina wspomniane na wstępie struktury K-line Minsky’ego. I podobnie jak to opisywał Minsky [MINSKY 1980], mogą one być skutecznym narzędziem gromadzenia i przechowywania wiedzy, uczenia korelacyjnego, podejmowania decyzji i innych funkcji poznawczych i zarządczych umysłu [MINSKY 1993].

Nowe, wyższego rzędu impresjony mogą się tworzyć pod wpływem nowych bodźców pobudzających, postrzegania nowych obiektów i nowych sytuacji. Ponieważ są one specyficzne dla określonych ścieżek pobudzeń i wysepek synaptycznych, tworzenie nowych impresjonów nie narusza struktury wzorców poprzednio utrwalonych w pamięci trwałej. Nowe impresjony mogą natomiast oddziaływać z dotychczasową siecią impresjonów. Nowe sygnały pobudzające mogą dzięki temu wykluczać lub wspierać przywoływane impresjony – modele [ABBOTT 2008]. Następuje dobrze nam znane porównanie modeli z perceptami. Zaakceptowanie wybranego impresjonu lub zespołu impresjonów oznacza zrozumienie nowych bodźców. Niewielkie modyfikacje impresjonu mogą być utrwalone w sąsiednich wysepkach synaptycznych i wiązkach, przewodzących informacje niezbędne mózgowi do tworzenia szerokiego modelu środowiska.



Na pozór wydaje się, że koncepcja FSS proponuje skuteczny neuronalno-synaptyczny mechanizm porównywania wzorców niezbędny dla skuteczności matematycznego modelu Perlovsky'ego. W dodatku to dopasowywanie wzorców pobudzeń, czyli impresjonów różnych poziomów hierarchicznej struktury przetwarzania informacji w mózgu, odbywa się bez matematycznych przekształceń według ustalonych z góry, złożonych algorytmów. Wysepki synaptyczne i ścieżki transmisji sygnałów neuronowych łatwo utożsamić z mikrokolumnami i kolumnami korowymi oraz z neuronowymi polami modelującymi NPM. Model FSS uwzględnia plastyczność sieci umożliwiającą uczenie korelacyjne. Pokazuje, jak mogą tworzyć się agregaty impresjonów na tyle bogate, że mogą odwzorować złożoność wszechświata. Dopuszcza sprzężenia zwrotne pobudzające całe „drzewa” impresjonów obejmujące miliardy synaps i neuronów. Pamiętać także należy, że wysepki synaptyczne sprzęgają neurony, jak również synapsy należące do innych, sąsiednich szlaków wstępujących, aferentnych, przez co pobudzane są równie bogate struktury impresjonów wysokich poziomów, tworząc – według Vadakkana – wyższe funkcje umysłowe.

Vadakkan używa modelu FSS i tworzonych przezeń impresjonów do objaśnienia wielu zjawisk mózgowych i umysłowych. Wskazuje, że mechanizm tworzenia impresjonów będzie podobny dla wszystkich istot żyjących w tym samym środowisku, jeśli przyjąć założenie, że procesy ewolucyjne wyposażyły je w zmysły umożliwiające poznanie tego środowiska. Wynika z tego adekwatność qualiów do rzeczywistości, które tak bardzo martwią filozofów umysłu [DENNETT 1988; CHALMERS 1995, 1996; RAMACHANDRAN & HUBBARD 2001]. Qualia mogą być objaśniane w ramach prezentowanego modelu jako impresjony drugiego rzędu, wrodzone dla poszczególnych gatunków, lecz później modyfikowane poprzez doświadczenie w trakcie rozwoju psychicznego. Ponieważ towarzyszą one stanom świadomym, to dotyczyć mogą jedynie gatunków o pewnym minimalnym stopniu świadomości. Model dopuszcza odwrotną sytuację: pojawienie się konglomeratu impresjonów najwyższego rzędu, będącego odpowiednikiem świadomości, bez impresjonów niższego rzędu reprezentujących qualia. Musiały one jednak uprzednio zaistnieć w celu utworzenia świadomości, choć później może się ona bez qualiów pewnych typów odbywać. Jest to zgodne z obserwacjami psychologicznymi umysłów naturalnych [CHALMERS 1998], kiedy umysł, dysponując wiedzą symboliczną, może ją przetwarzać logicznie i wyciągać wnioski, mimo że wrażenia (qualia), będące podstawą wiedzy symbolicznej, zostały już dawno zapomniane. Model przewiduje, że impresjony powstające

w trakcie rozwoju osobniczego odzwierciedlały będą dynamikę ich powstawania zbliżoną do występującej u innych osobników tego samego gatunku, w tym samym wieku, co spowoduje dużą jednorodność impresjonów w całej populacji.

Wielką zaletą modelu FSS jest wyjaśnienie powstawania wyobrażeń. Wyobrażenia powstają przez indukcję sprzężonych impresjonów po pobudzeniu przez bodźce tematycznie skojarzone. Zdolność wyobrażenia sobie obiektów niedostępnych aktualnie dla zmysłów jest niezbędna dla przewidywania skutków swego działania. Umożliwia więc planowanie, jak osiągnąć zamierzony cel, a więc i zachowania celowe. Wyobrażenie kombinacji obiektów, czyli agregacja indukowanych mechanizmem FSS impresjonów, może być podstawą kreatywności.

Jeśli impresjony generowane przez FSS odpowiedzialne są za powstawanie świadomości, jak sugeruje autor, to model można testować, przewidując, jakie skutki dla świadomości może powodować przerwanie FSS na skutek działania substancji chemicznych, narkotyków, psychotropów lub stanów urazowych lub patologicznych. Jednocześnie testować można zdolność modelu FSS do wytłumaczenia spotykanych zaburzeń w funkcjonowaniu przytomności. Autor dokonuje takiej pozytywnej weryfikacji na przykładzie fenomenu snu, działania środków anestetycznych, stanów narkotycznych, delirium, padaczki napadowej, niedotlenienia mózgu, hipoglikemii, chronicznych stanów utraty przytomności [BERNAT 2009] itp. [VADAKKAN 2010].

Kolejne dowody neurologiczne to utrata przytomności wskutek blokowania transmisji sygnałów sensorycznych w wyniku zawału wzgórza grzbietowego i innych jąder wzgórza, przez które przechodzą wszystkie szlaki aferentne pobudzeń sensorycznych. Wiadomo skądinąd, że deprywacja sensoryczna zakłóca świadomość. Autor wyjaśnia efekt asocjacyjnego zapamiętywania w przypadku braku przytomności, podświadomą reaktywacją FSS i powstawaniem nowych, zasocjowanych impresjonów. Najistotniejsze wydaje się autorowi wyjaśnienie zjawiska powiązania korelacyjnego różnych modalności (*binding problem*). Obiektywnie najważniejszym osiągnięciem modelu FSS jest zaproponowanie mechanizmu porównywania impresjonów różnych rzędów, co umożliwi konfrontowanie impresjonów-modeli przechowywanych w pamięci z nowo utworzonymi impresjonami dostarczonymi nam przez zmysły.

## 7. DOBRY MODEL, ALE NIE DZIAŁA

Model funkcjonalnego sprzężenia synaptycznego w dalszym ciągu nie pozwala zrozumieć i opisać, jak kształtuje się świadomość w mózgu. Widzimy narzędzia, które mogą być wykorzystane do tego opisu. Autorzy wskazują na konieczność występowania połączeń umożliwiających horyzontalną propagację pobudzeń i uczenie korelacyjne. Wskazują na wiele przesłańek, że takie procesy występują w behawioralnych cechach mózgu i psychologicznych zjawiskach umysłu. Z tym należy się zgodzić. Nie natrafiamy tu na sprzeczności z dotychczasowymi modelami procesów poznawczych, a raczej można ten model traktować jako rozwinięcie nurtu nominalistycznego, idei Minsky'ego i uzupełnienia modelu Perlovsky'ego.

Nie jest jednak jasne, dlaczego impresjony różnych poziomów miałyby tworzyć kombinatoryczne agregaty i permutować, tworząc złożony model jeszcze bardziej złożonej rzeczywistości? Co powoduje selekcję impresjonów władających chwilowo naszą świadomością? Jakie jest neuronalne podłoże pamięci trwałej i roboczej i jakie są relacje między nimi? Jakie są procedury zdobywania, klasyfikacji i dystrybucji wiedzy, samouczenia, uczenia pod nadzorem lub z krytykiem? Jak rejestrowane są epizody trwające przez zmienne interwały czasowe, jeśli niezdefiniowana została pamięć epizodyczna systemu? Jaka jest motywacja do myślenia? Autor nie wspomina tu o dążeniu poznawczym motywowanym instynktem wiedzy KI, jak to proponował Perlovsky.

Podobnie jak w modelu NPM nie wyjaśniono mechanizmu uwagi. Nie odniesiono się do oddziaływania umysłu z otoczeniem poprzez jego funkcje motoryczne oraz do wzajemnych relacji między intencjonalną ingerencją w środowisko i tworzeniem jego modelu w umyśle.

Autor koncepcji FSS, podążając za niektórymi badaczami, nie wiąże „uwagi” ze stanami świadomymi. Przypisuje jej tylko pewną rolę na etapie tworzenia FSS. To stanowisko wydaje się błędne, ponieważ nie wskazuje jak mechanizm wolicjonalny przełączania uwagi mógłby ujawniać teleologiczną naturę istoty świadomej, a więc dopuszczać działania intencjonalne i wolną wolę.

Vadakkan nie wskazuje algorytmu, który ustanawia reguły integracji i konsolidacji grup impresjonów, tak aby retrospekcje impresjonów-modeli odpowiadały obserwowanym qualiom. Istotne jest, którego rzędu warstwy neuronowej synapsy muszą asocjować, aby ustanowić impresjon adekwatny do złożoności obserwowanego obiektu. Nie znajdujemy odpowiedzi na takie pytanie w zaproponowanym modelu. Niezbędne byłyby także algorytmy

ustanawiające liczbę warstw neuronowych objętych kształtowaniem impresjonu, określające liczbę i rozmiar wysepek synaptycznych na poszczególnych poziomach hierarchii, algorytmy kombinacji i permutacji synaps, wymaganych do wygenerowania potencjałów czynnościowych, włączania nowych neuronów, dodawania i usuwania synaps oraz określania ścieżek równoległego przetwarzania i sposobu wypełniania funkcji hamujących. Szczególnie ta ostatnia rola jest istotna w sieci pełnej sprzężeń zwrotnych, które łatwo mogą prowadzić do wzbudzenia całego układu, co w mózгах naturalnych manifestuje się objawami patologicznymi (np. epilepsją).

Proponowane powyżej biologiczne i systemowe mechanizmy tworzenia FSS mogą być łatwo podważone. W każdym razie w formie sugerowanej przez Vadakkana, nie zostały dotychczas potwierdzone i dlatego autor postuluje poszukiwanie innych, koniecznych mechanizmów sprzężeń synaptycznych. Zadanie to stawia przed neurologami i biochemikami badającymi procesy mózgowe.

Co gorsza, można się zastanawiać, czy sprzężenia synaptyczne są niezbędne, bo przecież rolę transmisji wzorców – modeli z wyższych warstw sieci mogą pełnić aksony sprzężeń zwrotnych zstępujące na piętra niższe, bliższe polom sensorycznym, oraz tzw. aksony rekurencyjne (zob. np. [BRODZIAK 2001]). Natomiast sprzężenia horyzontalne mogą być realizowane przez aksony asocjacyjne, będące torami transmisji informacji między polami różnych modalności. Występowanie wymienionych pętli sprzężeń wykazano eksperymentalnie i zapewne odgrywają one istotną rolę w tworzeniu asocjacji multimodalnych oraz penetracji zasobów mentalnych, o czym będzie jeszcze mowa w dalszej części pracy. Pętle krótkozasięgowo tworzone przez aksony rekurencyjne rozpatrywane są także jako mechanizm pamięci krótkotrwałej, ponieważ pobudzenia powtarzane w tych pętlach mogą utrzymywać stan aktywacji przez wymagany czas kilku, kilkudziesięciu sekund. Czas być może niezbędny, ażeby przenieść informacje do pamięci trwałej. Liczba połączeń zwrotnych wydaje się jednak zbyt mała dla odzwierciedlenia i przywołania całego bogactwa postrzeganego środowiska. Natomiast sprzężenia międzysynaptyczne FSS mogłyby ten problem rozwiązać definitywnie, gdyby udało się wykazać jego występowanie w sieci neuronowej.

Nowe badania wskazują, że możliwe są inne tory transmisji asocjacji poprzecznych, horyzontalnych, zarówno długozasięgowo, pomiędzy kolumnami, polami mózgowymi w korze i różnymi regionami mózgu, jak i krótkozasięgowo, pomiędzy synapsami i sąsiednimi neuronami w mikrokolumnach. Przedstawiony model można znacznie ulepszyć.

Propozycja taka zostanie przedstawiona w następnej części pracy.

## BIBLIOGRAFIA

- ABBOTT Larry F. (2008), Theoretical Neuroscience Rising, „Neuron” Vol. 60, Issue 3, 489-495, 6 November.
- ANASTASSIOU Costas A., MONTGOMERY Sean M., BARAHONA Mauricio, BUZSÁKI György & KOCH Christof (2010), The Effect of Spatially Inhomogeneous Extracellular Electric Fields on Neurons, „The Journal of Neuroscience” 30, 1925-1936.
- ANASTASSIOU Costas A., PERIN Rodrigo, MARKRAM Henry, KOCH Christof (2011), Ephaptic Coupling of Cortical Neurons, „Nature Neuroscience” Vol. 14, No. 2, Febr.
- BAARS Bernard J. (1997), In the Theater of Consciousness, New York: Oxford University Press.
- BAARS Bernard J. (1988), A Cognitive Theory of Consciousness, Cambridge: Cambridge University Press.
- BERNAT James L. (2009), Chronic Consciousness Disorders, „Annual Review of Medicine”, Vol. 60, 381-392.
- BINZEGGER Tom, DOUGLAS Rodney J., MARTIN Kevan A.C. (2004), A Quantitative Map of the Circuit of Cat Primary Visual Cortex, „The Journal of Neuroscience” 24, 8441-8453.
- BISHOP Christopher M. (1996), Neural Networks for Pattern Recognition, New York: Oxford University Press.
- BRODZIAK Andrzej (2001), Neurophysiology of the Mental Image, „Medical Science Monitor” 7(3), 534-538.
- CHALMERS David J. (1995), Facing Up to the Problem of Consciousness, „Journal of Consciousness Studies” 2, 200-219.
- CHALMERS David J. (1996), The Conscious Mind: In Search of a Fundamental Theory, Oxford: Oxford University Press.
- CHALMERS David J. (1998), On the Search for the Neural Correlate of Consciousness, [w:] S. HAMEROFF, A. KASZNAK, and A. SCOTT (eds.), Toward a Science of Consciousness II: The Second Tucson Discussions and Debates, Cambridge: MIT Press, 219-230.
- COGGAN Jay S., BARTOL Thomas M., ESQUENAZI Eduardo, STILES Joel R., LAMONT Stephan, MARTONE Maryann E., BERG Darwin K., ELLISMAN Mark H., SEJNOWSKI Terrence J. (2005), Evidence for Ectopic Neurotransmission at a Neuronal Synapse, „Science” 309, 446-451.
- CRICK Francis (1997), Zdumiewająca hipoteza, tłum. B. Chacińska-Abramowicz, M. Abrahamowicz, Warszawa: Prószyński i S-ka.
- DEECKE Lüder (2012), There Are Conscious and Unconscious Agendas in the Brain and Both Are Important-Our Will Can Be Conscious as Well as Unconscious, „Brain Sciences” 2, 405-420.
- DEHAENE Stanislas & NACCACHE Lionel (2001), Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework, „Cognition” 79, 1-37.
- DEMING Ross, SCHINDLER John, PERLOVSKY Leonid I. (2009), Multi-Target/Multi-Sensor Tracking Using Only Range and Doppler Measurements, „IEEE Transactions on Aerospace and Electronic Systems” Vol. 45, No. 2, April.
- DENNETT Daniel C. (1988), Quining Qualia, [w:] A. MARCEL and E. BISIACH (eds.), Consciousness in Modern Science, Oxford: Oxford University Press, 121-158.
- DUCH Włodzisław (2010), Architektury kognitywne. Neurocybernetyka teoretyczna, ed. R. Ta-deusiewicz, rozdz. 14, s. 329-361.
- EDELMAN Gerald M. (1999), Przenikliwe powietrze, jasny ogień. O materii umysłu, tłum. J. Rączaszek, Warszawa: PIW 1999.
- FAVOROV Oleg V., KURSUN Olcay (2011), Neocortical layer 4 as a pluripotent function linearizer, „Journal of Neurophysiology” Vol. 105, No. 3 (March), 1342-1360.

- FELDMEYER Dirk, EGGER Veronica, LÜBKE Joachim, SAKMANN Bert (1999), Reliable Synaptic Connections between Pairs of Excitatory Layer 4 Neurones within a Single 'Barrel' of Developing Rat Somatosensory Cortex, „The Journal of Physiology” 521, 169-190.
- FELL Jürgen, KLAVER Peter, ELGER Christian E., FERNANDEZ Guillén (2002), The Interaction of Rhinal Cortex and Hippocampus in Human Declarative Memory Formation, „Review of Neuroscience” 13, 299-312.
- FELL Jürgen, FERNANDEZ Gillén, KLAVER Peter, ELGER Christian E., FRIES Pascal (2003), Is Synchronized Neuronal Gamma Activity Relevant for Selective Attention?, „Brain Research. Brain Research Reviews” 42, 265-272.
- FRIES Pascal, REYNOLDS John H., RORIE Alan E., DESIMONE Robert (2001), Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention, „Science” 291, 1560-1563.
- GROSSBERG Stephen (1982), *Studies of Mind and Brain*, Dordrecht, Holland: D. Reidel.
- GROSSBERG Stephen (2000), *Linking Mind to Brain: The Mathematics of Biological Intelligence*, CAS/CNS Technical Report Series, No 016.
- GROSBERG Stephen (2000), The Complementary Brain: Unifying Brain Dynamics and Modularity, „Trends in Cognitive Science” 4, 233-246.
- GRUBER Thomas, TSIVILIS Dimitris, MONTALDI Daniela, MULLER Matthias M. (2004), Induced Gamma Band Responses: An Early Marker of Memory Encoding and Retrieval, „Neuroreport” 15, 1837-1841.
- HAIKONEN PENTTI O. (2012), *Consciousness and Robot Sentience*, Singapore: World Scientific Publishing Co.
- HICKOK Gregory & POEPEL David (2007), The Cortical Organization of Speech Processing, „Nature Reviews Neuroscience” 8, 393-402.
- HORTON Jonathan C., ADAMS Daniel L. (2005), The Cortical Column: A Structure without a Function, „Philosophical Transactions of the Royal Society of London. Series B: Biological sciences” 360 (1456), 837-62.
- HURLEY Susan (1998), *Consciousness in Action*, Cambridge, MA: Harvard University Press.
- INGLE David (1982), *Analysis of Visual Behavior*, Cambridge, MA: MIT Press.
- JONES Edward G. (2000), Microcolumns in the Cerebral Cortex, „Proceedings of the National Academy of Sciences” Vol. 97, No. 10, 5019-5021.
- LAIRD Johnson (1993), *The Computer and the Mind*, 2<sup>nd</sup> ed., Fontana Press, UK.
- LISMAN John E., IDIART Marco A. (1995), Storage of 7 +/- 2 Short-term Memories in Oscillatory Subcycles, „Science” 267, 1512-1515.
- LOWE David (2003), Radial Basis Function Networks, [w:] *The Handbook of Brain Theory and Neural Networks*, 2<sup>nd</sup> ed., ed. by Arbib MA, Cambridge, MA: MIT Press.
- MAKINO Hiroshi, MALINOW Roberto (2009), AMPA Receptor Incorporation into Synapses during LTP: The Role of Lateral Movement and Exocytosis, „Neuron” 64, 381-390.
- VON DER MALSBURG Christoph (1981), *The Correlation Theory of Brain Function*, – Originally published July 1981 as Internal Report 81-2, Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, 3400 Göttingen, Germany.
- MARKRAM Henry, PERIN Rodrigo (2011), Innate Neural Assemblies for Lego memory, „Frontiers in Neural Circuits” Vol. 5, Art. 6.
- MILNER David A., GOODALE Melvyn A. (2008), *Mózg wzrokowy w działaniu*, Warszawa: Wydawnictwo Naukowe PWN. Seria: Biblioteka Psychologii Współczesnej.
- MILSTEIN Joshua N., KOCH Christof (2008), Dynamic Moment Analysis of the Extracellular Electric Field of a Biologically Realistic Spiking, „Neuron” Vol. 20, No. 8, 2070-2084.
- MINSKY Marvin (1980), K-Lines: A Theory of Memory, „Cognitive Science” 4, 117-133.
- MINSKY Marvin (1993), Book Review: Allen Newell, *Unified Theories of Cognition. Artificial Intelligence in Perspective*, Cambridge, MA: MIT Press, 59, 343-354.

- NOË Alva (2005), *Action in Perception*, Cambridge, MA: MIT Press.
- NOË Alva (2009), *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*, New York: Hill & Wang.
- PERLOVSKY Leonid I. (1997), *Multiple Sensor Fusion and Neural Network*. DARPA Neural Network Study, Lexington, MA: MIT Lincoln Laboratory.
- PERLOVSKY Leonid I. (1998), *Conundrum of Combinatorial Complexity*, „IEEE Transactions on Pattern Analysis and Machine Intelligence” Vol. 20, No. 6, June 1998.
- PERLOVSKY Leonid I. (2001), *Neural Networks and Intellect: Using Model Based Concepts*, New York: Oxford University Press.
- PERLOVSKY Leonid I. (2002), *Aesthetic and Mathematical Theories of Intellect*, Moscow: Iskusstvoznanie 2(02).
- PERLOVSKY Leonid I. (2005), *Music – The First Principle. Role of Music in Evolution of Consciousness*, „Musical Theater”, E-journal, [http://www.ceo.spb.ru/libretto/kon\\_lan/ogl.shtml](http://www.ceo.spb.ru/libretto/kon_lan/ogl.shtml) UTH
- PERLOVSKY Leonid I. (2006a), *Toward Physics of the Mind: Concepts, Emotions, Consciousness, and Symbols*, „Physics of Life Reviews” 3(1), 22-55.
- PERLOVSKY Leonid I. (2006b), *The Knowledge Instinct*, New York: Basic Books.
- PERLOVSKY Leonid I. (2007a), *Cognitive High Level Information Fusion*, „Information Sciences” 177, 2099-2118.
- PERLOVSKY Leonid I. (2007b), *Consciousness and Music*. St. Petersburg University for Philosophy and Religion, Invited Lecture, St. Petersburg, Russia.
- PERLOVSKY Leonid I. (2008a), *Religion from Purely Scientific Point of View*. Invited Lecture at Science and Religion Seminar Workshop, University of Arizona, Tempe, AZ.
- PERLOVSKY Leonid I. (2008b), *Music and Consciousness*, „Leonardo. Journal of Arts, Sciences and Technology” 41(4), 420-421.
- PERLOVSKY Leonid I. (2008c), *Sapience, Consciousness, and the Knowledge Instinct (Prolegomena to a Physical Theory)*, [w:] R.V. MAYORGA, L.I. PERLOVSKY, *Toward Artificial Sapience. Principles and Methods for Wise Systems*, London: Springer Verlag, 33-60.
- PERLOVSKY Leonid I. (2009a), *Cognitive Computational Intelligence: Data Mining, Financial Prediction, Tracking, Fusion, Language, Cognition, and Cultural Evolution*. Tutorial, the IASTED International Conference on Computational Intelligence. 17-19, Honolulu, HI. Aug.
- PERLOVSKY Leonid I. (2009b), *Intersections of Mathematical, Cognitive, and Aesthetic Theories of Mind*, „Psychology of Aesthetics, Creativity, and Art” Vol. 4(1), Feb. 2010, 11-17.
- PERLOVSKY Leonid I. (2010a), *Mathematical Equivalence of Evolution and Design*, „Journal of Systemics, Cybernetics and Informatics” Vol. 8, No. 6.
- PERLOVSKY Leonid I., ILIN Roman (2010), *Neurally and Mathematically Motivated Architecture for Language and Thought*, „The Open Neuroimaging Journal” 4, 70-80.
- PERLOVSKY Leonid I., DEMING Ross W. (2010), *Maximum Likelihood Joint Tracking and Association in a Strong Clutter*, „IEEE Transactions on Magnetics”.
- PERLOVSKY Leonid I. (2010b), *Musical emotions: Functions, origin, evolution*, „Physics of Life Reviews” 7(1), 2-27.
- PERLOVSKY Leonid I., CABANAC Arnaud, BONNIOT-CABANAC Marie-Claude, CABANAC Michel (2012), *Mozart Effect, Cognitive Dissonance and the Pleasure of Music*, „Quantitative Biology; Neurons and Cognition”, in press, deliv.: 15 Sept.
- PETERSEN Carl C.H., SAKMANN Bert (2000), *The Excitatory Neuronal Network of Layer 4 Barrel Cortex*, „Journal of Neuroscience” 20, 7579-7586.
- RAMACHANDRAN Vilayanur S., and HUBBARD Edward M. (2001), *Synaesthesia – a Window into Perception, Thought and Language*, „Journal of Consciousness Studies” 8, 3-34.
- TAYLOR John G. (1992), *Towards a Neural Network Model of Mind*, „Neural Network World” 2, 797-812.

- TAYLOR John G. (1998), Constructing the Relational Mind, „Psyche” 4(10), June 1998.
- TAYLOR John G. (1998), The Race for Consciousness, Cambridge, MA: MIT Press.
- TYE Kay M., STUBER Garret D., RIDDER Bram de, BONCI Antonello, JANAK Patricia H. (2008), Rapid Strengthening of Thalamo-amygdala Synapses Mediates Cue-reward Learning, „Nature” 453, 1253-1257.
- ULLMAN Shimon (1996), High Level Vision: Object Recognition and Visual Cognition, Cambridge, MA: MIT Press.
- THOMSON Alex M., WEST David C., WANG Yun, BANNISTER A. Peter (2002), Synaptic Connections and Small Circuits Involving Excitatory and Inhibitory Neurons in Layers 2–5 of Adult Rat and Cat Neocortex: Triple Intracellular Recordings and Biocytin Labelling in Vitro, „Cerebral Cortex” 12, 936-953.
- VADAKKAN Kunjumon I. (2010), Framework of Consciousness from Semblance of Activity at Functionally LINKed Postsynaptic Membranes, „Frontiers in Psychology (Consciousness Research)” 10.
- VADAKKAN Kunjumon I. (2011), Processing Semblances Induced through Inter-Postsynaptic Functional LINKs, Presumed Biological Parallels of K-lines Proposed for Building Artificial Intelligence, „Frontiers in Neuroengineering” 4(8).
- VADAKKAN Kunjumon I. (2012), The Nature of “Internal Sensations” of Higher Brain Functions May Be Derived from the Design Rules for Artificial Machines That Can Produce Them, „Journal of Biological Engineering” 6-21.
- VELMANS Max (2009), How to Define Consciousness – and How Not to Define Consciousness, „Journal of Consciousness Studies” 16(5), 139-156.
- ZWAAN Rolf A. & RADVANSKY Gabriel A. (1998), Situation models in language comprehension and memory, „Psychological Bulletin” 123(2), 162-185.

## ARCHITEKTURA ŚWIADOMOŚCI

### CZĘŚĆ I: LOGIKA I MORFOLOGIA SIECI NEURONOWEJ

#### Streszczenie

Celem pracy jest wskazanie na podstawowe procesy neurologiczne i biofizyczne konstytuujące kompleksowy model umysłu świadomego. Przedstawiono zarówno model obliczeniowy działania neuronowych pól tworzących mentalne reprezentacje rzeczywistości, jak i biologiczne koncepcje realizacji tych funkcji mózgu. Wyjaśniono motywacje ludzkiego i zwierzęcego działania. Przedstawiono koncepcje, jak powstaje świadomość oraz jak rozpoznajemy, że jesteśmy świadomi, co objaśnia także problem samoświadomości. Wskazano, że zrozumienie fenomenu świadomości usuwa antynomię teleologicznego i przyczynowego charakteru natury ludzkiej, co niweluje dualizm cielesnego i duchowego charakteru substancji tworzącej nasze umysły. W części pierwszej przedstawiono ograniczenia koncepcji obliczeniowego modelu rozpoznawania poprzez porównywanie i poszukiwanie podobieństw wzorców według Perlovsky’ego oraz wskazano na sposób rozwinięcia tej koncepcji poprzez uwzględnienie morfologii i biologicznych funkcji sieci neuronowej mózgow naturalnych. Pokazano, że neuronowe pola synaptyczne o strukturze hierarchicznej impresjonów w rozumieniu hipotezy Vadakkana mogą spełniać rolę neuronowych pól modelujących zaangażowanych w główne procesy psychiczne. Zauważono, że rozszerzony w ten sposób poziom analizy funkcji mózgu w dalszym ciągu nie pozwala na wyjaśnienie fenomenu świadomości i konieczne jest przeanalizowanie podłoża biochemicznego i molekularnego tych procesów, co zostanie przedstawione w następnych częściach pracy.



ARCHITECTURE OF CONSCIOUSNESS  
PART ONE: LOGIC AND MORPHOLOGY OF NEURAL NETWORK

## Summary

The aim of this paper is to indicate basic neural and biophysical processes which constitute the phenomenon of conscious brain. The paper includes the computational model of neural modelling fields (NMF), which form mental representations of reality as well as biological concepts regarding the way these brain functions are implemented. The motivation of human and animal activity is also explained. The concepts regarding the emergence of consciousness are discussed, along with the perception of being conscious and the issue of self-awareness. It was demonstrated that the understanding of the phenomenon of consciousness eliminates the antinomy of a teleological and causative character of human nature, which, in turn, concludes the discussion on the duality of the physical or spiritual nature of the matter of which our minds are created. In the first part of this essay limitation of the computational model of recognition through comparison and detection patterns similarities according Perlovsky dynamic logic applied to the NMF is discussed. Further improvement of this concept was demonstrated taking under consideration morphology and biological functions of the neural network of natural brains. Neural synaptic fields can play the role of NMF and create hierarchical structure of semblions according to Vadakkan hypothesis responsible for major psychical functions. Even deeper level of brain functions analyse hereunder doesn't explain the consciousness phenomenon. For this goal biochemical and molecular background of these processes is necessary what will be described in following parts of this essay.

*Summarised by Wiesław Galus*

**Słowa kluczowe:** samoświadomość, impresjon, instynkt rozumienia, model umysłu, neuronowe pole modelujące, pola synaptyczne.

**Key words:** consciousness, semblion, perception instinct, model of the brain, neural modeling field, synaptic fields.

**Information about Author:** WIESŁAW GALUS, D.Sc.— Eutherm Sp. z o.o.; address for correspondence: ul. Zachodzącego Słońca 46, PL 01-495 Warszawa; e-mail: w.galus@eutherm.eu