ISSN 0006-3509, Biophysics, 2009, Vol. 54, No. 2, pp. 204–207. © Pleiades Publishing, Inc., 2009.
Original Russian Text © E.V. Budilova, M.P. Karpenko, L.M. Kachalova, A.T. Terekhin, 2009, published in Biofizika, 2009, Vol. 54, No. 2, pp. 287–292.

COMPLEX SYSTEMS BIOPHYSICS

Memory Blocking: Facts, Problems, and Models

E. V. Budilova[a, b], M. P. Karpenko[a], L. M. Kachalova[a], and A. T. Terekhin[a, b]

[a] *Institute of Cognitive Neurology, Modern University for Humanities, Moscow, 109029 Russia*
[b] *Biological Faculty, Moscow State University, Moscow, 119899 Russia*
e-mail: terekhin_a@mail.ru

Received April 24, 2007; in final form, August 11, 2008

Abstract—The tip-of-the-tongue state, or memory blocking, is considered with regard to the feasibility of its neural network modeling. The results of psycholinguistic and neurobiological studies on memory blocking are reviewed, and basic problems that need be solved to comprehend this phenomenon are formulated. One such point is the dramatic discrepancy between the subjective assurance that an image is familiar and the inability to recollect it fully. To explain this discrepancy, we propose a biologically plausible neural network model of recognition, demonstrating cardinal superiority in the capacity of image recognition over its recollection.

Key words: tip-of-the-tongue state, Hopfield network, energy function, familiarity recognition, recollection

DOI: 10.1134/S0006350909020158

INTRODUCTION

Perhaps every one of us has more than once been in a situation when one was unable to recollect a word needed at the moment, being though absolutely sure it was not forgotten but kept safe “somewhere deep in memory,” which was indeed confirmed afterward as in a few hours or a few days the word that could not be recalled anyhow (usually a name of a person or a geographical place) itself “surfaced from memory.” This phenomenon, vividly described by Chekhov in his short story “A Horsy Surname” [1], since the late XIX century is marked in the scientific literature as a special psychological condition [2], called the “tip-of-the-tongue” state, or “blocking” [3, 4].

Investigation of spontaneous memory blocking, being of doubtless interest for basic science, is also important in the applied aspect, as it can expedite practical means of correcting this cognitive defect. The need for such means rises in the modern epoch in connection with its two characteristic features. One is the constantly growing extent of “informatization” of all aspects of life of the society, which places high demands for memory efficiency on actively working persons. The other is the demographic shift toward longevity and, consequently, the larger fraction of elderly people, whereas the risk of memory blocking increases with age.

In the present work, the problem of memory blocking is examined from the standpoint of the feasibility of building mathematical (computer) neural network (NNW) models of this phenomenon. First we review the results of experimental psycholinguistic and neurobiological studies concerned with memory blocking. Then we formulate the main questions ensuing from these results, and attempt some answers by building and assessing an appropriate NNW model.

EXPERIMENTAL FACTS RELATED TO BLOCKING

Even everyday experience reveals two most characteristic features of memory blocking: risk increasing with age, and more frequent blocking of proper nouns versus common nouns. Because of this combination, the inability to recall a name is often the main (or the sole) complaint of elderly people, marking the age-related changes in cognitive ability.

These rules are indeed objective and are confirmed experimentally. Proof has been presented [5–8] that the incidence of blocking positively correlates with age. In particular, when examinees were asked to register all cases of memory blocking over a month, the blocking frequency proved higher in older persons [5, 6]. To add, in all age groups the majority of cases involved proper nouns, mainly names and surnames of people out of contact for a long time. Common words that were blocked were those rarely used in the language. Pronounced deterioration of the ability to recall proper names has been noted in altitude hypoxia [9].

Convincing support of the association between the affiliation of a word with a certain semantic category and the possibility of its recollection has been obtained in observations of patients with brain disorders. Thus there was a case [10] when a patient showed no deterioration with respect to common nouns but was virtually unable to recall any proper names—be it people, pets, buildings, trade marks, etc. Conversely, other patients [11, 12] were incapable of recollecting common nouns but could remember names of famous persons or countries. In still other cases the patients could not recall personal names but remembered geographical names [13–17], to the extent when geographical names were the only category of retrievable words [16]. Memory disorders may concern concrete words only [18] or conversely, only abstract ones [19], only names of living beings [20–22] or only inanimate objects [21]. Cases are known with even narrower categories blocked—body parts [23, 24], fruits and vegetables [25], animals [26].

Observations of patients with partial brain lesions also allow some conclusions about the association of blocking in various word categories with particular regions of the brain. Thus there are reports on selective blocking of people's names upon local damage to the left temporal lobe [13–15, 27–30]. In one case [23] resection of the anterior part of the left temporal lobe impaired naming of body parts.

In a number of neurobiological studies, the brain regions associated with memory blocking were sought for with modern noninvasive techniques of registering brain activity (fMRI, PET, ERP, MEG), which can be applied to healthy persons. Thus [31] functional magnetic resonance investigation (fMRI) revealed that human name recollection is linked to the anterior part of the left temporal lobe, and that recently memorized names also require activity of the right temporal lobe. The anatomical localization of brain activity in recalling various word categories, such as verbs and nouns (and among nouns, narrow groups such as tools, food products, body parts, etc.) is confirmed by many other instrumental investigations [32–35].

The overall pattern of the processing of lexical information in the human brain that emerges from such studies can be summarized as follows [36]. First, the regions involved in lexical activity are not homogeneous but consist of small isolated spots of activity linked to various linguistic components (this first of all pertains to accessing phonetic information via the middle temporal gyrus, where the local activity is tightly connected with the semantics of the analyzed word). Second, such activity is not restricted to the classical lexically oriented regions (inferior frontal [Broca's] gyrus, Wernicke's zone and angular gyrus) but expand onto the superior left temporal gyrus, temporal pole, lingual and fusiform gyri, medial prefrontal zone and insula. Third, the localization of brain lexical activity is more adequately described in linguistic terms (phonology, syntax, semantics) than in terms of action (listening, pronouncing, reading).

PUZZLES OF MEMORY BLOCKING

Analysis of the above experimental facts gives rise to several questions about the particular neuropsychological manifestations of memory blocking; it is desirable to get answers at the level of the basic NNW mechanisms underlying these manifestations.

First, the blocking phenomenon vividly demonstrates the difference between the awareness of the person trying to reproduce some information that it is contained in his/her memory and the inability to retrieve this information. What are the NNW bases of this difference?

Second, the extent of blocking is associated with the semantics of the blocked information. What is the NNW basis of this linkage? And in this connection, what are the mechanisms whereby information of different semantic categories is located in different cortex regions?

Third, the blocked information is often recollected after some time. How can this event be explained in NNW terms?

Fourth, why do unfavorable factors (aging, hypoxia) selectively aggravate blocking of certain word categories, such as proper names? What specific changes are caused by these factors in the cerebral NNW performance?

Here we try to answer the first of these questions by building a NNW model of recognition.

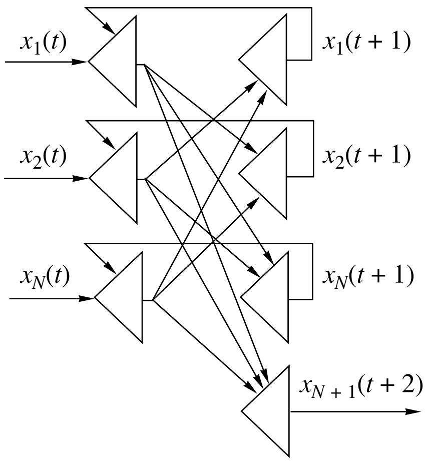
NEURAL NETWORK MODELING OF RECOGNITION AND RECOLLECTION

The dramatic contrast between being sure of possessing some information and being unable to extract it may be explained, at least partly, by the objective difference between the mechanism of familiarity recognition and the mechanism of recollection. This difference is revealed experimentally. For instance, having been presented several thousand images, examinees could almost unmistakably tell whether they have already seen the next image shown, whereas the capability of detailed reproduction was substantially lower [37].

NNW modeling of recognition is based on the finding that the energy function of a Hopfield network [38] takes larger values for memorized images than for newly presented ones [39].

A Hopfield network consists of N neurons interconnected symmetrically with synaptic weights $w_{ij} = w_{ji}$ (assuming $w_{ii} = 0$). The state of neuron i at moment t is set by the state variable $x_i(t)$ taking values –1 or 1, and the state of the whole network is defined by vector $X_t = (x_1(t), x_2(t), \ldots, x_N(t))$. The network dynamics is determined by the rules of change in neuron states and synaptic weights. If at moment t the weights $w_{ij}(t)$ and states $x_i(t)$ of all neurons are given, at the next moment $t + 1$ the state of neuron i will change according to the McCulloch–Pitts rule [40]:

$$x_i(t+1) = \operatorname{sgn}\left(\sum_{j=1}^{N} w_{ij}(t)x_j(t)\right), \qquad (1)$$



Scheme of a neural network for image recognition based on the familiarity function.

where sgn(x) is a function equaling –1 for nonpositive and +1 for positive argument values. The synaptic weights for all $j \neq i$ change by the Hebb rule [41]:

$$w_{ij}(t+1) = w_{ij}(t) + x_i(t)x_j(t), \qquad (2)$$

i.e., the weight increases by unity if neurons i and j are in the same state, and decreases by unity otherwise. For $j = i$ the weight is set to zero. The energy function of the Hopfield network with state vector $\{x_1(t), x_2(t), \ldots, x_N(t)\}$ and weight set $\{w_{ij}(t), i, j = 1, \ldots, N\}$ is

$$E(X_t) = -0.5\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(t)x_i(t)x_j(t). \qquad (3)$$

It has been shown [42] that the maximal number of images that can be presented to the Hopfield network and then recognized as familiar by the corresponding energy (3) with less than 1% error, equals $0.023N^2$; at large N this is orders of magnitude greater than the maximal number of images that can be memorized by the network to be fully reproducible, $0.145N$ [39]. The same work [42] also describes some architectures of NNWs that can provide biologically plausible calculations close to those required by (3), without substantial decline in the number of images memorizable for recognition.

We offer a new NNW model of recognition, also based on a modification of (3) but, in our opinion, more naturally matching the Hopfield network. The modification consists in replacing the inner sum in (3) with its sign (we also discard the multiplier –0.5 as insignificant in this context):

$$E^*(X_t) = \sum_{i=1}^{N} x_i(t)\,\mathrm{sgn}\left(\sum_{j=1}^{N} w_{ij}(t)x_j(t)\right), \qquad (4)$$

whereupon the appropriate calculation can be performed with (1), the basic formula for neuron state dynamics. After this (4) converts to

$$E^*(X_t) = \sum_{i=1}^{N} x_i(t)x_i(t+1). \qquad (5)$$

Hence, the "familiarity function" $E^*(t)$ simply equals the scalar product of neuron state vectors in two consecutive time steps. Maximal $E^*(t)$ is reached in the stationary points of the network, characteristic of memorized images, for which $x_i(t+1) = x_i(t)$; while for new images, when $x_i(t+1)$ is in fact a random transform of $x_i(t)$, the $E^*(t)$ is close to zero. The product (5) can also be obtained with (1), whereby in fact we calculate the scalar product of the $\{w_{ij}(t), j = 1, \ldots, N\}$ vector of weights from each neuron to i by vector $\{x_1(t), x_2(t), \ldots, x_N(t)\}$. We must simply replace $\{w_{ij}(t), j = 1, \ldots, N\}$ with $\{w_{ij}(t+1), j = 1, \ldots, N\} = \{x_1(t), x_2(t), \ldots, x_N(t)\}$ and $\{x_1(t), x_2(t), \ldots, x_N(t)\}$ with $\{x_1(t+1), x_2(t+1), \ldots, x_N(t+1)\}$. This can be done by introducing an additional "recognition neuron" $N+1$, the weights to which $\{w_{N+1,j}(t+1), j = 1, \ldots, N\}$ at step t in accordance with (2) are taken equal to $\{x_1(t), x_2(t), \ldots, x_N(t)\}$, assuming that earlier weights $\{w_{N+1,j}(t), j = 1, \ldots, N\}$ are "forgotten" and zeroed, and that neuron $N+1$ is in active state. Then at step $t+1$ according to (1) we perform scalar multiplication of these weights by input vector $\{x_1(t+1), x_2(t+1), \ldots, x_N(t+1)\}$ and take the sign of the result

$$x_{N+1}(t+2) = \mathrm{sgn}\left(\sum_{j=1}^{N} w_{N+1,j}(t+1)x_j(t+1)\right). \qquad (6)$$

This completes the recognition, giving an active state $x_{N+1}(t+1) = +1$ for a familiar input image $\{x_1(t), x_2(t), \ldots, x_N(t)\}$ or inactive state $x_{N+1}(t+1) = -1$ for an unfamiliar one. The advantage of this scheme over those mentioned above [42] is that calculations at step t are performed not by an additional special network, but by the basic Hopfield network plus just one terminal recognition neuron. The network is outlined in the figure.

It can be shown that the memory capacity of a NNW based on the familiarity function $E^*(t)$ is $0.0185N^2$, insignificantly lower than the $0.023N^2$ [42] obtained with the energy function $E(t)$. Thus, a quadratic dependence of the number of recognizable images on the neuron number can be obtained with a simpler NNW mechanism; this argues in favor of that in some form such a mechanism can be realized in the brain, and is responsible for the contrast between recognition and recollection in memory blocking.

The details of calculating the network capacity are given elsewhere [43].

REFERENCES

1. A. P. Chekhov, *A Horsy Surname* (Peterburgskaya Gazeta, St. Petersburg, 1885) [in Russian].
2. W. James, *The Principles of Psychology* (Holt, New York, 1890).
3. B. L. Schwartz, *Tip-of-the-Tongue States: Phenomenology, Mechanism, and Lexical Retrieval* (Lawrence Erlbaum Association, New Jersey, 2001).
4. D. L. Schacter, *The Seven Sins of Memory – How the Mind Forgets and Remembers* (Houghton Mifflin Co, Boston, 2001).
5. D. M. Burke, D. G. MacKay, J. S. Worthley, and E. Wade, J. Memory Language **30**, 542 (1991).
6. M. K. Heine, B. A. Ober, and G. K. Shenaut, Psychol. Aging **14**, 445.
7. M. Evrard, Brain Language **81**, 174 (1999).
8. D. M. Burke and M. A. Shafto, Curr. Direct. Psychol. Sci. **13**, 21 (2004).
9. G. Pelamatti, M. Pascotto, and C. Semenza, Cortex **39**, 97 (2003).
10. A. W. Ellis, A. W. Young, and E. M. R. Critchley, Brain **112**, 1469 (1989).
11. L. Cipolotti, J. E. McNeil, and E. K. Warrington, Memory **1**, 2891 (1993).
12. C. Semenza and M. T. Sgaramella, Memory **1**, 265 (1993).
13. P. McKenna and E. Warrington, J. Neurol. Neurosurg. Psychiatry **43**, 781 (1980).
14. C. Semenza and M. Zettin, Nature **342**, 678 (1989).
15. F. Lucchelli and E. De Renzi, Cortex **28**, 221 (1992).
16. E. K. Warrington and F. Clegg, Memory **1**, 281 (1993).
17. M. Hittmair-Delazer, G. Denes, C. Semenza, and M. C. Mantovan, Neuropsychologia **32**, 465 (1994).
18. E. K. Warrington, Brit. J. Psychol. **72**, 175 (1981).
19. L. K. Tyler, H. E. Moss, and F. Jennings, Neuropsychology **9**, 354 (1995).
20. E. K. Warrington and T. Shallice, Brain **107**, 829 (1984).
21. C. Sacchett and G. W. Humphreys, Cogn. Neuropsychol. **9**, 73 (1992).
22. E. DeRenzi and F. Lucchelli, Cortex **30**, 3 (1994).
23. M. Dennis, Brain Language **3**, 147 (1976).
24. K. Suzuki, A. Yamadori, and T. Fujii, Neurocase **3**, 193 (1997).
25. J. Hart, R. S. Berndt, and A. C. Caramazza, Nature **316v**, 439 (1985).
26. A. C. Caramazza and J. R. Shelton, J. Cogn. Neurosci. **10**, 1 (1998).
27. R. Fukatsu, T. Fujii, T. Tsukiura, et al., Neurology **52**, 1096 (1999).
28. D. M. Harris and J. Kay, Cortex **31**, 575 (1995).
29. P. Verstichel, L. Cohen, and G. Crochet, Neurocase **2**, 221 (1996).
30. M. Reinkemeier, H. J. Markowitsch, M. Rauch, and J. Kessler, Neuropsychologia **35**, 677 (1997).
31. T. Tsukiura, T. Fujii, R. Fukatsu, et al., Cogn. Neurosci. **14**, 922 (2002).
32. M. Spitzer, K. K. Kwong, W. Kennedy, and B. R. Rosen, Neuroreport **6**, 2109 (1995).
33. A. Martin, C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby, Nature **379**, 649 (1996).
34. E. Warburton, R. J. S. Wise, C. J. Price, et al., Brain **119**, 159 (1996).
35. H. Damasio, T. J. Grabowski, D. Tranel, et al., Nature **380**, 499 (1998).
36. H. J. Neville and D. Bavelier, Curr. Opin. Neurobiol. **8**, 254 (1998).
37. L. Standing, Quarterly J. Exp. Psychol. **25**, 207 (1973).
38. J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
39. D. J. Amit, *Modeling Brain Functions* (Cambridge Univ. Press, Cambridge, 1989).
40. W. S. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).
41. D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
42. R. Bogacz, M. W. Brown, and C. Giraud-Carrier, J. Comput. Neurosci. **10**, 5 (2001).
43. E. V. Budilova, M. P. Karpenko, L. M. Kachalova, and A. T. Terekhin, Biofizika **54**, 500 (2009).