

БЛОКИРОВАНИЕ ПАМЯТИ: ФАКТЫ, ПРОБЛЕМЫ, МОДЕЛИ

© 2009 г. Е.В. Будилова* **, М.П. Карпенко*, Л.М. Качалова*, А.Т. Терехин* **

**Институт когнитивной нейрологии Современной гуманитарной академии, 109029, Москва, Нижегородская ул., 32;*

***Биологический факультет Московского государственного университета им. М.В. Ломоносова, 119899, Москва, Воробьевы Горы*

E-mail: terekhin_a@mail.ru

Поступила в редакцию 24.04.07 г.

После доработки 11.08.08 г.

Исследован феномен блокирования памяти с точки зрения возможности построения нейросетевых моделей этого явления. Приведен обзор результатов психолингвистических и нейробиологических экспериментальных исследований, касающихся блокирования памяти. Сформулированы основные вопросы, решение которых способствовало бы объяснению этого явления. Одним из таких вопросов является наличие психологически парадоксального расхождения между субъективной уверенностью в том, что некоторый наблюдаемый образ знаком, и невозможностью воспроизвести его полностью. Для объяснения этого расхождения предложена биологически правдоподобная нейросетевая модель узнавания, демонстрирующая кардинальное превосходство в способности к узнаванию образа по сравнению со способностью к его воспроизведению.

Ключевые слова: блокирование памяти, сеть Хопфилда, функция энергии, функция знакомости, воспроизведение, узнавание.

Наверное, каждый из нас не раз оказывался в ситуации, когда не мог в нужный момент вспомнить какое-то слово, будучи при этом в полной уверенности, что оно не забыто и находится в полной сохранности «где-то в глубине памяти», что, действительно, потом подтверждалось тем, что через несколько часов или дней слово, которое мы никак не могли вспомнить (как правило, фамилию или географическое название), само «всплывало в памяти». Этот феномен, ярко описанный А.П. Чеховым в рассказе «Лошадиная фамилия» [1], по меньшей мере, с конца XIX века выделяется в научной литературе как особое психологическое состояние [2], называемое «состоянием на кончике языка» (англ. «tip-of-the-tongue state» – ср. с русским «вертится на языке»), или «блокированием» [3,4].

Исследование явления спонтанного блокирования памяти, представляя несомненный интерес для фундаментальной науки, важно и с прикладной точки зрения, поскольку может помочь разработать практические методики преодоления этого когнитивного дефекта. Потребность в таких методиках особенно возрастает в современную эпоху в связи с двумя ее характерными признаками. Первый – это постоянно растущая степень информатизации всех сторон

жизни общества, которая предъявляет повышенные требования к эффективному функционированию памяти людей, вовлеченных в активную деятельность. Второй – демографический сдвиг в сторону увеличения продолжительности жизни и, как следствие, увеличение числа пожилых людей, при том, что с возрастом повышается риск возникновения ситуаций блокирования памяти.

В данной работе проблема блокирования памяти исследована с точки зрения возможности построения математических (компьютерных) нейросетевых моделей этого явления. Сначала дан обзор результатов психолингвистических и нейробиологических экспериментальных исследований, касающихся блокирования памяти. Затем сформулированы основные вопросы, вытекающие из рассмотрения этих результатов, и предприняты попытки ответить на некоторые из них путем построения и анализа нейросетевых моделей процессов запоминания и воспроизведения.

ЭКСПЕРИМЕНТАЛЬНЫЕ ФАКТЫ, СВЯЗАННЫЕ С БЛОКИРОВАНИЕМ

Уже на уровне повседневного опыта выявляются две наиболее характерные особенности

блокирования памяти – повышение риска блокирования с увеличением возраста индивида и более частое блокирование собственных имен по сравнению с нарицательными. Совместное проявление этих особенностей приводит к тому, что учащение случаев блокирования при попытке воспроизведения собственных имен становится основной (и часто единственной) жалобой пожилых людей, что свидетельствует о возрастных изменениях их когнитивных способностей.

Эти закономерности, действительно, носят объективный характер и подтверждаются экспериментально. В работах [5–8] приводятся результаты экспериментов, доказывающих положительную зависимость частоты блокирования от возраста. В частности, в работах [5,6] показано, что когда испытуемым разного возраста предлагалось в течение месяца фиксировать все случаи блокирования слов, частота блокирования была выше у более пожилых людей. Кроме того, оказалось, что во всех возрастных группах основная доля блокирований приходилась на собственные имена – в основном на имена и фамилии знакомых, с которыми давно не было контакта. Блокированные слова, не являющиеся собственными именами, характеризовались низкой частотой встречаемости в языке. В работе [9] отмечено большее снижение способности к воспроизведению собственных имен при высотной гипоксии.

Убедительные подтверждения связи принадлежности слова к определенной семантической категории с возможностью его воспроизведения получены при наблюдении за пациентами с мозговыми нарушениями. Например, в случае, описанном в работе [10], у пациента не наблюдалось ухудшения памяти в отношении нарицательных имен, но он был почти полностью неспособен к воспроизведению любых собственных имен – имен и фамилий людей, кличек животных, названий известных зданий, брендов и т.д. Напротив, в случае, описанном в работах [11,12], пациенты были неспособны к воспроизведению обычных слов, но могли вспомнить имена известных людей и названия стран. В ряде случаев пациенты не могли воспроизвести имена и фамилии людей, но сохраняли способность к воспроизведению географических названий [13–17]. И, наоборот, известен случай, когда географические названия были единственной категорией слов, которые пациент мог вспомнить [16]. Нарушения памяти могут касаться только конкретных слов [18] или, наоборот, только абстрактных [19], только названий живых существ [20–22] или только названий предметов [21]. Известны случаи, когда нару-

шения памяти были связаны с еще более узкими категориями слов – названиями частей тела [23,24], фруктов и овощей [25], животных [26].

Наблюдения за пациентами с частичными мозговыми повреждениями позволяют также сделать некоторые выводы о связи нарушений воспроизведения различных категорий слов с определенными участками мозга. Так, в ряде работ сообщается об избирательном нарушении воспроизведения имен людей при локальных повреждениях левой височной доли [13–15,27–30]. В работе [23] описывается случай, когда резекция передней части левой височной доли привела к затруднениям в назывании пациентом частей тела.

В ряде нейробиологических исследований для выявления участков мозга, деятельность которых может быть связана с блокированием воспроизведения, были использованы современные неинвазивные технические средства регистрации его активности (fMRI, PET, ERP, MEG), которые могут быть применены для исследования работы мозга здоровых людей. В частности, в работе [31] с помощью функционального магнитно-резонансного исследования (fMRI) мозга было показано, что воспроизведение имен людей связано с активностью передней части левой височной доли мозга, причем при воспроизведении недавно запомненных имен активна также и правая височная доля. Анатомическая локализация активности мозга при воспроизведении различных категорий слов, таких как глаголы и существительные (а внутри существительных – более узкие семантические группы таких как инструменты, продукты питания, части тела и др.) подтверждается и многими другим инструментальными исследованиями [32–35].

Общая картина обработки лексической информации в мозге человека, получаемая на основе анализа результатов неинвазивных инструментальных наблюдений, может быть резюмирована следующим образом [36]. Во-первых, области, вовлеченные в лексическую деятельность, не являются однородными, а состоят из небольших изолированных пятен активности, связанных с различными языковыми компонентами (прежде всего, это касается осуществления доступа к фонетической информации через посредство средней височной извилины, локальная топографическая структура активности которой тесно связана с семантикой анализируемого слова). Во-вторых, эта активность не ограничивается только классическими лексически ориентированными областями (область Брока в нижней лобной доле, область Вернике на стыке верхней части височной и теменной долей

и угловая извилина), а распространяется также на верхнюю извилину левой височной доли, височный полюс, язычную и веретенообразную извилины, среднюю префронтальную область и островок. В-третьих, локализация лексической активности мозга более адекватно описывается в лингвистических терминах (фонология, синтаксис, семантика), нежели в терминах действий (слушание, произнесение, чтение).

ЗАГАДКИ БЛОКИРОВАНИЯ ПАМЯТИ

При анализе приведенных выше экспериментальных фактов возникает ряд вопросов, касающихся конкретных нейропсихологических проявлений блокирования памяти, ответ на которые желательно было бы получить на уровне выявления лежащих в основе этих проявлений базовых нейросетевых механизмов.

Во-первых, феномен блокирования ярко демонстрирует различие между знанием пытающегося воспроизвести некоторую информацию субъекта о том, что она содержится в его памяти, и невозможностью извлечь эту информацию. Каковы нейросетевые механизмы этого различия?

Во-вторых, степень блокирования связана с семантикой блокируемой информации. Какова нейросетевая основа связи семантики со степенью блокирования? В частности, в связи с этим возникает вопрос о том, каковы механизмы локализации семантически различных категорий информации в разных участках коры головного мозга?

В-третьих, часто блокированная информация через некоторое время воспроизводится. Как объяснить этот процесс в терминах работы нейронных сетей?

И, наконец, в-четвертых, почему неблагоприятные факторы (старение, гипоксия) усиливают в большей степени блокирование определенных семантических категорий слов, например, собственных имен? Какие изменения, специфически влияющие на блокирование, эти факторы вызывают в функционировании нейронных сетей мозга?

Здесь мы попробуем ответить на первый из заданных вопросов путем построения нейросетевой модели узнавания.

НЕЙРОСЕТЕВОЕ МОДЕЛИРОВАНИЕ УЗНАВАНИЯ И ВОСПРОИЗВЕДЕНИЯ

Резкий контраст между знанием субъекта о том, что некоторая информация содержится в его памяти и невозможностью извлечь эту информацию, ярко проявляющийся при блокиро-

вании памяти, может быть объяснен, по крайней мере частично, объективным различием между механизмом узнавания («familiarity recognition») и механизмом воспроизведения («recall», «recollection»). Это различие обнаруживается экспериментально. Так, в ряде психологических тестов после предъявления испытуемым нескольких тысяч изображений они могли почти безошибочно ответить на вопрос о том, было или не было им ранее показано следующее предъявляемое изображение, тогда как возможности детального воспроизведения были существенно ниже [37].

Нейросетевое моделирование процесса узнавания основано на наблюдении, что функция энергии сети Хопфилда [38] принимает большие значения для запомненных образов по сравнению с вновь предъявляемыми [39].

Сеть Хопфилда состоит из N нейронов, связанных друг с другом симметричными связями с весами $w_{ij} = w_{ji}$, причем предполагается, что $w_{ii} = 0$. Состояние нейрона i в момент времени t задается переменной состояния $x_i(t)$, принимающей значение -1 или 1 , а состояние всей сети – вектором $X_t = (x_1(t), x_2(t), \dots, x_N(t))$. Динамика сети определяется правилами изменения состояний ее нейронов и весов соединяющих их синаптических связей. Если в момент времени t веса сети $w_{ij}(t)$ и состояния $x_i(t)$ всех ее нейронов заданы, то в следующий момент $t+1$ состояние нейрона i изменится в соответствии с правилом Мак-Каллока и Питтса [40]:

$$x_i(t+1) = \text{sign} \left(\sum_{j=1}^N w_{ij}(t)x_j(t) \right), \quad (1)$$

где $\text{sign}(x)$ – функция, равная -1 для неположительных значений аргумента и 1 – для положительных. Что касается синаптических весов, то для всех $j \neq i$ они изменяются в соответствии с правилом Хебба [41]:

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

т.е. вес увеличивается на единицу, если нейроны i и j находятся в одинаковых состояниях, и уменьшается на единицу, если – в разных. Для $j = i$ вес устанавливается равным нулю. Функция энергии сети Хопфилда, характеризующейся вектором состояний $\{x_1(t), x_2(t), \dots, x_N(t)\}$ и набором весов $\{w_{ij}(t), i, j = 1, \dots, N\}$ задается выражением:

$$E(X_t) = -0,5 \sum_{i=1}^N \sum_{j=1}^N w_{ij}(t)x_i(t)x_j(t). \quad (3)$$

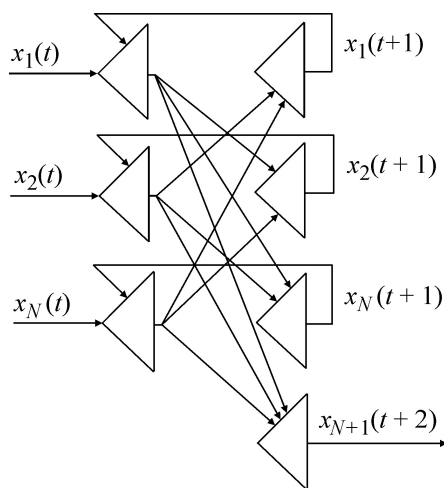


Схема нейронной сети для узнавания образов, основанная на вычислении значения знакомости, соответствующей образу, предъявленному для узнавания.

В работе [42] показано, что максимальное количество образов, которые могут быть предъявлены сети Хопфилда и затем с ошибкой менее 1% узнаны как знакомые на основании соответствующего им значения энергии (3), равно $0,023N^2$, что при больших значениях N на порядки выше, чем максимальное число образов, которые могут быть запомнены сетью с возможностью последующего полного воспроизведения – $0,145N$ [39]. В работе [42] описывается также ряд архитектур нейронных сетей, которые биологически правдоподобно могут произвести вычисления, близкие к требуемым формулой (3), без существенного снижения максимального числа образов, запоминаемых с возможностью последующего узнавания.

Мы предлагаем новую нейросетевую модель узнавания, также основанную на модификации формулы (3), но, на наш взгляд, более естественно согласующуюся с сетью Хопфилда. Модификация состоит в замене внутренней суммы в (3) ее знаком (мы отбрасываем также несущественный в данном контексте множитель – 0,5):

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

в результате чего соответствующее вычисление становится возможным осуществлять с помощью формулы (1) – базовой формулы, определяющей динамику состояний нейронов в сети Хопфилда. После применения формулы (1) формула (4) приобретает вид:

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

Из выражения (5) следует, что функция $E^*(t)$ – назовем ее «знакомостью» (ср. с англ. «familiarity») – равна просто скалярному произведению векторов состояний нейронов в два последующих временных шага. Максимальное значение $E^*(t)$ достигается в стационарных точках сети, характерных для запомненных образов, для которых выполняется равенство $x_i(t+1) = x_i(t)$, тогда как для новых образов, когда $x_i(t+1)$ фактически является случайным преобразованием $x_i(t)$, значение $E^*(t)$ близко к нулю. Для вычисления скалярного произведения (5) можно также воспользоваться формулой (1), с помощью которой фактически вычисляется скалярное произведение вектора $\{w_{ij}(t), j = 1, \dots, N\}$ весов связей от каждого нейрона к нейрону i на вектор $\{x_1(t), x_2(t), \dots, x_N(t)\}$. Нужно просто заменить вектор $\{w_{ij}(t), j = 1, \dots, N\}$ на $\{w_{ij}(t+1), j = 1, \dots, N\} = \{x_1(t), x_2(t), \dots, x_N(t)\}$, а $\{x_1(t), x_2(t), \dots, x_N(t)\}$ – на $\{x_1(t+1), x_2(t+1), \dots, x_N(t+1)\}$. Это можно осуществить введением дополнительного «нейрона узнавания» $N+1$, веса к которому $\{w_{N+1,j}(t+1), j = 1, \dots, N\}$ на шаге t в соответствии с формулой (2) полагаются равными вектору $\{x_1(t), x_2(t), \dots, x_N(t)\}$, если предположить, что более ранние веса $\{w_{N+1,j}(t), j = 1, \dots, N\}$ «забыты» и равны нулю, а нейрон $N+1$ находится в активном состоянии. Тогда на шаге $t+1$ в соответствии с формулой (1) осуществляется скалярное умножение этих весов на входной вектор $\{x_1(t+1), x_2(t+1), \dots, x_N(t+1)\}$ и взятие знака от результата

$$x_{N+1}(t+2) = \operatorname{sign} \left(\sum_{j=1}^N w_{N+1,j}(t+1) x_j(t+1) \right). \quad (6)$$

Взятие знака в формуле (6) завершает узнавание, приводя $x_{N+1}(t+1)$ в активное состояние $x_{N+1}(t+1) = +1$ в случае знакомого входного образа $\{x_1(t), x_2(t), \dots, x_N(t)\}$ и в неактивное $x_{N+1}(t+1) = -1$ – для незнакомого. Преимуществом рассмотренной схемы по сравнению с предложенными в работе [42] является то, что вычисления на шаге t производятся не специальной дополнительной сетью, а основной сетью Хопфилда, а дополнительным является лишь конечный узнающий нейрон.

На рисунке приведена схема описанной сети.

Можно показать, что емкость памяти нейронной сети, основанной на использовании

функции знакомости $E^*(t)$, равна $0,0185N^2$, т.е. снижается незначительно по сравнению с значением $0,023N^2$ [42], получаемым при использовании для узнавания функции энергии $E(t)$. Таким образом, квадратичная зависимость числа узнаваемых образов от числа нейронов может быть получена с помощью более простого нейросетевого механизма, что является доводом в пользу того, что такой механизм в каком-то виде может быть реализован в нейросетевой структуре мозга и что именно он ответственен за парадоксальный контраст между узнаванием и воспроизведением, возникающий при блокировании памяти.

Детали вычисления емкости сети, осуществляющей узнавание на основании вычисления знакомости, задаваемой формулой (5), приведены в работе [43].

СПИСОК ЛИТЕРАТУРЫ

1. А. П. Чехов, *Лошадиная фамилия* (Петербургская газета, СПб., 1885).
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 and Language* **30**, 542 (1991).
6. M. K. Heine, B. A. Ober, and G. K. Shenaut, *Psychology and Aging* **14**, 445.
7. M. Evrard, *Brain and Language* **81**, 174 (1999).
8. D. M. Burke and M. A. Shafto, *Current Directions in Psychological Science* **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. of Neurology, Neurosurgery, and 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, *British J. of 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, *Cognitive Neuropsychology* **9**, 73 (1992).
22. E. DeRenzi and F. Lucchelli, *Cortex* **30**, 3 (1994).
23. M. Dennis, *Brain and 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. Cognitive 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, *Current Opinion in Neurobiology* **8**, 254 (1998).
37. L. Standing, *Quarterly J. Exp. Psychology* **25**, 207 (1973).
38. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
39. D. J. Amit, *Modeling Brain Functions* (MA: 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. Computat. Neurosci.* **10**, 5 (2001).
43. Е. В. Будилова, М. П. Карпенко, Л. М. Качалова и А. Т. Терехин, *Биофизика* (в печати).

Memory Blocking: Facts, Problems, and Models

E.V. Budilova* **, M.P. Karpenko*, L.M. Kachalova*, and A.T. Teriokhin* **

**Institute of Cognitive Neurology, Modern University for Humanities, Nizhegorodskaya ul. 32, Moscow, 109029 Russia*

***Department of Biology, Lomonosov Moscow State University, Vorob'evy Gory, Moscow, 119899 Russia*

The tip-of-the-tongue state, or memory blocking, has been investigated from the point of view of possibilities of its neural network modeling. The results of neuropsychological and neurobiological studies on memory blocking have been reviewed, and basic problems whose solution could contribute to the comprehension of this phenomenon have been formulated. One of such questions is the existence of a paradoxical discrepancy between the subjective assurance that a pattern is familiar and the impossibility to recollect it fully. To explain this discrepancy, we propose a biologically plausible neural network model of recognition demonstrating a radical superiority of the capacity for the recognition of a pattern as compared with its recollection.

Key words: memory blocking, Hopfield network, energy function, familiarity recognition