

PrimeCat: Neuronowy model konsekwencji *primingu* kategorialnego¹

Marek Drogosz

Wydział Psychologii, Uniwersytet Warszawski

Marcin Jaworski

Instytut Psychologii, Uniwersytet Gdański

PRIMECAT:

A NEURAL NETWORK MODEL OF CATEGORIAL PRIMING CONSEQUENCES

Abstract. We propose theoretical and simulation models of the alternative category inhibition in categorial priming tasks. Experiments (Jaworski & Drogosz, 1998) show that relative facilitation of category activated by priming is mostly based on the mechanism of inhibition of alternative category. These results can be easily explained in neural model assuming that there are categorial units in the brain, which inhibit each other. Our neural network simulations prove that even in simpler neural systems – without „gnostic” units – inhibition of alternative categories can be the basic and probably the only one answer on priming.

An ordinary single-bodied animal can only move in one direction at a time, and this tends to constrain it to work toward only one goal at a time
Minsky, 1986, s. 166

Powyższe proste zdanie znanego teoretyka umysłu i sztucznej inteligencji, Marvin'a Minsky'ego, wyraża najbardziej podstawowy problem, z jakim musi poradzić sobie dowolny system poznawczy – konieczność nieustannego dokonywania wyboru: które z niezliczonych bodźców docierających ze środowiska są istotne; która z wielu możliwych interpretacji jest trafniejsza; które z tysięcy możliwych działań należy podjąć... Oczywiście jest, że większość z tych wyborów musi być dokonywana automatycznie, na bazie najszybszych i najbardziej podstawowych mechanizmów, aby zapewnić organizmowi efektywność reakcji. Czy jednak jest możliwe, aby efektywne reakcje systemu tak złożonego jak ludzki mózg na niezwykle bogactwo sygnałów płynących ze środowiska mogły opierać się na prostych i automatycznych mechanizmach? Nie jest to łatwe do wyobrażenia, ale mamy już narzędzie, które pozwala lepiej zrozumieć emergentne własności systemów złożonych i relacje pomiędzy całością a częścią – symulacje komputerowe.

PRIMING Z UŻYCIEM KONKURENCYJNYCH KATEGORII: PRZEWAGA MECHANIZMU HAMOWANIA

W naszych ostatnich badaniach (Jaworski, 1995; Jaworski, Drogosz, 1998) zajęliśmy się naturą konsekwencji *primingu*, który jest procedurą eksperymentalną pozwalającą w precyzyjny sposób badać złożone i subtelne zjawiska poznawcze, czego dowodzi bogata literatura (np. Zajonc, 1980; 1984; Higgins, Bargh, Lombardi, 1985). Wyniki naszych badań sugerują, że często stwierdzany w następstwie zastosowania *primingu* efekt asymilacji, czyli ułatwienia (facylitacji) używania przez badanych kategorii interpretacyjnych zgodnych z *primingiem*, może być wywołany przede wszystkim kontrastem spowodowanym utrudnieniem (zahamowaniem) stosowania kategorii alternatywnych. Ponieważ w badaniach tych stosowaliśmy *priming* podprogowy, bodźce różniące się wyłącznie na wymiarze deskryptywnym (spójne ewaluatywnie), a także precyzyjny wskaźnik czasów reakcji, uzasadniony wydaje się wniosek, że mechanizm hamowania kategorii alternatywnych wobec aktywowanych może być jednym z bazowych mechanizmów kategoryzacji, realizowanym na poziomie podstawowych procesów neurofizjologicznych. Uzasadnienie tej tezy za pomocą badania behawioralnego nie jest jednak oczywiście możliwe. Dlatego postanowiliśmy przeprowadzić komputerowe symulacje modelu dynamicznej sieci neuronowej zbudowanego na podstawie wyników naszych badań. Jako wzór posłużyły nam też wcześniejsze symulacje „zwykłego” efektu ekspozycji (*mere exposure*), których rezultaty wydawały się zachęcające (Drogosz, Nowak, 1995).

¹ Niniejszy artykuł powstał w ramach realizacji grantu KBN nr 1-P106-004-06.

Korespondencję dotyczącą artykułu należy kierować do Marka Drogosza pod adres: Wydział Psychologii UW, ul. Stawki 5/7, 00-183 Warszawa, e-mail: mardrog@sci.psych. uw.edu.pl.

Po pierwsze, przyjmujemy, że przetwarzanie bodźców jest procesem dynamicznym, tj. polega raczej na trwającym w czasie poszukiwaniu stanu równowagi systemu (atraktora) niż na mechanicznej i jednoznacznie określonej odpowiedzi na docierające z zewnątrz pobudzenie. Oprócz faktów neurofizjologicznych (Skarda, Freeman, 1987; Miyashita, Chang, 1988) świadczą o tym również udane symulacje zjawisk psychologicznych, oparte na modelach dynamicznych (por. Lewenstein, Nowak, 1989a; 1989b; Nowak, Lewenstein, 1990; Drogosz, Nowak, 1995).

Drugim ważnym założeniem o charakterze przede wszystkim logicznym jest integralność systemów przetwarzania, tj. dostępność alternatywnych kategorii interpretacyjnych wewnątrz jednego określonego systemu. W innym przypadku – tzn. gdyby każda kategoria tworzyła odrębny podsystem – układ nerwowy działałby w sposób niezwykle nieekonomiczny, ponieważ ta sama informacja byłaby przetwarzana jednocześnie w wielu różnych miejscach. Istnieją zapewne w mózgu struktury działające w taki właśnie nieekonomiczny, redundantny sposób, ale prawdopodobnie są one albo wysoce wyspecjalizowane (np. kolumny kory wzrokowej reagujące na specyficzne cechy percepcyjne bodźca), albo spełniają tylko rolę „zapasową”, zwiększając plastyczność układu i jego odporność na zaburzenia. Wydaje się, że dla zapewnienia kontroli i sterowania procesem przetwarzania konieczna jest zdolność systemu poznawczego do natychmiastowej detekcji kategorii już w czasie podstawowej obróbki bodźca. Do tego sposobu myślenia dobrze pasują znane zjawiska psychologiczne: efekt pierwszeństwa rozpoznawania słów (*word superiority effect*) oraz efekt pierwszeństwa rozpoznawania kategorii. Warto tutaj przypomnieć tezę Zajonca (1980; 1984) o istnieniu dwóch równoległych systemów przetwarzania: poznawczego i emocjonalnego. Choć nasze założenie o integralności przetwarzania nie kłóci się z tą tezą, ponieważ system poznawczy i system emocjonalny mają odrębne funkcje i zajmują się różnymi aspektami informacji, to jednak trafniejsza i bardziej spójna z naszymi postulatami wydaje się oparta na badaniach mózgu propozycja LeDoux (1989; 1994). Zakłada ona, że na pewnym etapie przetwarzania następuje aktywizacja emocjonalnych kategorii interpretacyjnych i ewentualne uruchomienie szybkiej, afektywnej ścieżki przetwarzania.

Trzecie założenie jest z jednej strony silnie związane z poprzednim, z drugiej wynika z podstawowych faktów neurofizjologicznych. Chodzi mianowicie o to, że zarówno krótko-, jak i długotrwała wiedza danego systemu neuronowego (również dotycząca dostępnych kategorii interpretacyjnych) jest zapisywana poprzez zmiany siły i struktury połączeń synaptycznych, same zaś neurony pełnią funkcję integrującą i dystrybuującą informację. Na założeniu tym oparta jest większość modeli neuronowych, jakie powstały w ciągu ostatnich 40 lat (por. np. Rosenblatt, 1958; Minsky, Papert, 1969; Rumelhart, McClelland, 1986), ale w niniejszym artykule ma ono swoje szczególne znaczenie, dlatego jest warto podkreślenia.

MODEL „C” – IMPLEMENTACJA KATEGORII W SIECI NEURONOWEJ

Modelem, który spełnia wszystkie powyższe założenia, jest zaproponowany przez Nowaka i Lewensteina (1990) tzw. model „C” (zob. też Żochowski, Nowak, Lewenstein, 1993).

Jego podstawę stanowi dynamiczna sieć atraktorowa typu Hopfielda (1982). Jest to sieć nie mająca warstw: każdy „neuron” (zazwyczaj przyjmujący tylko stany binarne) jest połączony ze wszystkimi pozostałymi za pomocą „synaps” o zróżnicowanej „sile przekazu” (wadze). Charakterystyczną własnością tej sieci jest proces rozpoznawania bodźców, który polega na dynamicznym poszukiwaniu najlepszego stanu równowagi (spoczynku) całego systemu lub, inaczej mówiąc, stanu minimalnej energii.

Nowak i Lewenstein zaproponowali proste i pomysłowe uzupełnienie tego modelu. Mianowicie przydzielili losowo poszczególnym „połączeniom synaptycznym” różne „kolory”, tworząc w ten sposób kilka podsieci wewnątrz systemu, funkcjonującego w dalszym ciągu jako integralna całość (neurony, czyli jednostki przetwarzające, pozostały jakby „na zewnątrz” podsieci albo też, innymi słowy, zachowały przynależność do wszystkich podsieci), wykorzystująca w danym momencie tylko część połączeń do zapamiętywania napływającej informacji. Okazało się, że nie tylko nie ogranicza to zdolności pamięciowych modelu, ale bardzo wyraźnie je poprawia! Aby jednak proces odtwarzania informacji (rozpoznawania) mógł skutecznie przebiegać we właściwej podsieci, Nowak i Lewenstein wprowadzili dodatkowy parametr globalnej wagi podsieci, sprzężony ze średnim poziomem aktywacji w podsieci: im niższy poziom aktywacji, tym niższa waga całej podsieci, a więc mniejsze znaczenie w procesie przetwarzania bodźca. Rozwiązanie takie jest w pewnym stopniu uzasadnione – np. zupełnie naturalnie pojawia się w takiej sieci efekt pierwszeństwa kategorii – ma jednak istotne wady, do których nawiązemy w dalszej części artykułu.

PRIMECAT: NEURONOWY MODEL KONSEKWENCJI PRIMINGU KATEGORIALNEGO
WYJŚCIOWY MODEL HAMOWANIA INTERKATEGORIALNEGO
OPARTY NA JEDNOSTKACH KATEGORIALNYCH

Łatwo nasuwającą się interpretacją zjawiska hamowania alternatywnych kategorii interpretacyjnych w wyniku *primingu* może być istnienie wzajemnie hamujących się jednostek kategorialnych. Jednostki takie byłyby wyspecjalizowanymi neuronami lub grupami neuronów, które uzyskują zdolność detekcji określonej kategorii w trakcie wielokrotnego rozpoznawania egzemplarzy konkurencyjnych kategorii przez system poznawczy. Taki sposób myślenia o mózgu – jako systemie wyposażonym w jednostki gnostyczne – jest bardzo rozpowszechniony od czasu ukazania się klasycznej pracy Konorskiego (1969).

Rys. 1. Schemat hipotetycznego modelu sieci neuronowej z dwiema podsieciami połączeń (zaznaczone ciągłymi i przerywanymi krzywymi) oraz „jednostkami kategorialnymi” (zaznaczone paskami); strzałki symbolizują pobudzenie płynące do jednostek kategorialnych; czarne kropki – synapsy hamujące. Ze względu na czytelność rysunku zaznaczono tylko część połączeń.

Hipotetyczny system dysponujący jednostkami kategorialnymi (zob. rys. 1) funkcjonowałby w następujący sposób: z sieci rozpoznającej dany bodziec płynie pobudzenie do neuronów kategorialnych, które się wzajemnie hamują; silniej pobudzony neuron „wygrywa” i wysyła zwrótnie impulsy hamujące (lub nawet blokujące) połączenia synaptyczne należące do konkurencyjnych podsieci-kategorii.

Zauważmy, że funkcjonalna charakterystyka naszego hipotetycznego modelu w dużym stopniu pokrywa się z właściwościami modelu „C”. Sprawdzanie średniego pobudzenia w podsieci jest odpowiednikiem wysyłania pobudzenia do neuronu kategorialnego; impulsy hamujące płynące od neuronu kategorialnego do połączeń należących do konkurencyjnej podsieci działają tak, jak zmiana globalnej wagi tej podsieci.

Symulacje modelu „C” wykonane przez Nowaka i Lewensteina (1990) oraz Żochowskiego, Nowaka, Lewensteina (1993) pokazują, że zaimplementowany w taki sposób mechanizm hamowania interkategorialnego efektywnie wywiązuje się z zadania kontroli procesu przetwarzania i pozwala na detekcję kategorii jeszcze przed ostatecznym rozpoznaniem bodźca. Ponadto z punktu widzenia neurofizjologicznego wydaje się, że podobne struktury, oparte na hamującym sprzężeniu zwrotnym, są często spotykanym elementem anatomicznym w układzie nerwowym. Należy zatem postawić pytanie, czy istnieje w ogóle potrzeba poszukiwania innych rozwiązań?

OGRANICZENIA MODELU Z JEDNOSTKAMI KATEGORIALNYMI

Model oparty na rywalizujących neuronach kategorialnych ma jednak istotne ograniczenia i słabości. Po pierwsze, jest całkowicie mechanicznym przełożeniem idei konkurujących kategorii na język neuronowy, tzn. rozwiązuje problem poprzez dodanie zewnętrznego mechanizmu, którego wyłącznym zadaniem jest właśnie zahamowanie aktywności „przegrywających” kategorii. Jest więc w istocie pewnego rodzaju formalnym i symulacyjnym opisem zjawiska hamowania interkategorialnego, a nie jego wyjaśnieniem. Nie jest też propozycją jakiegoś nowego mechanizmu neuropoznawczego, gdyż ze względu na modułarną budowę (czy też posługiwanie się globalnymi parametrami) proces na poziomie neuronowym odpowiada dokładnie zachowaniu się całej sieci – nie mamy więc do czynienia z emergencją.

Po drugie, model mający wyspecjalizowane jednostki kategorialne jest nieekonomiczny, i to co najmniej w dwóch aspektach: dużej liczby połączeń pełniących funkcje czysto strukturalne, tzn. nie biorących udziału w magazynowaniu informacji (połączenia od neuronu kategorialnego do konkurencyjnych podsieci), oraz znacznego wydłużenia czasu przetwarzania bodźców (konieczne uaktywnienie pełnej pętli sprzężenia zwrotnego).

Po trzecie, ograniczeniem tego modelu jest to, że cała informacja o zawartości kategorii musi być zapisana nie tylko w stosunkowo licznych połączeniach synaptycznych danej podsieci (np. 5000, jeśli system składa się ze 100

MAREK DROGOSZ, MARCIN JAWORSKI

neuronów i 2 podsieci), ale również powtórnie w niewielu połączeniach aktywujących neuron kategoryalny (maks. zaledwie 100 w tych samych warunkach). Jest to więc „wąskie gardło”, które zdecydowanie ogranicza zdolność uczenia się systemu, zwłaszcza jeśli w odrębnych kategoriach zapisywane są podobne informacje. Należy tutaj zaznaczyć, że zarzuty nieekonomiczności i ograniczenia pamięci nie odnoszą się do oryginalnego modelu „C”, w którym hamujące sprzężenie zwrotne oparte było na czysto matematycznym parametrze globalnej wagi podsieci, a nie na neuronach kategoryalnych. Neurofizjologiczne przesłanki użycia tego parametru są jednak dość niejasne. Zaproponowaliśmy tutaj pewną jego interpretację ze względu na jej walory analityczne. Możliwa jest również inna interpretacja, którą ostatecznie zdecydowaliśmy się zastosować w naszych eksperymentach symulacyjnych. Jest ona mniej intuicyjna, ale daje lepszy wgląd w zależność pomiędzy mikroprocesami a emergentnymi własnościami systemu neuronalnego.

MODEL PRIMECAT:

RYWALIZACJA KATEGORII NA BAZIE HAMOWANIA LOKALNEGO

Model PrimeCat powstał w celu zbadania możliwych mechanizmów neuronowych *primingu* kategoryalnego, w wyniku którego następuje zahamowanie kategorii alternatywnych wobec aktywizowanych. Powstał on na bazie modelu „C” Nowaka i Lewensteina. Zrezygnowaliśmy z rozwijania modelu z wyspecjalizowanymi neuronami kategoryalnymi ze względu na jego opisane wyżej ograniczenia, a także – co ważniejsze – dlatego że, jak sądziliśmy, konsekwencje *primingu* mogą być związane z bardziej podstawowymi i mniej specyficznymi procesami neuronalnymi.

Takimi absolutnie podstawowymi procesami są z pewnością zjawiska zachodzące w obrębie pojedynczego drzewka dendrytycznego, czyli połączeń synaptycznych dochodzących do jednego neuronu. W ostatnich latach odkryto wiele mechanizmów, które całkowicie zmieniają nasze wyobrażenia o funkcji synaps i dendrytów w przetwarzaniu informacji: zamiast o pasywnym przestrzenno-czasowym sumowaniu pobudzenia, myślimy teraz o aktywnej detekcji specyficznych czasowo-przestrzennych wzorców pobudzenia oraz zdolności do błyskawicznej adaptacji (por. Koch, 1997). Dokładne omówienie tych mechanizmów przekracza ramy niniejszego artykułu, ale wspomnieć należy o przynajmniej dwóch, ważnych dla naszego modelu.

Pierwszy to rywalizacja pobudzeń docierających do pojedynczego neuronu, która może być realizowana przez rodzaj „krzyżowego wyłączania” (*cross-exclusion*; por. Minsky, 1986), czyli bezpośrednie połączenia hamujące pomiędzy aksonami należącymi do różnych podsieci (zob. rys. 2), a także np. przez propagację impulsu nerwowego „w górę” drzewka dendrytycznego (*antidromic spike invasion*). Drugim jest zdolność przynajmniej niektórych synaps do szybkiej adaptacji, czyli „zmiany swojej efektywnej wagi w sposób ciągły w czasie przetwarzania” (Koch, 1997). Doszliśmy do wniosku, że właśnie te działające na lokalnym poziomie mechanizmy mogą być odpowiedzialne za zachowanie całego systemu w reakcji na *priming*, tj. za zahamowanie kategorii alternatywnych, a w konsekwencji opóźnienie lub wydłużenie reakcji na reprezentujące je bodźce.

Rys. 2. Schemat mechanizmu lokalnego hamowania interkategoryalnego na przykładzie dwóch podsieci-kategorii (połączenia należące do jednej podsieci zaznaczono linią ciągłą, do drugiej – przerywaną; połączenia hamujące są oznaczone cieńszymi liniami zakończonymi czarną kropką).

Aby uzyskać w naszym modelu PrimeCat czysto lokalne oddziaływania, zmodyfikowaliśmy oryginalny model „C” w taki sposób, że zamiast parametru globalnej wagi podsieci, wprowadziliśmy parametr plastyczności, tj. krótkotrwałych zmian zdolności przewodzenia impulsów dla każdej pojedynczej synapsy. Rywalizacja podsieci w modelu PrimeCat może się więc odbywać wyłącznie na poziomie lokalnym, w obrębie połączeń dochodzących

PRIMECAT: NEURONOWY MODEL KONSEKWENCJI PRIMINGU KATEGORIALNEGO

do pojedynczego neuronu.

Kolejnym elementem, który wprowadziliśmy do modelu, jest możliwość dwojakiego charakteru oddziaływań interkategorialnych: podsieć, która najsilniej aktywuje dany neuron może uzyskać przewagę nie tylko dzięki osłabieniu połączeń od innych podsieci do tego neuronu, ale też poprzez wzmocnienie swych własnych połączeń. Chcieliśmy w ten sposób sprawdzić, czy lokalne mikroprocesy hamowania i facylitacji mają odmienny wpływ na całościowy przebieg przetwarzania.

Ostatnim ważnym elementem modelu PrimeCat jest zdefiniowanie *primingu*: zakładamy, że jest to niepełny (niezakończony rozpoznaniem) proces przetwarzania bodźca poprzedzającego, który pozostawia po sobie ślad w postaci krótkotrwałych zmian zdolności przewodzenia synaps (dzięki wyposażeniu ich w cechę plastyczności), co z kolei ma wpływ na przetwarzanie bodźców bezpośrednio następujących.

Dokładniejszy formalny opis modelu PrimeCat zamieściliśmy w Aneksie.

MAREK DROGOSZ, MARCIN JAWORSKI

PRZEBIEG I WYNIKI SYMULACJI Z UŻYCIEM MODELU PRIMECAT

Symulacje przeprowadziliśmy z użyciem 220 zbudowanych niezależnie sieci, składających się z 200 neuronów. Każda z nich miała dwie podsieci-kategorie (co odpowiada dwóm kategoriom, jakie badaliśmy w eksperymencie psychologicznym), w których zapamiętywane było po 10 bodźców (były one losowane, aby zachować maksymalną ortogonalność). Następnie wzorce były rozpoznawane w czterech różnych sytuacjach: 1) bez *primingu*; 2) *primingu* z użyciem bodźca z tej samej kategorii; 3) *primingu* z użyciem bodźca z alternatywnej kategorii oraz 4) *primingu* z użyciem nieznanego, losowego bodźca. Dodatkowo manipulowaliśmy kilkoma innymi parametrami: siłą lokalnego hamowania i facylitacji, czasem *primingu* oraz stopniem wstępnego zniekształcenia bodźca rozpoznawanego.

Zmienną zależną, którą badaliśmy, był średni czas reakcji (rozpoznania bodźca), przy założeniu wymaganej zgodności z oryginałem na poziomie 95%.

Uzyskane w naszych symulacjach wyniki można streścić w następujących punktach:

1. Konsekwencją *primingu*, wynikającą ze strukturalnych cech systemu neuronowego niezależnie od proporcji sił lokalnego hamowania i facylitacji, jest zawsze zahamowanie lub wydłużenie czasu przetwarzania bodźców należących do alternatywnych kategorii (rys. 3-7); nigdy natomiast nie zachodzi przyśpieszenie przetwarzania bodźców należących do kategorii aktywizowanych przez *priming*.
2. Kluczowym procesem, na którym opiera się rywalizacja alternatywnych kategorii interpretacyjnych, jest hamowanie lokalne (rys. 3 i 4); proces facylitacji (wzmocnienia) kategorii aktywizowanej poprawia co prawda efektywność systemu (rys. 4), ale zastosowanie wyłącznie facylitacji bez hamowania powoduje paradoksalny efekt wydłużenia czasu przetwarzania bodźców należących do kategorii aktywizowanej.
3. Wpływ skuteczności (czasu) *primingu* w normalnych warunkach rozpoznawania (tzn. przy pewnym poziomie zaburzeń, szumu) jest nieistotny; ujawnia się natomiast w sytuacji idealnego przetwarzania bez zaburzeń: efekt wyraźnego zahamowania alternatywnej kategorii pojawia się dopiero przy dłuższym, bardziej skutecznym *primingu*; im łatwiej więc sieć neuronowa może przetwarzać bodźce, tym skuteczniejszych procedur *primingu* należy użyć, by efekt ten osiągnąć (rys. 6 i 7).

PRIMECAT: NEURONOWY MODEL KONSEKWENCJI PRIMINGU KATEGORIALNEGO

Rys. 3. Średni czas rozpoznania bodźców dla różnych rodzajów *primingu* przy włączonym mechanizmie lokalnego hamowania, a wyłączonej facylitacji (hamowanie = 0,2, facylitacja = 0, zniekształcenie bodźca = 20%).

Rys. 4. Średni czas rozpoznania bodźców dla różnych rodzajów *primingu* przy włączonych mechanizmach lokalnego hamowania i facylitacji (hamowanie = 0,2, facylitacja = 0,2, zniekształcenie bodźca = 20%).

Rys. 5. Średni czas rozpoznania bodźców dla różnych rodzajów *primingu* przy wyłączonym mechanizmie lokalnego hamowania, a włączonej facylitacji (hamowanie = 0, facylitacja = 0,2, zniekształcenie bodźca = 20%).

Rys. 6. Średni czas rozpoznania bodźców dla różnych rodzajów *primingu* w sytuacji krótkiego (nieskutecznego) *primingu* i w warunkach idealnej „percepcji” bodźców (czas *primingu* = 0,5, MC, zniekształcenie bodźca = 0%).

Rys. 7. Średni czas rozpoznania bodźców dla różnych rodzajów *primingu* w sytuacji długiego (skutecznego) *primingu* i w warunkach idealnej „percepcji” bodźców (czas *primingu* = 1, MC, zniekształcenie bodźca = 0%).

DYSKUSJA

Przed wszystkim musimy stwierdzić, że rezultaty symulacji przeszły nasze oczekiwania, gdyż otrzymaliśmy wyniki zgodne z naszymi eksperymentami psychologicznymi natychmiast po zaimplementowaniu modelu opartego na lokalnej rywalizacji kategorii, bez zwykłej w takich przypadkach fazy prób i błędów. Uznajemy to za kolejne potwierdzenie słuszności myślenia o pewnych aspektach przetwarzania informacji w kategoriach dynamicznych systemów neuronowych.

Udało się nam pokazać zarówno we wcześniejszych eksperymentach, jak i w opisanej wyżej symulacji komputerowej, że to właśnie zahamowanie kategorii alternatywnej wobec aktywizowanej jest naturalną konsekwencją *primingu*. Nie zachodzi nic takiego, jak usprawnienia czy przyspieszenie przetwarzania bodźców należących do kategorii aktywizowanej, gdyż system, który nie został specjalnie „zmylony”, przetwarza wszystkie wzorce – również nie poprzedzone *primingiem* – w taki sam optymalny sposób, na zasadzie użycia minimalnej energii. Trzeba jednak podkreślić, że mówimy tu tylko o jednorodnym, silnie zintegrowanym systemie. Można przypuszczać, że w przypadku kilku systemów kategoryalnych komunikujących się ze sobą (np. kategorii różniących się nie tylko deskryptywnie, ale również ewaluatywnie, albo bodźców o różnych modalnościach zmysłowych) podstawowy mechanizm supresji alternatywnych kategorii może być obudowany dodatkowymi mechanizmami facylitującymi określone sposoby przetwarzania informacji. Mogłoby to tłumaczyć, dlaczego w niedawnych eksperymentach Macrae i in. (1995) okazało się, że pewne rodzaje kategoryzacji sterowane są zarówno przez mechanizmy hamujące, jak i pobudzające – w swoich procedurach eksperymentalnych wykorzystywali oni złożone bodźce o różnych modalnościach zmysłowych.

Nieoczekiwanym i interesującym rezultatem jest paradoksalny na pierwszy rzut oka efekt wydłużenia czasu przetwarzania przy *primingu* zgodnym z kategorią bodźca i włączonym jedynie mechanizmem lokalnego wzmocnienia (facylitacji) aktywizowanej kategorii oraz braku hamowania kategorii konkurencyjnej. W tych warunkach wydłużenie czasu przetwarzania wydaje się ostatnią rzeczą, jakiej moglibyśmy się spodziewać. A jednak ma to swoje logiczne uzasadnienie. Mechanizm hamowania oddziałuje wyłącznie na kategorie alternatywne, zmieniając czasowo charakterystykę tworzących je połączeń; połączenia aktywizowanej kategorii pozostawia w niezmienionym stanie pełnej gotowości do przetwarzania dowolnego, należącego do niej bodźca. Mechanizm facylitacji oddziałuje na połączenia kategorii aktywizowanej, zmieniając czasowo strukturę jej połączeń tak, że faworyzowany jest właśnie przetwarzany bodziec, a pamięć pozostałych członków kategorii zostaje częściowo zamazana. Być może w ten właśnie sposób otrzymujemy odpowiedź, dlaczego do realizacji procesów kategoryzacji znacznie lepiej nadaje się mechanizm hamowania.

BIBLIOGRAFIA

- Amit, D. J. (1989). *Modeling brain function: The world of attractor neural networks*. Cambridge: Cambridge University Press.
- Arbib, M. (1995). *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press.
- Drogosz, M., Nowak, A. (1995). Symulacyjna teoria efektu ekspozycji: sieć neuropodobna EXAC. *Przegląd Psychologiczny*, 38, 65-84.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: John Wiley.
- Higgins, E. T., Bargh, J. A., Lombardi, W. J. (1985). Nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 59-69.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of National Academy of Science, USA*, 79, 2554-2558.
- Jaworski, M. (1995). Wyznaczniki interpretacji informacji społecznej: przystępność kategorii i obrona samooceny. *Przegląd Psychologiczny*, 38, 305-321.

MAREK DROGOSZ, MARCIN JAWORSKI

- Jaworski, M., Drogosz, M. (1998). Hamowanie alternatywnej struktury poznawczej jako mediator efektu asymilacji. *Przegląd Psychologiczny*, 41(1-2), 135-150.
- Koch, C. (1997). Computation and the single neuron. *Nature*, 385, 207-210.
- Konorski, J. (1969). *Integracyjna działalność mózgu*. Warszawa: PWN.
- LeDoux, J. E. (1989). Cognitive-emotional interactions in the brain. *Cognition and Emotion*, 4, 267-289.
- LeDoux, J. E. (1994). Emocje, pamięć, mózg. *Świat Nauki*, 8, 34-41.
- Lewenstein, M., Nowak, A. (1989a). Recognition with self-control in neural networks. *Physical Review*, A40, 4652-4664.
- Lewenstein, M., Nowak, A. (1989b). Fully connected neural networks with self-control of noise levels. *Physical Review Letters*, 62, 225-228.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, 69, 397-407.
- Minsky, M. (1986). *The society of mind*. New York: Simon and Schuster.
- Minsky, M., Papert, S. (1969). *Perceptions*. Cambridge: MIT Press.
- Miyashita, Y., Chang H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331, 68-70.
- Nowak, A., Lewenstein, M. (1990). Excitation neural networks (niepublikowany maszynopis).
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386-408.
- Rumelhart, D. E., McClelland, J. L. (1986). *Parallel distributed processing: Exploration in microstructure of cognition* (vol. 1). Cambridge: MIT Press.
- Skarda, C. A., Freeman W. J. (1987). How brains make chaos to make sense of the world. *Behavioral and Brain Sciences*, 10, 161-195.
- Zajonc, R. B. (1980). Feeling and thinking. Preferences need no inferences. *American Psychologist*, 35, 151-175.
- Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, 39, 117-123.
- Zochowski, M. R., Nowak, A., Lewenstein, M. (1993). Memory that tentatively forgets. *Journal of Physics A*, 26, 2453-2460.

PRIMECAT: NEURONOWY MODEL KONSEKWENCJI PRIMINGU KATEGORIALNEGO
ANEKS

Opis formalny modelu PrimeCat

Model bazowy: autoasocjacyjna, atraktorowa, dynamiczna, w pełni połączona sieć neuronowa typu Hopfielda (1982) z binarnymi neuronami oraz połączeniami pogrupowanymi w podsieci (Nowak, Lewenstein, 1990).

Reguła uczenia: podstawowa reguła Hebb'a (Hebb, 1949; Amit, 1989; Arbib, 1995) stosowana niezależnie dla każdej podsieci połączeń (wzór 1):

(1)

gdzie: w_{ij}^C – waga połączenia pomiędzy neuronami i-tym i j-tym w podsieci C; N – liczba neuronów; P – liczba uczonych bodźców; s – stan neuronu.

Reguła rozpoznawania: asynchroniczna metoda Monte Carlo.

Reguła aktywacji neuronu: funkcja progowa (próg = 0) sumy pobudzeń w najsilniejszej podsieci (wzory 2, 3 i 4):

(2)

(3)

(4)

gdzie: I_i^C – siła pobudzenia z podsieci C; p_{ij} – plastyczność synapsy i-j; h – stała hamowania/facylitacji.