Eliminatywizm i konstrukty psychologiczne

Włodzisław Duch
Katedra Informatyki Stosowanej i Laboratorium Neurokognitywne
Uniwersytetu Mikołaja Kopernika

1. Problem

Jednym z centralnych problemów badań kognitywistycznych jest relacja pomiędzy mentalnymi konstruktami, które pozwalają opisywać w intersubiektywny sposób na poziomie werbalnym obserwowane zjawiska mentalne i behawioralne, a obiektywnie mierzalnymi cechami procesów, które je tworzą. W efekcie mamy niekończące się spory o „świadomość", „myślenie", „zrozumienie" czy „inteligencję", zwłaszcza w kontekście rozwoju sztucznej inteligencji w ostatnich latach. Pojęcia fizyczne, takie jak „energia", „promieniowanie" czy „magnetyzm" używane są w pseudonaukowych teoriach w bezsensowny sposób. Argumenty posługujące się pojęciami, które nie są dobrze zdefiniowane, nie przyczyniają się do lepszego zrozumienia procesów umysłowych. Zauważył to już Alan Turing w słynnym artykule „Computing Machinery and Intelligence" (Turing, 1950). Nie można odpowiedzieć na pytanie „czy maszyna może być inteligentna" analizując sposób użycia słów "inteligencja" i "maszyny". Można to zrobić w obrębie określonej teorii, w której pojęcia będą jednoznacznie zdefiniowane, np. teorii symbolicznych systemów opartych na wiedzy w ramach sztucznej inteligencji (Newell, 1990).

Odwoływanie się do intuicji w przypadku wieloznacznych pojęć prowadzi do niekończących się dyskusji, ale nie pomaga znaleźć odpowiedzi. Było to jedną z motywacji empiryzmu logicznego jak i operacjonalizmu, sformułowanego w 1927 roku przez P.W. Bridgmana w książce „The Logic of Modern Physics". Takie podejście, podkreślające konieczność analizy procesów, zdarzeń, działań, a nie odwoływania się do statycznych pojęć czy abstrakcyjnych teorii, stało się popularne nie tylko w fizyce, ale przede wszystkim w psychologii. Nie można jednak całkowicie odrzucić pojęć teoretycznych i skupić się tylko nad operacyjnym ustalaniem warunków eksperymentu, gdyż trudno będzie zachować ciągłość pojęć, których używamy do opisania świata. Każdy nowy eksperyment wykonywany jest w nieco innym kontekście, ale próbujemy go opisać w terminach już nam znanych. Sam Bridgeman zdawał sobie z tego sprawę pisząc o złożoności świata zbyt wielkiej by dała się odwzorować za pomocą werbalizowalnych struktur. Nawet w fizyce pojęcie energii uległo reifikacji. „Nie da się w prosty sposób zwerbalizować wszystkich sytuacji, w których przejawiają się różne aspekty energii" (Bridgman, w: Frank 1956).

Powstaje więc pytanie, czy pojęcia psychologii potocznej i psychologii naukowej pozwolą nam opisać rzeczywistość i na jak dokładny opis możemy mieć nadzieję? Czy można zwerbalizować zachowania nieliniowych układów dynamicznych o dużym stopniu złożoności, które pozwalają coraz lepiej opisywać zachodzące w mózgu procesy? Czy te procesy można opisać w terminach jednoznacznych pojęć, a więc posługiwać się logiką klasyczną by stwierdzić, czy miały miejsce? Czy też należy wyeliminować wszystkie pojęcia psychologii potocznej? Czym je zastąpić i jak w sensowny sposób budować teorie potrzebne by się porozumieć?

Pewne pojęcia zniknęły z rozważań naukowych fizyków, chemików czy psychologów jeszcze przed XX wiekiem, ale nadal są używane w metaforycznym lub religijnym sensie.

Chociaż eter nie istnieje słuchamy radia „na falach eteru", ale to tylko niegroźna metafora. „Bioenergia" i „biopole" są znacznie bardziej szkodliwe, stwarzając pozorne wyjaśnienie dla działań bioenergoterapeutów. Dualistyczne intuicje, traktowanie mózgu/ciała i umysłu można dostrzec w artykułach na temat neuronauki nawet w najlepszych czasopismach specjalistycznych (Mudrik, i Maoz, 2015). Utrudnia to głębsze zrozumienie relacji pomiędzy światem psychicznym a fizycznym. Nie utożsamiamy się z całym organizmem, nie traktujemy wszystkich procesów w nim zachodzących jako części „siebie".

2. Eliminatywizm

Które z potocznie używanych pojęć dotyczących stanów mentalnych mają głębszy sens, a które należy wyeliminować? Dyskusje na temat eliminatywizmu mają długą tradycję, ale dotyczyły tylko niektórych, wybranych pojęć. Hume (podobnie jak Budda 2500 lat temu) uznał, że trwała jaźń nie istnieje. „Dusza" utraciła wszystkie funkcje, które jej przypisywano, więc stała się pojęciem pustym, pozbawionym desygnatu. Nie miało to wpływu na język potoczny. Dopiero w połowie XX wieku zaczęto jednak powątpiewać w sens większości pojęć potocznych. Sellars w artykule „Empiricism and the Philosophy of Mind" (1956) podkreślił, że nasze rozumienie umysłu nie jest oparte na bezpośrednim dostępie do mechanizmów jego działania tylko na kulturze, która dostarcza nam naiwnych pojęć i ram teoretycznych. Mogą to więc być pojęcia całkowicie błędne, chociaż pełnią istotną rolę w komunikacji między ludźmi. Feyerabend (1963) uznał niematerialną naturę zdroworozsądkowych pojęć mentalnych, a więc dowolna wersja fizykalizmu powinna pokazać, że stany mentalne, takie jak przekonania czy wrażenia, nie istnieją. Skoro wiemy, że stany fizjologiczne istnieją, po co odwoływać się do stanów mentalnych, wystarczą ich fizjologiczne korelaty – pisał Quine (1960). Wyjaśnienie stanów mentalnych w oparciu o pojęcia opisujące procesy fizjologiczne powinno być wystarczające by uznać, że złość czy fizyczny ból jest tożsama ze stanem mentalnym. Ciekawe jak by ocenił chroniczny ból psychosomatyczny osób, u których nie udaje się znaleźć żadnych fizjologicznych przyczyn. Muszą być za to odpowiedzialne procesy zachodzące w mózgu, ale nawet dzisiaj trudno je wykryć. Ten przykład pokazuje, że wewnętrzna interpretacja aktywności mózgu jest rzeczywistym procesem, mającym określone skutki behawioralne, a określenie „ból chroniczny" nie da się wyeliminować z języka opisu stanów organizmu, chociaż lekarze-specjaliści będą mu mogli nadać bardziej precyzyjny charakter.

Podobnie stało się z takim ogólnymi pojęciami jak uwaga czy pamięć. Okazało się, że warto rozróżnić różne formy pamięci, zależnie od procesów zachodzących w mózgu: czasu trwania – pamięć długotrwała, krótkotrwała, natychmiastowa, ultrakrótka, pętla fonologiczna, czy funkcji – pamięć operacyjna, robocza, retrospektywną lub prospektywna, rozpoznawcza, deklaratywna, epizodyczna, ikonograficzna, autobiograficzna, semantyczna, proceduralna. Pamięć może być jawna bądź utajona (nieświadamialna), dotyczyć odruchów warunkowych (gotowości reakcji, dyspozycyjności), habituacji-sensytyzacji (nieasocjacyjna) oraz torowania (priming). Może dotyczyć percepcji, języka, ruchu lub emocji. Każde zdarzenie inicjowane zewnętrznie lub zachodzące wewnętrznie w organizmie może zostać zapamiętane ale mechanizmy za tym stojące są bardzo różne. Można się więc spierać, że pojęcie „pamięć" do niczego się konkretnego nie odnosi, stanowi kategorię nadrzędną dla wielu bardzo różnych procesów. Podręczniki neuronauk jak i podręczniki medyczne stają się z roku na rok coraz grubsze, opisujemy procesy biologiczne coraz dokładniej, ale zamiast eliminować pojęcia raczej je uszczegóławiamy.

Nie wydaje się więc, by radykalny materialistyczny eliminatywizm miał szansę powodzenia. Pojęcia potoczne są punktem wyjścia dla neuronauk. Różne pojęcia rzeczywiście znikły z naukowych rozważań, w tym większość pojęć wprowadzonych przez Freuda, jednakże nie były to pojęcia psychologii potocznej, tylko nieudane próby wprowadzenia nowych konstruktów psychologicznych. Język neuronauk staje się coraz bardziej precyzyjny i coraz bardziej odległy od pojęć potoczne psychologii. Język układów dynamicznych, konektomiki, procesów w sieciach neuronowych (network neuroscience, Bassett i Sporns, 2017) pozwala coraz lepiej opisać zachodzące w mózgu procesy na poziomie, który pozwala na interpretacje na poziomie mentalnym. Pamięć utajona to aktywacja mózgu wpływająca na zachowanie, lecz zbyt słaba by ją świadomie dostrzec na poziomie mentalnym, o jej istnieniu można jedynie wnioskować obserwując jej wpływ na działanie danej osoby. Pamięć epizodyczna to reaktywacja podobnych obszarów mózgu, które były aktywne w czasie przypominanego epizodu. Uwaga to synchronizacja oscylacji różnych obszarów mózgu, pozwalająca na szybsze reakcje. Potrafimy tworzyć coraz lepsze interfejsy mózg-komputer pozwalające odczytać intencje działania, ale również wyobrażenia wzrokowe. Mamy coraz więcej dowodów na to, że o inteligencji jak i chorobach mózgu, takich jak autyzm czy ADHD, decyduje sprawna współpraca różnych grup neuronów ze sobą. Jednakże całkowita zmiana naszego sposobu patrzenia na świat i sprowadzenia pojęć mentalnych do obiektywnie mierzalnych procesów fizjologicznych jest nierealna. Zamiast „zwrócił uwagę" nie będziemy mówić, które obszary mózgu się uaktywniły, to jest zbyt szczegółowy poziom, nieprzydatny do komunikacji pomiędzy ludźmi, ale ważny dla ekspertów.

3. Neuronauki

Paul Churchland w artykule "Eliminative Materialism and the Propositional Attitudes" (1981) wyraził przekonanie, że postęp w badaniach mózgu wyeliminuje większość pojęć psychologii potocznej. Również Patricia Churchland w znanej książce "Neurophilosophy" (1986) argumentowała, że psychologia potoczna nie jest lepiej uzasadniona niż fizyka potoczna z czasów Arystotelesa, z której niewiele pozostało. Ludzie mylili się w stosunkowo prostych sprawach, dotyczących ruchu, ciepła, ognia, pogody, chorób, planet i gwiazd, dlaczego więc mieliby stworzyć dobrą teorię zjawisk mentalnych, które są znacznie bardziej złożone? W książce z 1995 roku „The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain" Paul Churchland opisał sposób kodowania percepcji jak i tworzenia się przekonań w sieciach neuronowych i przybliżony opis tego procesu za pomocą modeli wektorowych. Mózg nie uczy się formułek logicznych, nowe pojęcia nabierają stopniowo znaczenia, zrozumienie rośnie wraz ze wzrostem liczby skojarzeń i formowania się teorii, wzajemnie pobudzających się aktywacji sieci neuronowych. Takie procesy trudno uchwycić za pomocą formuł klasycznej logiki, a więc w sposób dający się wyrazić werbalnie. Są to procesy ciągłe (Spivey, 2007), wiec nie dają się opisać werbalnie, za pomocą skończonej liczby symboli. Utrwalanie się nowych pojęć w siatce już istniejących jest procesem stopniowym formowania się atraktorów neurodynamiki sieci, pojawienia się silnie zsynchronizowanych aktywnych grup neuronów, które dzięki mechanizmom „k zwycięzców bierze wszystko" (kWTA, k-Winner-Takes-All), lub innych mechanizmów hamowania zapewniających powstawanie unikalnych stanów (cf. O'Reilly i inn. 2012) wygaszają na pewien czas alternatywne procesy.

Dzięki temu chwilowy stan mózgu odróżnia się od tła słabiej pobudzonych obszarów, może zostać jednoznacznie zinterpretowany, można do niego dołączyć werbalną, symboliczną aktywację realizowaną przez obszary specjalizujące się w funkcjach językowych. Formuły logiczne nie mogą jednak opisać w poprawny sposób procesów skojarzeniowych. Logika rozmyta, w której można uwzględnić stopień prawdziwości stwierdzeń, tylko w niektórych sytuacjach pozwala na lepszy opis zachodzących procesów. Aproksymacja funkcji realizowa-

nych przez sieci neuronowe za pomocą wyrażeń logiki klasycznej lub rozmytej jest trudna i nie zawsze możliwa, zwłaszcza w przypadku sieci rekurencyjnych, których aktywność zmienia się w czasie (Duch, Setiono i Żurada, 2004). Alternatywą jest oparcie się na regułach wykorzystujących podobieństwo rozkładów aktywacji elementów sieci neuronowej, lub innych ocenach podobieństwa bodźców (Duch, 2005a). Shepard próbował w ten sposób znaleźć uniwersalne prawa percepcji (Shepard, 1987, 1994).

Doświadczenia wewnętrznego nie można więc w pełni opisać, metoda fenomenologiczna nie prowadzi do głębszego zrozumienia stanów mentalnych. Wiele przykładów, ilustrujących ten problem można znaleźć w książkach (Hurlbrut i Schwitzgabel, 2007; Schwitzgabel 2011). Słynne powiedzenie Wittgensteina „Granice mojego języka oznaczają granice mojego świata" jest tylko częściowo prawdziwe. Świat, który potrafimy opisać werbalnie jest zaledwie modelem rzeczywistości, obejmującej stany umysłu, których nie można w pełni opisać werbalnie, włączając w to emocje, uczucia, wrażenia, wyobrażenia, intuicyjne skojarzenia. Wspólny artykuł Churchland & Churchland (2002) nie ma już zdecydowanego, redukcjonistycznego charakteru zmierzającego do eliminacji potocznych pojęć. W mózgu powstaje model zjawisk zachodzących w świecie, indywidualny rozsądek tworzy model tych zjawisk. Nauka robi to samo tylko jako wysiłek zbiorowy bardziej szczegółowo, więc model naukowy pozwala na lepsze przewidywania. Stany mentalne opisywane przez potoczne pojęcia to realnie istniejące procesy, zachodzące w mózgu i w całym organizmie. W tym sensie są to dwie strony tego samego medalu, można je opisywać za pomocą różnych pojęć, jednakże potrzeba pragmatycznego używania języka do celów komunikacji nie pozwoli na eliminację pojęć potocznych, chociaż będą one mniej precyzyjne.

Reprezentacje wewnętrzne można opisać w przestrzeni aktywności grup neuronów (Duch, 2005, 2010). Mamy tu do czynienia nie z izomorfizmem stanów środowiska i stanów mózgu, ale z podobieństwem pomiędzy samymi stanami, tzn. to co podobne w świecie jest reprezentowane przez podobne aktywacje w mózgu. Churchlandowie piszą dość niejasno o potrzebie pewnej „reorientacji semantyki" w oparciu o relacje pomiędzy modelem mentalnym i rzeczywistością, w stronę zgodną z neuronaukami. Dlaczego powstały takie a nie inne modele? Muszą być bliższe rzeczywistości. To dość zaskakująca konkluzja, bo ich głównym zadaniem jest oczywiście zwiększenie szans na przeżycie, a nie wierniejsze odwzorowanie świata zewnętrznego. Perspektywa ewolucyjna nie pojawia się jednak w tym artykule. Paul Churchland w książce z 2007 roku poświęcił neurosemantyce cały rozdział a jako przykład omówił w dwóch rozdziałach percepcję koloru. W przypadku percepcji wzrokowej nie widać szans redukcji stanów mentalnych do neuronalnych, pomimo licznych iluzji wzrokowych i efektów związanych z postrzeganiem koloru. Widać za to jak zrozumienie tych zjawisk nie jest możliwe na poziomie psychologii potocznej.

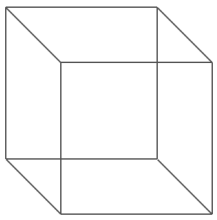
4. Fenomika neurokognitywna

Jeszcze większe trudności ma psychologia potoczna z bardziej złożonymi konstruktami, takimi jak osobowość i temperament. Teorie temperamentu z czasów Hipokratesa zostały powiązane z pobudliwością neuronów już za czasów Pawłowa, rozróżniono wiele typów osobowości, poznano mechanizmy regulacyjne związane z neurotransmiterami i genetyką (Strelau, 2008). Psycholodzy opracowali różne narzędzia psychometryczne próbując opisać „wymiary temperamentu" w oparciu o konstrukty psychologiczne, np. regulacyjna teoria temperamentu Strelaua używa takich cech jak „żwawość" i „perseweratywność", oraz cztery cechy związane z energetycznymi aspektami: aktywność, reaktywność, wrażliwość sensoryczna oraz wytrzymałość. Ten styl badań można nazwać zstępującym (top-down), od ogólnych pojęć psychologii potocznej do coraz bardziej szczegółowego opisu weryfikowanego statystycznie za pomocą kwestionariuszy i analiz behawioralnych. W jakim stopniu można w ten spo-

sób scharakteryzować działanie mózgu/umysłu? Alternatywą jest podejście wielopoziomowe, uwzgledniające kompletny fenotyp człowieka, od poziomu genetycznego, molekularnego, ścieżek sygnalizacji komórkowej, rodzaju i własności neuronów oraz ich sieci, do szybkich reakcji związanych z percepcją i aktywizacji względnie stabilnych cech całego systemu, opisywanych przez psychologię eksperymentalną, nauki behawioralne i raporty introspekcyjne.

Psychiatria przez długi okres próbowała opisać zespoły zaburzeń na poziomie obserwowalnych objawów, zebranych w podręcznikach klasyfikujących wszystkie psychopatologie (obecnie to DSM V z roku 2013 i ICD 10 z 1992 roku). Okazało się to jednak niewystarczające, albowiem „proponowane kategorie, oparte na objawach i zespołach, mogą nie pozwolić uchwycić fundamentalnych mechanizmów odpowiedzialnych za dysfunkcję" (Insel i inn. 2010). Dlatego obecne propozycje zmierzają do stworzenia zupełnie innego system klasyfikacji w oparciu o zaburzenia pracy 5 rozległych sieci neuronowych, zaangażowanych w mechanizmy regulujące homeostazę i pobudzenie (Arousal/Regulatory), mechanizmy poznawcze, nagrody (Positive Valence Systems), utraty (Negative Valence Systems), oraz interakcji społecznych (Social Processes Systems). Każda z tych sieci realizuje bardzo wiele funkcji, np. mechanizmy poznawcze obejmują zarówno uwagę, różne formy percepcji, różne rodzaje pamięci, posługiwania się językiem, kontroli działania (reakcji, wyboru celu, oceny jego skutków). Są to konstrukty znane z psychologii, ale na ich działanie w konkretnym organizmie wpływ mają procesy na poziomie genetycznym, molekularnym (neurotransmitery i neuromodulatory), komórkowym, mikroobwodów neuronalnych i ich interakcji, aktywacji podsieci strukturalnego konektomu, neuroplastyczności, procesów neurofizjologicznych związanych z oscylacjami EEG, oraz układu kardiowascularnego dostarczającego neuronom energię. Te procesy pozwalają na zdefiniowanie „jednostek analizy", wymiarów w których można dokonać pełnego opisu. Fenomika neuropsychiatryczna nie próbuje więc wyeliminować ale chce uszczegółowić konstrukty psychologiczne na wielu poziomach zbierając wszystkie informacje w macierzy RDoC (Research Domain Criteria) łączącej konstrukty z jednostką analizy. Projekt rozwija się od 2010 roku, ale daleko jest jeszcze do jego końca. Psychologia i nauki o uczeniu się (learning sciences) nie sformułowały jeszcze swojego planu rozwoju fenomiki neurokognitywnej, zadawalając się częściowymi opisami fenotypów (Duch, 2013).

Czy konstrukty psychologiczne wystarczą by sformułować predyktywne modele zachowania? Jaki język, aparat pojęciowy pomoże nam najbardziej by powiązać te konstrukty z procesami w mózgu decydującymi o zachowaniu? To zależy od perspektywy czasowej: reakcja neuronów na pobudzenie może być bardzo szybka, np. w przypadku iluzji wzrokowych nie zauważamy czasu przejścia pomiędzy dwoma sposobami interpretacji tego samego obrazu. Sześcian Kanizsy raz widzimy bliższą stroną w lewej górnej części a raz w prawej dolnej. Zmiana synchronizacji neuronów odpowiedzialnych za interpretację jest bardzo szybka i jej nie zauważamy. Zmiany rozwojowe zachodzą w różnym tempie od narodzin do śmierci. Uczenie się możliwe jest dzięki neuroplastyczności i do pewnego stopnia neurogenezie. Jednakże w stosunkowo krótkim odcinku czasu mamy do czynienia z siecią neuronów, w której zachodzą szybkie procesy w czasie milisekund, przy względnie stabilnych połączeniach strukturalnych opisywanych przez konektom. Dlatego badanie sieci, ich struktury i neurodynamiki, rozchodzenia się aktywacji w tych sieciach, ma obecnie pierwszoplanowe znaczenie. Powiązanie procesów zachodzących w sieciach z konstruktami psychologicznymi jest jeszcze dalekie od doskonałości, ale elektroencefalografia i metody neuroobrazowania oraz neuronauk obliczeniowe coraz lepiej ukazują procesy związane z uwagą, uczeniem się czy pamięcią. Konstrukty psychologiczne powinny odzwierciedlać mechanizmy ich implementacji przez mózgi, inaczej teorie na nich oparte będą ograniczone.



Alternatywą dla podejścia zstępującego, startującego od konstruktów psychologicznych, jest próba aproksymacji procesów fizycznych zachodzących w sieciach neuronowych, które można próbować skategoryzować i opisać w symboliczny sposób. Świadome skupienie uwaga związanej z percepcją można powiązać z pobudzaniem kory zmysłowej przez płaty czołowe. Percepcja wymaga powstania stanu atraktorowego angażującego neurony kory zmysłowej i obszarów skojarzeniowych, nadających znaczenie perceptom. Liczba teoretycznie możliwych stanów, które mogą powstać w mózgu jest astronomicznie duża, ale stany realnie powstające ograniczone są przez strukturalne własności sieci neuronów oraz przez funkcjonalnie dostępne kwazi-stacjonarne rozkłady pobudzeń. Konstrukty takie jak „wyczerpanie ego" (ego depletion, Baumeister i inn., 1998) wynikają z desynchronizacji neuronów związanej z ich zmęczeniem. Intensywne procesy myślowe wymagają czasem oderwania się od tematu, co pozwala zaangażowanym w nie neuronom odpocząć. Jestem zmęczony trudno mi się skupić, czyli nie synchronizują mi się neurony w odpowiednich obszarach mózgu. Jeśli cierpię z powodu ADHD to neurony zbyt szybko się desynchronizują, a to zależy od ich budowy. Możemy w ten sposób przejść na poziom molekularny, jednocześnie zastępując psychologiczne konstrukty pojęciami pozwalającymi na opis układów dynamicznych, których wartości dają się obiektywnie zmierzyć, a modele na nich oparte pozwolą na bardziej szczegółowe przewidywania.

Chociaż wstępujące podejście ma swoje zalety trudno się spodziewać, by taki język całkowicie wyeliminował potoczne określenia, chociaż częściowa eliminacja w specjalistycznej literaturze jest możliwa. Synergetyka (Haaken, 1983) to teoria samoorganizacji i emergencji pokazująca jak w układach złożonych, takich ja mózgi, pojawia się na poziomie makroskopowym uporządkowana struktura, którą można opisywać w przestrzeni o znacznie mniejszej liczbie parametrów, np. utożsamianych z wymiarami charakteryzującymi temperament. Teoria synergetyki znalazła zastosowanie zarówno do opisu zachowań sensomotorycznych (Jantzen i Kelso, 2007) jak i koordynacji społecznych interakcji pomiędzy ludźmi (Coey i inn, 2016). Na poziomie mikroskopowym mamy do czynienia z neuronami, których aktywacja zależy od parametrów związanych z wewnętrzną dynamiką i zewnętrznym otoczeniem. Zmiana tych parametrów powoduje pojawienie się nowych wzorców makroskopowej aktywności, które widoczne są za pomocą metod neuroobrazowania. W tego typu układach nie mamy do czynienia z liniową przyczynowością – mózg nie tworzy umysłu – ale z przyczynowością kołową, bo przejście pomiędzy stanami makroskopowymi można scharakteryzować za pomocą parametrów porządku w niskowymiarowych przestrzeniach, określających zmianę na poziomie mikroskopowym. W tej teorii procesy mentalne i fizyczne wzajemnie się warunkują, stany mentalne zmieniają stany fizyczne mózgu a te kolei wpływają na stany mentalne. Co więcej, ostatnie modele matematyczne pokazują, że zależności przyczynowe na makropoziomie mają większą moc przewidywania niż na poziomie mikroskopowym (Hoel, 2017). Redukcjonizm i eliminatywizm nie jest więc dobrym rozwiązaniem.

5. Mózg i ja

Nasz sposób myślenia o umyśle i mózgu przesiąknięty jest dualizmem wynikającym z potocznej psychologii i jak pokazała Mudrik i Maoz (2014) widać to nawet w publikacjach znakomitych neuronaukowców, takich jak Baars, Damasio, Frith, Gazzaniga, Koch, LeDoux, Rizzolatti, Sternberg i wielu innych. Liczne przykłady cytowane w ich pracy to wyrażenia „mózg zna nasze decyzje zanim jeszcze sami je poznamy", „mózg nas oszukuje", lub „mózg mnie do tego zmusił". Tego typu wyrażenia przeciwstawiające „ja" i mózg, który je tworzy, można oczywiście wyeliminować, pisząc „procesy zachodzące w mózgu spowodowały takie działanie, które nie było świadome". Próba obiektywnego opisu prowadzi jednak do pomijania roli „ja", intencjonalnych procesów pozwalających w świadomy sposób kontrolować za-

chowanie organizmu. Dualizm ma konsekwencje moralne, gdyż powinniśmy się czuć odpowiedzialni za wszystko, co robimy, nawet jeśli niektóre z procesów kierujące naszym działaniem nie są świadome. Wobec uwag napisanych powyżej pewien dualizm wynikający z emergencji nie jest pozbawiony sensu. Świadome „ja", z którym się utożsamiamy, wie tylko o nielicznych procesach zachodzących w mózgu.

Definiowanie dobrych konstruktów psychologicznych jest jednak bardzo trudnym zadaniem. Mózgi ludzkie różnią się znacznie, kontekst prowadzonych eksperymentów ma czasem duże znaczenie i nie wszystkie zmienne da się kontrolować, a sam eksperyment zmienia osobę badaną, więc stabilności powtarzalność rezultatów jest trudno osiągnąć. Z tego powodu Smedslund (2016) doszedł do wniosku, że psychologia nie może być w pełni nauką empiryczną. Używając ogólnych konstruktów psychologicznych zakładamy, że stosują się one do każdego badanego, a nawet do niektórych gatunków zwierząt. Tworząc teorie musimy pracować na wysokim poziomie abstrakcji, nie możemy używać odrębnych pojęć dla każdego badanego. Trudno jest jednak znaleźć dostatecznie homogeniczną grupę osób, zwłaszcza gdy pracuje się z dziećmi czy osobami z różnymi zaburzeniami. Strach czy ból ośmiornicy jest do pewnego stopnia funkcjonalnie podobny do naszych doświadczeń, ale też znacznie odmienny. Jak zauważył Varela (1996) fenomenologiczna struktura doświadczenia ludzkiego i fizykalne ujęcie procesów poznawczych na poziomie neuronalnym wzajemnie się warunkują. Neurofenomenologia może być tu dobrą metodologią poszukiwań lepszych konstruktów psychologicznych.

6. Kategoryzacja i psychologiczne teorie

Ograniczenia werbalnych modeli zachowania opartych na konstruktach psychologicznych widać nawet w stosunkowo prostych przypadkach uczenia się i kategoryzacji. Psycholodzy badając powstawanie stereotypów stworzyli „uwagową teorię uczenia się kategorii" (Sherman i inn. 2009). Badania dotyczą iluzorycznych korelacji i zaskakujących efektów, takich jak kategoryzacja wbrew częstościom bazowym (inverse base-rate effects, IBRE). W wielu publikacjach przedstawiono empiryczne wyniki uczenia się kategorii w różnych warunkach. W najprostszym przypadku mamy do czynienia z listą objawów i chorób. Oznaczmy przez C (częstą) i R (rzadką) pojawiające się choroby, oraz przez PC (zawsze dla C), PR (zawsze dla R), oraz I (nieistotne, przypadkowe) listę objawów. Np. na liście mamy:

Gorączka, Wysypka: Grypa; czyli $(PC, I) \rightarrow C$

Ból gardła, Katar: Angina; czyli $(PR, I) \rightarrow R$

Niech C występuje 3 razy częściej niż R, w kontekście różnych nieistotnych objawów I, które mogą wystąpić zarówno dla PC jak i PR. Po pokazaniu dłuższej listy osoby badane uczą się kategoryzować zespoły objawów, przypisując do nich choroby. Następnie pytane są o diagnozę dla kombinacji symptomów. Dla pojedynczych objawów odpowiedzi są jednoznaczne, $PC \rightarrow C$; $I \rightarrow C$, dla kombinacji $(PC, I) \rightarrow C$. Jest to zgodnie z oczekiwaniami opartymi na częstościach występowania chorób i ich objawów. Również dla kombinacji (PC, I, PR) 60% odpowiedzi to choroba częsta C. Jednakże jeśli zabraknie nieistotnego objawu, czyli zapytamy o przewidywania dla (PC, PR) 60% odpowiedzi to choroba rzadka. To zaskakujący wynik, wbrew częstości bazowych. Badania tego typu są ważne dla zrozumienia błędów poznawczych. Jak wyjaśniają to psycholodzy? Kruschke (1996) interpretował wyniki tego i podobnych eksperymentów odwołując się do uczenia (częstość decyduje) i przenoszenia uwagi na cechy rozróżniające. Faktem jest, że kategoryzacji uczy się sieć neuronowa w mózgu a proces uczenia się i odpowiedzi jest zależny od nieliniowej neurodynamiki, która opisuje zachowanie się neuronów w sieci. Trzeba więc zbadać jak w modelu kategoryzacji tworzą się ślady pamięci (atraktory neurodynamiki) i jak w zależności od warunków początkowych (py-

tań kontrolnych) taka sieć zareaguje, czyli jak będzie wyglądać trajektoria obrazująca zmiany stanu sieci (Dobosz i Duch, 2010; Duch i Dobosz, 2011). Symulacje takiego modelu neuronowego (Duch i Dobosz, w przygotowaniu) można opisać werbalnie, jednak próba opisu zachowania takiego systemu na poziomie psychologicznych konstruktów jest ryzykowna (w pracy Duch, 1996 użyty został prosty układ dynamiczny, a nie model neuronowy). Odpowiedź „przychodzi nam do głowy" bo powstają odpowiednie skojarzenia w naszym mózgu. Próbujemy racjonalizować przyczyny takich skojarzeń, ale nie mamy dobrego wglądu w naszą maszynerię poznawczą.

Mamy 5 parametrów: {C, R, I, PC, PR}, pierwsze 3 na wejściu i dwa wyjściowe. Pomiędzy wejściem i wyjściem mamy większą grupę neuronów, których aktywacja może przyjmować różne konfiguracje. Uczenie się listy powoduje powstawanie stabilnych rozkładów prawdopodobieństwa aktywacji neuronów sieci. Ponieważ objawy I są różne w przestrzeni aktywacji neuronów powstają obszary (różne konfiguracje), w których stany sieci interpretowane są jako PC lub PR. Są to baseny atrakcji neurodynamiki, w zależności od stanu początkowego aktywności sieci jej dynamika zmienia się tak, że powstają specyficzne konfiguracje interpretowane na wyjściu sieci jako PC lub PR.

Możemy w tym przypadku porównać interpretację psychologiczną z neurodynamiką. Co więcej, mamy prace eksperymentalne z użyciem EEG (potencjały wywołane, Wills i inn. 2014) jak i prosty model koneksjonistyczny przesuwania uwagi EXIT (Kutzner i Fiedler, 2015), podważające interpretację uwagowej teorii uczenia się kategorii. Porównajmy obydwie interpretacje, psychologiczna za pracami (Kruschke, 1996; Sherman i inn., 2009).

Przypadek	Neurodynamika	Psychologia
Uczenie: (PC, I) $\rightarrow$ C	Częste powtarzanie PC w kontekście I tworzy rozległy basen atrakcji do konfiguracji interpretowanych jako C.	Objawy PC, I są typowe bo pojawiają się często.
Uczenie: (PR, I) $\rightarrow$ R	Basen atrakcji dla R to mniejszy obszar, konfiguracje do niego należące odróżniają się od C nadając PR większą wagę.	Uwaga przesuwa się na objaw PR bo I jest niejednoznaczny.
Pytania: I $\rightarrow$?	Trajektorie neurodynamiki częściej prowadzą do większego basenu atrakcji C.	Zgodnie z częstościami bazowymi C będzie częściej przypominane.
Pytania: PC+PR+I $\rightarrow$?	Trajektorie neurodynamiki wpadają częściej do basenu C bo I częściej występowało z PC.	Zgodnie z częstościami bazowymi C będzie częściej przypominane.
Pytania: PC+PR $\rightarrow$?	Waga PR jest większa, brak I powoduje częstsze wpadanie trajektorii do basenu R.	Uwaga skupia się nad objawem PR, bo jest bardziej jednoznaczny.

W ostatnim przypadku wyjaśnienie oparte na skupieniu uwagi nad PR można podważyć zauważając, że PC jest też objawem jednoznacznym, dyskryminującym pomiędzy C i R. Uwagowa teoria uczenia się kategorii zakłada, że cechy dla rzadszej kategorii muszą mieć większą wagę, są dystynktywne. Tak wynika również z modelu neuronowego i to wystarcza by kom-

binacja PC+PR wprowadzała układ w basen atrakcji dla R. Mechanizm przesuwania uwagi nie jest tu potrzebny i to pokazały nowsze badania eksperymentalne. Niestety ta tendencja powoduje powstawanie stereotypów dotyczących mniejszości, rzadziej pojawiających się przypadków, odróżniania „swoich" i „obcych". Na szczęście nie jesteśmy prostą siecią neuronową i nasze spontaniczne skojarzenia mogą być zmodyfikowane przez dalsze skojarzenia wynikające z głębszych przemyśleń. Psycholodzy opracowali Test Utajonych Skojarzeń (Implicit Association Test), który ma pomóc w badaniu nieświadomych reakcji i porównać ich wynik z końcowymi decyzjami.

7. Mózg i umysł

Mamy z jednej strony fizyczne procesy zachodzące w mózgu, które coraz lepiej udaje się nam zrozumieć i powiązać z percepcją, pamięcią, wyobrażeniami, inteligencją czy zaburzeniami psychicznymi. Odczytywanie intencji stanowi podstawę interfejsów mózg-komputer. Ilość informacji w sygnałach fMRI czy EEG jest ogromna, a na jej podstawie określamy jedynie proste kategorie. Mamy więc mapowanie dynamicznych stanów mózgu na konstrukty psychologiczne. Możemy dokonać wizualizacji zmian stanów mózgu w przestrzeni aktywności zlokalizowanych w wybranych obszarach (Duch i Dobosz, 2011). Z drugiej strony chcielibyśmy z nim powiązać opis przestrzeni fenomenologicznych zdarzeń na poziomie mentalnym, a więc określić transformację pomiędzy obiema przestrzeniami i zachodzącymi w nich procesami. Wówczas moglibyśmy wybrać język, w którym chcemy opisywać stany poznawcze. Taki geometryczny model jest jednak dość odległym celem (Duch, 1997; 2012).

W wielu sytuacjach interpretacje psychologiczne i modele oparte na dopasowywaniu parametrów nie będą w stanie uchwycić całej złożoności procesów zachodzących w złożonych sieciach neuronowych. Neuronauki na pewno pogłębią i uszczegółowią pojęcia psychologiczne. Jednakże nawet jeśli uda się nam stworzyć poprawny model obliczeniowy całego mózgu i będziemy w stanie przewidzieć jego zachowanie nie wyeliminuje to z języka potocznego większości stosowanych obecnie pojęć. Jest wiele przykładów pokazujących problemy podejścia wstępującego (bottom-up), od szczegółowych fizycznych symulacji do próby pojęciowego opisu zjawisk. Fizyka i chemia kwantowa pozwala na obliczanie własności cząsteczek rozwiązując odpowiednie równania, nie doprowadziło to jednak do zauważalnej zmiany języka, którym posługują się chemicy prowadzący badania eksperymentalne. Meteorolodzy mają numeryczne modele atmosfery i mogą na obrazach zobaczyć, że na północnej półkuli pasaty wieją z północnego wschodu, jednakże nadal prosta odpowiedź fizyków – bo działa siła Coriolisa – pozwala to lepiej zrozumieć. Sytuacja w naukach kognitywnych pozostanie zapewne podobna.

Podziękowania: Badania opisane w tej publikacji wspierał grant Narodowego Centrum Nauki UMO-2016/20/W/NZ4/00354.

8. Literatura

Bassett, D. S., and Sporns, O. (2017). Network neuroscience. Nature Neuroscience, 20(3), 353–364.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the active self a limited resource? Journal of Personality and Social Psychology, 74(5), 1252–1265.

Churchland, P.M. (1981). Eliminative Materialism and the Propositional Attitudes. The Journal of Philosophy, 78(2), 67–90.

Churchland, P. (1986). Neurophilosophy. MIT Press.

Churchland, P. M. (1995). The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain. Cambridge, Mass.: The MIT Press.

Coey, C.A., Varlet, M., & Richardson, M.J. (2012). Coordination dynamics in a socially situated nervous system. Frontiers in Human Neuroscience, 6. https://doi.org/10.3389/fnhum.2012.00164

Churchland, P. S., & Churchland, P. M. (2002). Neural worlds and real worlds. Nature Reviews. Neuroscience, 3(11), 903–907.
Dobosz, K., & Duch, W. (2010). Understanding neurodynamical systems via fuzzy symbolic dynamics. Neural Networks, 23(4), 487–496.
Duch, W (1996). Computational physics of the mind. Computer Physics Communication 97: 136-153
Duch, W (1997). Platonic model of mind as an approximation to neurodynamics. In: Brain-like computing and intelligent information systems, ed. S-i. Amari, N. Kasabov (Springer, Singapore 1997), chap. 20, pp. 491-512
Duch, W., Setiono, R., & Żurada, J. M. (2004). Computational intelligence methods for rule-based data understanding. Proceedings of the IEEE, 92(5), 771–805.
Duch, W. (2005). Brain-inspired conscious computing architecture. *Journal of Mind and Behavior* 26(1-2): 1-22.
Duch, W. (2005a). Rules, similarity, and threshold logic. Behavioral and Brain Sciences, 28(1), 23–23.
Duch, W. (2010). Reprezentacje umysłowe jako aproksymacje stanów mózgu. Studia z kognitywistyki i filozofii umysłu (red. W. Dziarnowska i A. Klawiter), 3, 5–28.
Duch, W., & Dobosz, K. (2011). Visualization for understanding of neurodynamical systems. Cognitive Neurodynamics, 5(2), 145–160.
Duch, W. (2012). Mind-Brain Relations, Geometric Perspective and Neurophenomenology, American Philosophical Association Newsletter 12(1), 1-7.
Duch, W. (2013). Brains and education: towards neurocognitive phenomics. In Learning while we are connected. Eds. N. Reynolds, M. Webb, M.M. Sysło, V. Dagiene. (Vol. 3, pp. 12–23). Toruń, Poland: Nicolaus Copernicus University Press.
Feyerabend, P., 1963, "Mental Events and the Brain," Journal of Philosophy 40:295–6.
Frank, Philipp G., ed. 1956. The Validation of Scientific Theories. Boston: Beacon Press; reprinted in 1961 by Collier Books, New York.
Haken, H. (1983). Advanced synergetics: instability hierarchies of self-organizing systems and devices. Springer-Verlag.
Hoel, E. P. (2017). When the Map Is Better Than the Territory. Entropy, 19(5), 188. https://doi.org/10.3390/e19050188
Hurlburt, R.T., Schwitzgabel, E. (2007). Describing Inner Experience? MIT Press.
Insel, T., Cuthbert, B., Garvey, M., et al. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. American Journal of Psychiatry, 167(7), 748–751.
Jantzen, K.J., & Kelso, J.A.S. (2007) Neural coordination dynamics of human sensorimotor behavior: A Review. In V.K Jirsa & R. MacIntosh (Eds.) Handbook of Brain Connectivity. Heidelberg: Springer.
Kruschke, J. K. (1996). Base rates in category learning. Journal of Experimental Psychology: Learning, Memory and Cognition, 22, 3-26.
Kutzner, F. L., & Fiedler, K. (2015). No correlation, no evidence for attention shift in category learning: different mechanisms behind illusory correlations and the inverse base-rate effect. Journal of Experimental Psychology. General, 144(1), 58–75.
Mudrik, L., & Maoz, U. (2015). „Me & my brain": exposing neuroscience's closet dualism. Journal of Cognitive Neuroscience, 27(2), 211–221.
Newell, A. (1990). Unified Theories of Cognition. Harvard University Press.
O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors (2012). Computational Cognitive Neuroscience. Wiki Book, 1st Edition. URL: http://ccnbook.colorado.edu
Quine, W.V.O., 1960, Word and Object. Cambridge, MA: MIT Press
Schwitzgabel, E. (2011). Perplexities of consciousness. MIT Press.
Sellars W., 1956, "Empiricism and the Philosophy of Mind," w: Feigl H. i Scriven M. (red). The Foundations of Science and the Concepts of Psychology and Psychoanalysis: Minnesota Studies in the Philosophy of Science, Vol. 1. Minneapolis: Uni. of Minnesota Press: 253–329.
Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: a common model for category accentuation and illusory correlation. Journal of Personality and Social Psychology, 96(2), 305–323.
Spivey, M.J. (2007). The Continuity of Mind. Oxford University Press.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science, 237, 1317–1323.
Shepard, R. N. (1994). Perceptual-Cognitive Universals as Reflections of the World. Psychonomic Bulletin & Review, 1, 2–28.
Smedslund, J. (2016). Why Psychology Cannot be an Empirical Science. Integrative Psychological & Behavioral Science, 50(2), 185–195.
Strelau, J. (2008). Temperament as a Regulator of Behavior: After Fifty Years of Research. Eliot Werner Publications.
Turing, A. (1950). Computing Machinery and Intelligence. Mind, 236, 433–460.
Varela, F. J. (1996). Neurophenomenology: a methodological remedy for the hard problem. Journal of Consciousness Studies, 3(4), 330–349. Tł. Polskie R. Poczobut, Avant 1 (2010) 31-73.
Wills, A. J., Lavric, A., Hemmings, Y., & Surrey, E. (2014). Attention, predictive learning, and the inverse base-rate effect: Evidence from event-related potentials. NeuroImage, 87 (Supp. C), 61–71.

Dedykowane Andrzejowi Klawiterowi z okazji 45-lecia pracy naukowej. Pamiętam (pewnie niezbyt dokładnie) uwagę, którą zrobił po moim referacie: świadomość nie problem, znacznie większy będziemy mieli z osobowością. Powoli do tego dojrzewamy.