

Анализ роли понятий «образ» и «символ» в моделировании процесса мышления средствами нейрокомпьютинга

**О.Д. Чернавская, Д.С. Чернавский, В.П. Карп, А.П. Никитин,
Я.А. Рожило**

Аннотация

Обсуждаются понятия «образ» и «символ», а также их функции в *мыслительной системе*, состоящей из связанных нейропроцессоров. Показано, что образная подсистема играет ведущую роль в записи и хранении информации. Введение символьной подсистемы обеспечивает переход к условной (*семантической*) информации, что в итоге позволяет осуществлять коммуникацию с окружающей средой. Парадигма внимания в рассматриваемом подходе реализуется за счет параметрического влияния символьной подсистемы на образную. Показано, что эффект специализации нейронов естественно воспроизводится за счет самоорганизации системы. Предлагается система нелинейных динамических уравнений, связывающих символьную и образную подсистемы, которая потенциально способна описать «ход мысли» в индивидуальной мыслящей системе. Предлагается возможная интерпретация понятий интуиции, подсознания, логического мышления в терминах функций образных и символьных нейронов.

On the role of “pattern” and “symbol” concepts for simulation of the thinking process via neurocomputing

O.D. Chernavskaya, D.S. Chernavskii, V.P. Karp, A.P. Nikitin, Y.A. Rozhilo

Abstract

The concepts of “pattern” and “symbol” and their functions are discussed in the context of “thinking” system of combined neuroprocessors. It is shown that the pattern subsystem does play a key role in *recording* and *storage* of information. The *symbol* subsystem initiation provides the transition to conventional *semantic* information and communications with environment. The paradigm of *attention* within the scheme presented is secured by a *parametric effect* of symbol subsystem on the pattern one. It is shown that the effect of neuron specialization is reproduced due to self-organization of the whole system. The system of dynamical nonlinear equations is proposed to combine the pattern and symbol subsystems and thus, to describe the “*train of thought*” within an individual thinking system. It is shown that the concepts of intuition and logic could be interpreted in terms of specific functions of pattern and symbol neurons.

1. Введение

Данная работа продолжает активные исследования авторов [1-3] в области моделирования процесса мышления средствами нейрокомпьютинга, в частности, в контексте Динамической Теории Информации (ДТИ) [4]. Цель настоящей статьи — обсудить понятия *образ* и *символ*, их формирование, взаимодействие, а также их роль в процессе мышления.

Несмотря на многолетнюю популярность исследований процесса мышления, устоявшегося, общепринятого определения самого мышления не существует. В работе [3] было предложено следующее определение: *мышление есть самоорганизующийся процесс восприятия, обработки, сохранения, а также генерации и распространения информации без постороннего вмешательства*. Имеется в виду, что естественный обмен информацией с внешней средой, разумеется, присутствует, но запись и обработка информации производится самой *мыслительной системой*, которая формируется в процессе *самоорганизации*. Поскольку речь идет о моделировании процесса мышления средствами нейрокомпьютинга, под «мыслительной системой» мы будем понимать *систему связанных нейропроцессоров*, различающихся характером связей между формальными нейронами.

Ключевыми понятиями в моделировании процесса мышления являются *образ* и *символ*. Их смысл интуитивно ясен, однако заслуживает уточнения.

Термин *восприятие* обычно подразумевает участие *рецепторных* нейронов (*reception* = получение, принятие) и, в частности, периферической нервной системы. Однако принципы работы и, соответственно, моделирования, рецепторных, а также аффлекторных систем, с одной стороны, и «*внутренних*» нейронов, с другой, слишком сильно разнятся, чтобы пытаться в рамках единой модели воспроизвести все процессы (по сути «объять необъятное»). В наших работах мы всегда ограничиваемся анализом принципов работы *только «внутренних» нейронов*, и в этом контексте восприятие трактуется как *запись информации*, полученной от рецепторных нейронов. Такая трактовка неявно подразумевает *кодирование* этой информации, т.е. выбор средств записи, *удобных* для данной системы.

Проблема кодирования применительно к нейропроцессорам имеет три уровня.

1). *Первичное кодирование* — запись информации в виде «*цепочки*» *активированных нейронов*, составляющих *образ объекта*. Такое представление практически взято из физиологии: в результате взгляда на реальный объект возбуждается некоторый набор «внутренних» нейронов; при повторном предъявлении того же объекта возбуждаются те же самые нейроны. Механизмы этого процесса, т.е. передачи информации от рецепторных нейронов к «внутренним», обсуждаются в специальной литературе; этой проблемы мы не

касаемся. Далее будем считать, что реальный объект активирует некоторую определенную цепочку нейронов, которую и будем называть *образом* объекта.

2). **Внутреннее кодирование** — сопоставление образу объекта (явления, ситуации) его *символа*, иными словами, *локализация образа*. Проблема сопоставления некоему объему информации единственного символа существует во многих областях деятельности. В обычных компьютерах это происходит путем присвоения извне данному объекту его *названия* (например, оператор задает имя файла). Нас, однако, интересует возможность смоделировать процесс формирования символа *без постороннего вмешательства*.

3). Образование *общего кода* (например, языка). Символьная информация становится особенно конструктивной, если она является *условной*, т.е. одни и те же символы (слова, понятия и т.д.) имеют *одинаковый смысл* внутри какого-либо сообщества. Механизм образования общего кода является в большой степени *социальной* проблемой, что выходит за рамки исследования; социальный аспект обсуждался в литературе, где рассматривались различные модели этого процесса (см. [4]). Здесь мы будем предполагать, что символы, сформировавшиеся в конкретной системе, *потенциально* являются *словами*, т.е. средством коммуникации с окружающей средой.

В работе [3] подробно обсуждался вопрос о том, *что* должна делать мыслительная система (или «аппарат мышления»). Здесь мы подробно обсудим, *как* могут выполняться некоторые из необходимых функций, а именно — *восприятия* (записи) и *сохранения* (запоминания) информации; функция *обработки* информации исследуется на предмет *роли образов и символов* в ее исполнении.

2. Нейропроцессоры

Нейропроцессором будем называть пластину, населенную *формальными нейронами* общим числом n . Понятие *пластина* вводится, во-первых, исходя из «традиционной» идеологии нейрокомпьютинга и того, что именно так работают нейропроцессоры, используемые в практических целях. Во-вторых, как показано ниже, такое представление удобно для *структурирования* информации. При этом равно важную роль играют как *внутрипластинные*, так и *межпластинные связи*. Число N связей (каждого типа) оценивается как $1 \ll N \ll n \cdot (n-1)$, т.е. принимается, что каждый из n нейронов пластины связан со многими, но *не со всеми* нейронами как на «своей», так и на других пластинах.

2.1. Формальный нейрон как понятие заслуживает специального обсуждения. Изначально под формальным нейроном понимали бистабильный элемент, который может существовать *стационарно* в активном (+1), либо в пассивном (-1 или 0) состояниях. Следствием такого подхода было резко отрица-

тельное восприятие нейрокомпьютинга физиологами. Реальный нейрон — гораздо более сложное «устройство», «организм в организме», плоды жизнедеятельности которого (метаболиты) влияют как на мозг, так и на весь организм; роль нейрона отнюдь не сводится к «рефлекторному» ответу на внешнее воздействие [5].

Для учета реальных особенностей нейрона была предложена более сложная модель Фицхью-Нагумо [6], а также ряд других (см. [7] и ссылки *ibid*), описывающих накопление потенциала, фазы возбуждения и последующей релаксации, латентную фазу и т.д. Однако, по сути, эти модели не слишком сильно отличаются от «оригинального» формального нейрона, если понимать последний не упрощенно, а рассматривать как объект, описываемый *нелинейным динамическим уравнением*, и учитывать не только стационарные состояния, но и всю динамику процесса. Более того, начиная исследование с достаточно простого объекта, мы далее можем его усложнять (что планируется продолжать и в дальнейшем), вводя параметры, описывающие *внутреннее состояние*. Так или иначе, важнейшим свойством реального нейрона является накопление мембранного потенциала, т.е. суммирование воздействий от связанных с ним нейронов как той же пластины, так и других пластин системы — какое-то свойство вполне может быть отражено через уравнения, описывающие формальный нейрон.

2.2. Процессор Хопфилда¹ — линейная аддитивная сеть, описываемая моделью:

$$\frac{dH_i(t)}{dt} = -\frac{1}{\tau_i^{H_i}} [H_i + \beta_i (H_i^2 - 1) - H_i^3] + \sum_{j \neq i}^N \Omega_{ij} H_j + Z(t) \xi(t) \quad (1)$$

где $H_i(t)$ — переменная, описывающая состояние нейрона; τ_i^H — характерное время активации нейронов данного типа; β_i — параметр, связанный с порогом возбуждения нейрона; Ω_{ij} — матрица связей между нейронами; $i, j = 1 \dots n$. Последний член в (1) описывает случайное воздействие (*шум*), где $Z(t)$ — амплитуда, а $\xi(t)$ — случайная функция. Наличие шума обеспечивает спонтанную активацию не связанных друг с другом цепочек-образов, что позволяет интенсивно использовать практически всю информацию, записанную на пластине; в определенном смысле это отвечает *параллельному способу* обработки информации (см. [2]). Процессор с шумом называют еще процессором Хопфилда-Больцмана [10].

Связи Ω_{ij} модифицируются при *обучении* процессора, т.е. при *активации*

¹ Этот процессор предложен Дж. Хопфилдом в [8]. Однако, в литературе известны высказывания (см. [7]), что аналогичные идеи предлагались ранее другими авторами (прежде всего, в [9]). Однако название процессора устоялось, невзирая на упомянутую полемику.

образов. Это свойство играет важнейшую роль и будет подробно обсуждаться ниже. Еще одно ключевое свойство процессора Хопфилда — *ассоциативность*. Цепочки, имеющие *общие нейроны*, связаны, так что возбуждение одной влечет за собой возбуждение и другой. Отметим еще один важный терминологический момент: фазовое пространство уравнения (1) является одновременно (и по сути) «пространством признаков» на языке теории распознавания.

2.3. Процессор локализации (кодирования) обычно связывают с именем С. Гроссберга [11], однако в его работах было исследовано множество вариантов процессоров с нелинейным взаимодействием, так что выделить один определенный затруднительно. Мы будем называть *процессором Гроссберга* простой вариант процессора, обеспечивающего *локализацию*, т.е. преобразование какой-либо группы из нескольких возбужденных нейронов в один. Его математическое выражение [12]:

$$\frac{dG_k(t)}{dt} = \frac{1}{\tau_k} \{ -(\alpha_k - 1)G_k + \alpha_k G_k^2 - G_k^3 \} - \sum_{k \neq l} \Gamma_{kl}(t) G_k G_l \quad (2)$$

где $G_k(t)$ — переменная, отражающая состояние нейрона, τ_k — характерное время активации; α_k — параметр, описывающий его внутреннее состояние. «Нейрон Гроссберга» обладает свойством *самовозбуждения* и имеет активное (+1) и пассивное (0) стационарные состояния; взаимодействие таких нейронов нелинейное и *подавляющее* (благодаря знаку «−»). Связи Γ_{kj} обеспечивают взаимодействие *внутри* пластины.

В начальном, необученном состоянии все связи считаются равными: $\Gamma_{kl}(0) = \Gamma_{lk}(0) = \Gamma_0$. В процессе обучения связи изменяются (τ_Γ — характерное время обучения связей Γ):

$$\frac{d\Gamma_{kl}(t)}{dt} \propto \frac{\Gamma_0}{\tau_\Gamma} \{ G_k \cdot G_l (G_k - G_l) \}. \quad (3)$$

Отметим, что в силу выбранных стационарных состояний $\{0,1\}$ в процессе обучения связей участвуют только *активные* нейроны. Это означает, что при активации некоторого конкретного набора M нейронов Гроссберга ($M \ll N$), именно эти M нейронов сливаются в один из них, *выбранный случайно*. Именно этот нейрон становится *символом* данного набора.

3. Образ: Запись информации и запоминание.

Само понятие *информация* имеет несколько определений. Мы будем использовать то, которое было предложено Г. Кастлером: *информация есть запомненный выбор одного варианта из множества возможных и равноправных*. Это определение не только не противоречит распространенным в научной литературе дефинициям термина, но и, в отличие от них, дает содер-

жательное представление о том, *как* информация возникает. А именно, необходимо *сделать* (т.е. *записать*) *выбор* и *запомнить* его.

В нейροкомпьютинге обе задачи могут быть решены при помощи процессора Хопфилда. Однако важно подчеркнуть (и это подробно обсуждалось в [2, 3]), что эти задачи **дуальны**, и, следовательно, должны реализовываться в **двух разных подсистемах**.

3.1. Первичная запись: формирование образов

При обучении нейροпроцессора информация *возникает*: внешнее воздействие специфическим образом преобразуется и превращается в *индивидуальную информацию*. Это — самый первый и основной *акт генерации информации* (творчества). Принцип преобразования и закон обучения напрямую связаны с целевой функцией процессора.

Задача записи информации в системе нейροпроцессоров, как и все прочие, должна решаться посредством самоорганизации. Иными словами, следует постулировать некие основополагающие принципы и далее исследовать, как они могут быть реализованы. Указанные принципы должны быть, по возможности, близки к реальному, человеческому процессу мышления, поэтому для проверки их адекватности вполне уместно привлекать житейские наблюдения, аналогии и фольклор.

Здесь будем использовать два основных принципа записи информации.

- Чем *чаще* активируется конкретный образ, тем **сильнее** связи между нейронами, ему соответствующими (согласно общеизвестному «повторенье — мать ученья»).

- Одинаковые (или подобные) элементы разных объектов возбуждают **одни и те же** нейроны (аналогия: правило экономии записывающего материала, минимизация объема носителей информации). Тогда сходные в чем-то образы оказываются «сцепленными» общими нейронами, и, следовательно, возбуждение одной из цепочек, отвечающей конкретному образу, может повлечь за собой активацию и другого образа, ассоциативно связанного с ним.

Оба положения естественно удовлетворяются именно в процессоре Хопфилда, причем свойство ассоциативности (и так имманентно присущее этому типу процессора) приобретает содержательный смысл: ассоциация как артефакт становится ассоциацией в обычном понимании, т.е. косвенной семантической связью.

Первый принцип выполняется, если обучение связей описывается уравнением:

$$\frac{d\Omega_{ij}(t)}{dt} \sim \frac{\Omega_0}{\tau_\Omega} (H_i(t) + 1)(H_j(t) + 1)\Phi(t) \quad (4)$$

где Ω_0 и τ_Ω — параметры, характеризующие процесс обучения, $\Phi(t)$ — монотонная интегрируемая функция, обладающая эффектом насыщения (по-

сколькучу обучение не безгранично). Отметим, что согласно (4) обучаются связи только между *активными* нейронами, остальные участия не принимают.

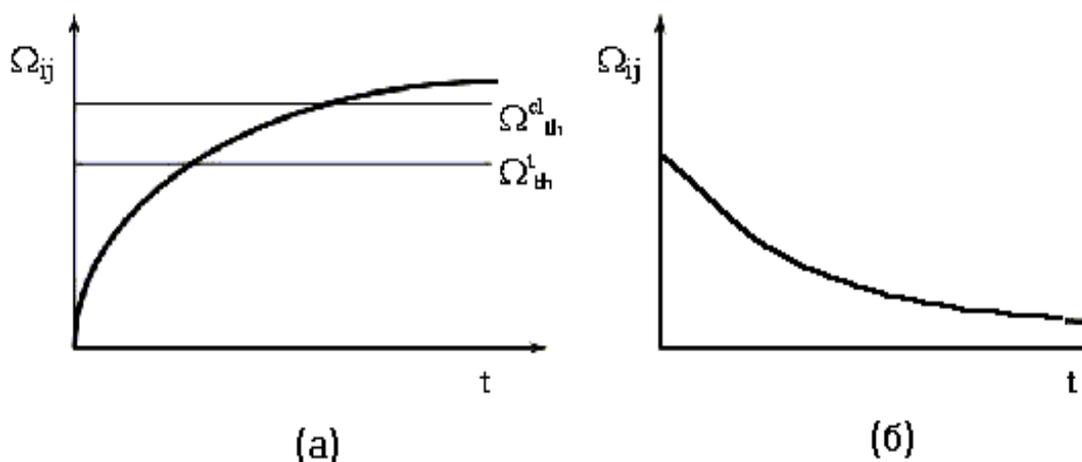


Рис. 1. Зависимость силы связи от времени и/или активности обучения для (а) «альтернативного» варианта обучения и (б) обучения «по Хопфилду».

Таким образом, информация об объекте будет записана в форме его *образа*, а еще точнее — в форме *обученных связей* между нейронами. Итак, задача *записи информации* может быть решена при помощи процессора Хопфилда, обученного согласно (4).

3.2. Запоминание и хранение важной информации

Проблема *запоминания*, казалось бы, уже решена при первичной записи. Однако смысл запоминания подразумевает, что оно должно быть селективно, а именно — из всего записанного надо отобрать лишь нужное для субъекта, иначе память будет переполнена. Иными словами, должна быть решена еще и *проблема забывания*.

Учитывая принципы самоорганизации, можно предложить несколько способов забывания, имеющих право на существование и, вероятно, реализующихся в разных системах.

Один из путей — осуществить вариант «*угасания неактуального*»: связи, не используемые в течение долгого времени, «впадают в спячку» по некоторому (экспоненциальному) закону. При повторном обращении эти связи могут восстанавливаться на прежнем уровне.

Другой путь («*отбрасывание ненужного*») может быть реализован в пластине Хопфилда, обученной по-другому закону, как и было предложено самим Дж. Хопфилдом. Пусть (см. рис. 1б) внутрисластинные связи *изначально велики и одинаковы*. Обучение заключается в том, что связи между нейронами, возбужденными при предъявлении данного объекта, и всеми *другими* нейронами пластины *слабеют*. Таким образом, избыточные связи «отбрасываются»,

реализуется фильтр на отсечение лишнего, а запоминается *только один вариант* образа объекта, остальные варианты *забываются*.

Закон обучения связей в этом случае имеет вид:

$$\Omega_{ij}(t) = \Omega_0 \left\{ 1 - \frac{1}{2T_0} \int_0^t [1 - H_i(t')H_j(t')] \zeta(t') dt' \right\} \quad (5)$$

$$\int_0^{\infty} \zeta(t') dt' \leq 1$$

Таким образом, проблема записи и хранения информации может быть решена при помощи *двух последовательных процессоров* хопфилдовского типа (см. рис. 2). Далее первую пластину будем называть пластиной «*первичных*» (или «*размытых*») *образов* (H_0), вторую — пластиной «*типичных образов*» ($H_{тип}$).

Процесс записи и селекции можно представить следующим образом. Пусть на пластину H_0 записываются *все образы*, даже встреченные однократно. Образ конкретного объекта предстает в *разных вариантах* и потому записи этих вариантов слегка отличаются друг от друга (т.е. каждый раз активируются близкие, но не одни и те же нейроны). Связи между нейронами, вообще говоря, не слишком сильны (такие связи мы будем называть «*серыми*»). В результате возникает *пятно*, внутри которого, однако, просматривается контур, составленный из нейронов, задействованных наиболее часто, так что связи именно между ними оказываются *сильнее остальных* (такие связи будем называть «*черными*»).

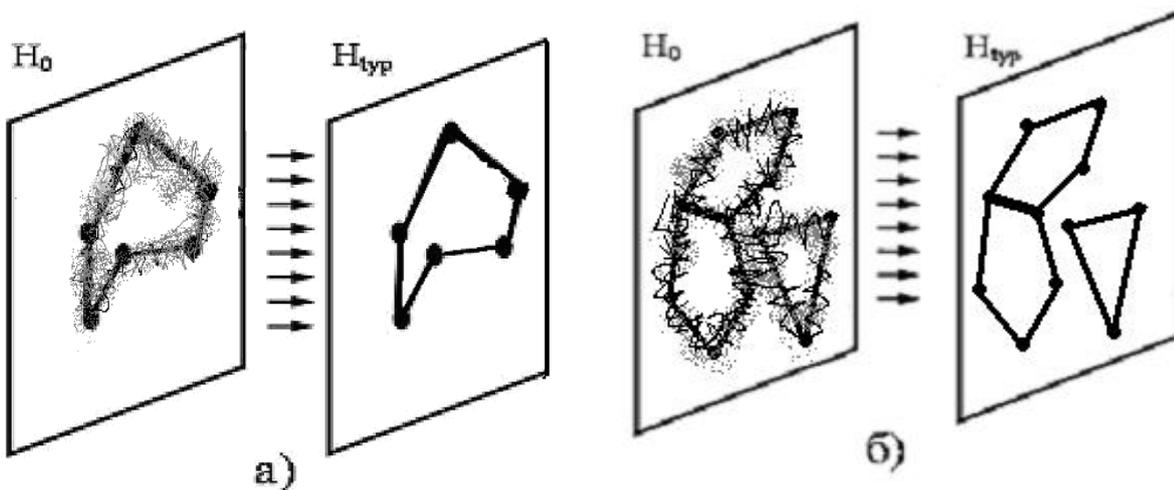


Рис. 2. Схема перехода первичного (размытого) образа в типичный образ (отобранный для запоминания) в случае изолированного образа (а) и набора образов (б)

Можно предложить следующий алгоритм *запоминания нужного* (см. рис. 2). Если величина силы связей нейронов в контуре, представляющем наиболее активную репрезентацию данного объекта, превышает некоторый *порог* Ω_{th}^{typ} , данная цепочка *копируется* (передается прямыми межпластинными связями) на другую пластину. Вторая пластина N_{typ} также хопфилдовского типа, но обучена согласно (5) и содержит только *типичные* образы, слабые *серые* связи отсекаются на этапе передачи. Таким образом, вторая пластина работает как фильтр для отсека ненужного. Связи между нейронами цепочек на пластине N_{typ} *apriori* сильнее, чем связи во всех первичных образах на пластине N_0 , кроме того, который был выбран на роль типичного.

Подчеркнем, что часть информации при такой процедуре теряется, причем может теряться и *ассоциативность* образов (на рис. 2б слева два нижних образа имеют общие серые связи; справа общих связей между ними нет). Это не значит, что такая информация не нужна вообще, просто в данный момент записи она оказалась *не самой актуальной*. Однако в реальной мыслительной системе она должна оставаться доступной на пластине N_0 , что еще раз подтверждает необходимость наличия двух подсистем.

Отметим, что образы на пластине N_{typ} возникают *позже* и *в результате* (с некоторой задержкой) формирования цепочек на первой пластине. Иными словами, на этапе формирования всей мыслительной системы эта пластина (N_{typ}) оказывается ведомой, а пластина «размытых» образов N_0 — ведущей.

Итак, задачи и *записи*, и *запоминания* могут быть, в принципе, решены посредством *только образов* — двух пластин Хопфилда — без привлечения дополнительных процессоров. Переход *образ* → *типичный образ* сопровождается потерей части информации и фактически представляет собой ее *кодирование* (по крайней мере, сжатие). Дальнейшее структурирование памяти, т.е. выделение подмножеств из множества типичных образов, необходимо и возможно, однако для этого нужна *идея*, или *принцип* такой структуры. Иначе говоря, образы в хранилище памяти должны группироваться «*со смыслом*».

Приведем аналогию с хранением информации в персональных компьютерах. Если *образ* ассоциировать с *файлом*, то при небольшом их количестве используется одна папка-директория. Если же файлов много, то возникает путаница, поиск нужного затягивается, поэтому разумно выделить группы файлов, связанных неким *общим смыслом*, и распределить их по нескольким директориям. Как именно строить структуру папок для хранения и распределять по ним файлы решает *сам пользователь*, сообразуясь с удобством и только. Такой выбор *условен* (не продиктован объективными законами) и *индивидуален*, т.е. являет собой пример *самоорганизации пользователя*.

Здесь впервые возникает потребность в *семантической* (смысловой), *условной* информации, первым шагом к которой является *символ*.

4. Символ: запись и хранение семантической информации.

Согласно ДТИ, весьма важную роль играет *условная информация* — «выбор, возникший в коллективе (обществе) в результате договоренности» [4]. Однако роль «общества» может играть и сама система («коллектив нейронов»), если под «договоренностью» понимать самоорганизацию.

Когда возникает потребность и возможность в формировании символа в рассматриваемой системе? Оговоримся сразу: этот процесс, как и любой акт самоорганизации, строго индивидуален, причем имеется в виду не вмешательство «бессмертной души», а широкий спектр возможностей, имеющих право на реализацию. В частности, символы могут не образовываться вообще. Однако наиболее естественно предположить (следуя аналогии с распределением файлов), что сигналом к формированию символов будет большое количество записанных образов и их «перемешивание».

4.1. Формирование символа изолированного образа.

Рассмотрим сначала наиболее простой процесс — формирование символа *изолированного* типичного образа.

Пусть обучение носит декларативный характер («обучение с учителем»): некий объект предъявляется многократно и изолированно (см. рис. 2а). При этом *типичный* образ формируется относительно быстро (но и в этом случае запись образа объекта может и должна иметь несколько вариантов) и уже *осознанно*: «учитель» фактически сообщает системе, что данный типичный образ отвечает одному реальному объекту и, соответственно, одному символу.

Тогда формирование символа происходит следующим образом (см. рис. 3а). Образ с пластины $H_{\text{тип}}$ передается (параллельным переносом, прямыми межпластинными связями) на пластину локализации $G_{\text{тип}}$, где все нейроны кроме одного, выбранного случайно, деактивируются, а оставшийся активным нейрон становится *символом данного образа*. В результате этого происходит *первичное обучение связей* G_{ij} , (см. [12]), так что при вторичной активации тех же нейронов произойдет быстрая (за время, меньшее, чем характерное время первичного обучения) *свертка* в тот же самый нейрон — *символ* образа. Далее межпластинные связи нейронов-участников данного образа *перезамыкаются* именно на нейрон-победитель (рис. 3а, нижний блок).

Процесс обучения межпластинных связей символ \leftrightarrow образ, Ψ , может быть описан уравнением, аналогичным тому, которое принималось для записи информации (4), где в обучении участвуют только активные нейроны:

$$\frac{d\Psi_{ik}(t)}{dt} \propto \frac{1}{\tau_{\Psi}} G_k (H_i + 1) \quad (6).$$

После установления таких связей активность нейронов разных типов («образных», т.е. нейронов типа Хопфилда, и «символьных», т.е. нейронов типа Гроссберга) оказывается взаимосвязанной.

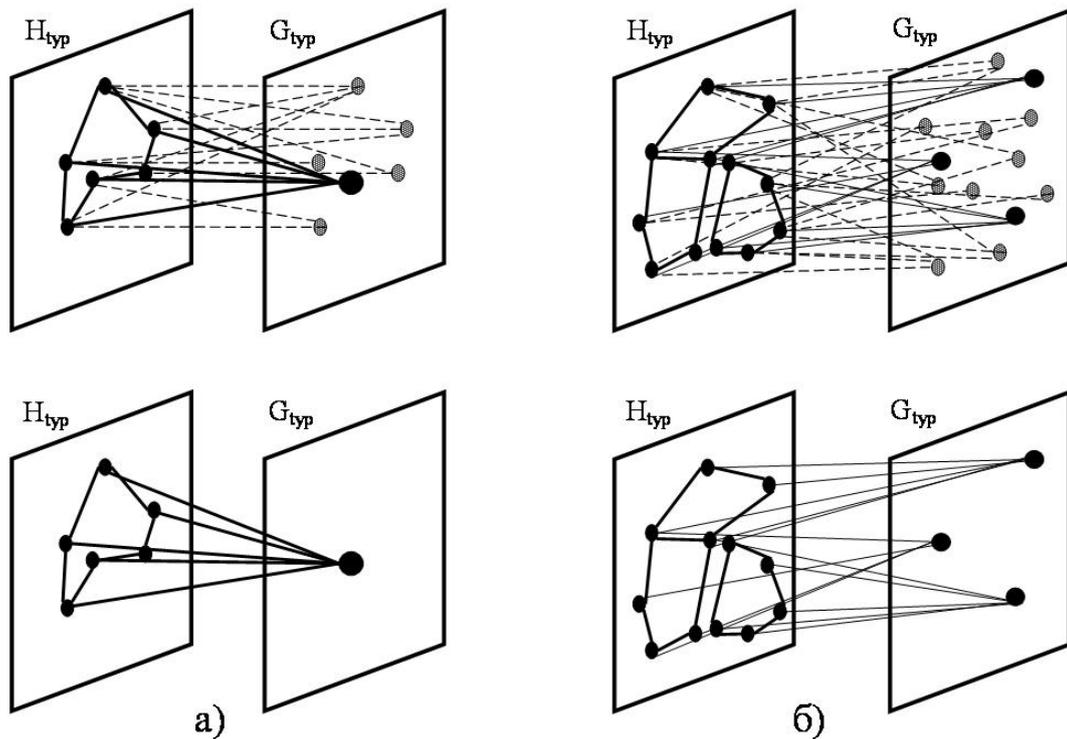


Рис. 3. Схема формирования символа образа при
 (а) «обучении с учителем» (изолированный образ) и
 (б) «обучении без учителя» (одновременный анализ нескольких образов)

Для чистоты процедуры *может быть* предусмотрена **проверка**, т.е. *декомпозиция* (развертка) образа из символа и *сравнение* его с исходным типичным образом (такая процедура обсуждалась в [12]). Однако сама необходимость проверки требует пояснения. В рамках предложенной схемы и формальных нейронов Гроссберга случайность выбора — *абсолютна*. Это само по себе красиво: описанный процесс прямо указывает на *генерацию информации*, т.е. акт сугубо творческий. Более того, выбор, сделанный таким образом, не требует *доводов и доказательств*, это и есть та самая *своя условная информация*, защищать которую — основная цель мыслящего организма (см. [3]). В этой связи процедура проверки, обсуждаемая в [12], казалось бы, смысла не имеет.

С другой стороны, если пытаться соотнести этот воображаемый процесс с реальностью, то механизм перезамыкания связей не столь однозначен. Как уже говорилось, мы принимаем изначально, что число межпластинных связей N велико, но меньше, чем $n \cdot (n-1)$, что отвечало бы ситуации «все связаны со всеми». Поэтому может оказаться, что у нейрона-«победителя», выбранного механизмом конкуренции, действующим согласно уравнению (2), нет одной или нескольких нужных связей с нейронами-«прародителями» данного образа. Тогда развертка даст образ, отличающийся от типичного.

Сказанное хорошо соотносится и с представлениями о реальном нейроне

как о «сложном организме» [5], в рамках которых процесс установления связей («настройка синапсов») отнюдь не тривиален. Связи строятся так, как нейронам «удобно», и процесс настройки регулируется не уравнением (2), а метаболитами. Как следствие, далеко не каждый нейрон, выбранный на роль символа, сможет установить «контакт» со всеми нейронами своего образа.

Универсальный механизм, который может действовать как в воображаемых системах, так и в реальных — метод «*попытка + проверка*» (или «проб и ошибок»). Можно предложить два способа описания этой процедуры.

- Образование символа только за счет *конкуренции на пластине Гроссберга с обязательной последующей проверкой*. Одна из возможных схем формирования символа с проверкой представлена на рис. 3с.

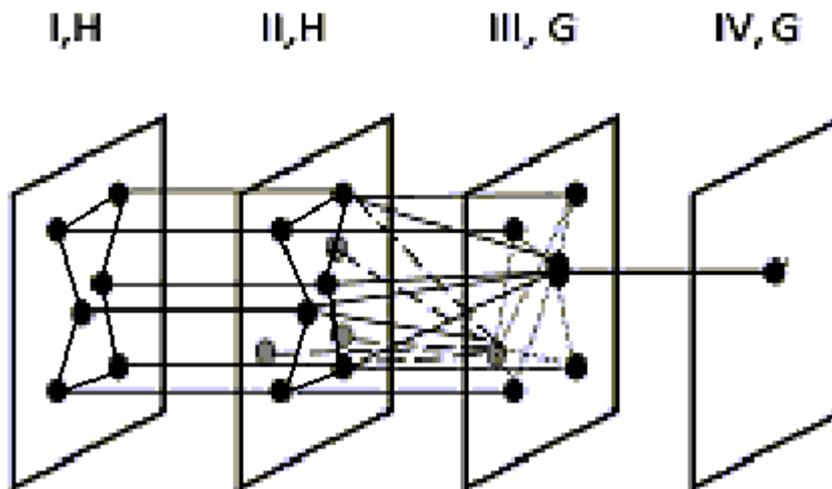


Рис. 3 с. Схема формирования символа изолированного образа с проверкой.

Здесь предусмотрена специальная вставка двух дополнительных пластин, часто называемая (см. например [1]) комплексом *Back Propagation (BP)*. Пластина II, H предназначена именно для декомпозиции символа, сформированного первоначально на «пробной» пластине III, G типа Гроссберга. Если результат декомпозиции недостаточно *похож* на исходный типичный образ, т.е. активны не только «нужные» нейроны, но и лишние, или, напротив, не активны нужные, возникает значительная *невязка* (определяемая стандартным образом через, например, метод наименьших квадратов), символ разрушается и процедура повторяется заново. Если достигнут положительный результат, то символ-победитель передается на конечную пластину IV, G и запоминается там.

- Локализация образа за счет *одновременной конкуренции межпластинных связей и символов*. На языке нейрокомпьютинга это означает, что процесс *настройки связей* должен происходить *одновременно* (или даже *раньше*), чем конкурентная борьба в процессоре локализации, а побеждать должен не совсем *случайный* нейрон, а тот, который способен обеспечить полноту об-

раза. При этом механизм выбора нейрона-победителя, обсуждавшийся выше (и подробно обсуждавшийся в [12]), должен быть модифицирован: фактор *случайности* в самом начале обучения пластины Гроссберга следует заменить на *преимущество* (возможно минимальное) того нейрона Гроссберга, который наиболее «удобен» для связи с данным образом. Дальнейшая конкурентная борьба идет так, как было описано, и окончательно закрепляет положение выбранного — он становится *символом* данного образа. На языке уравнений это значит, что уравнения (2) и (6) должны решаться одновременно. Подобная система являет собой пример системы дифференциальных уравнений для связанных переменных, решение которой неустойчиво. Этот математический эффект как раз и может описать неопределенность и неоднозначность выбора символа. Так или иначе, проблема требует технической доработки при относительной идеологической ясности.

Заметим, что подобная процедура (в любом из описанных вариантов) может *не дать ожидаемого результата*, а образ не получит своего символа. Тогда сам типичный образ должен быть уточнен (при помощи «диалога» с множеством первичных образов H_0). Однако вполне допустимо, что определенная часть типичных образов (не слишком часто вовлекаемых в процесс) не получает своего символа вообще.

После того, как символ все-таки образовался и межпластинные связи символа с его нейронами-прародителями установились, характер взаимодействия символ \leftrightarrow образ наиболее естественно представить себе следующим: активация символа приводит к активации его образа, и наоборот. Естественно предположить, что такие же связи устанавливаются и с прообразом типичного образа в множестве *первичных* размытых образов. Кроме того, устанавливается *параметрическая* связь нейрона-символа с его образом: параметр α модифицируется таким образом, чтобы данный гроссберговский нейрон уже не мог стать символом еще какого-либо образа $\alpha_k \rightarrow \alpha_k(\{\Psi_{ik}\})$ (см. [12]).

4.2. Формирование символов «смешанных» образов

Большой интерес представляет вариант, приближенный к реальности. Предположим, что образы предъявляются и, соответственно, записываются хаотично, т.е. пластина первичных образов представляет собой некую «паутину», в которой некоторые «нити» более «толстые», другие — менее, но что есть что — неизвестно.

Тогда самоорганизованное формирование символов приводит не только и не столько к сжатию информации, сколько к *осознанию* того факта, что вся эта паутина есть набор *K отдельных образов*. Подчеркнем, что при формировании пластины *типичных образов* такой проблемы не возникает: отбираются наиболее «толстые» нити, а что к чему относится — вопрос не встает. Но для формирования символов в том случае, когда образы пересекаются (имеют общие

нейроны), необходимо *отделить* их друг от друга, что есть акт высокого творчества. В результате этого акта генерируется совершенно новый тип информации — **семантическая информация**, т.е. **осмысление** факта *цельности* определенного набора нейронов (*образ*) и его *отдельности* от других образов.

Аналогом подобной ситуации в классической теории распознавания образов может рассматриваться проблема кластеризации, когда на множестве заданных объектов (неоднородных по сути и характеризующихся общим набором признаков) требуется выявить ***K*** отдельных образов, соответствующих ***K*** подмножествам похожих объектов (возможно и с нечеткими границами). Процедура кластеризации обычно предшествует собственно процедуре распознавания классов (кластеров). Поэтому еще от Бонгарда процедура поиска распознающего набора признаков, определяющего класс, называется **формирование названия класса** (см. [13]), что достаточно обоснованно согласуется по смыслу с термином **символ класса**.

Задача может решаться при помощи того же механизма, что и в предыдущем случае (когда семантическое содержание в процесс закладывалось «учителем»). Типичные образы передаются на пластину локализации. Если это происходит *строго одновременно*, тогда все они могут получить *только один общий символ*, и такая ситуация тоже имеет право на существование (воспринимается вся «картинка», а не отдельные ее элементы). Если же процесс в каком-то смысле *хаотичен*, то возникает *перемешивающий слой* (см. рис. 3б, верхний блок), в котором пробуются различные варианты распределения символов между данным набором активных нейронов, а метод «попытка + проверка» становится *единственным механизмом* образования символов.

В результате система «понимает», что к чему относится, а установившееся распределение связей (см. рис. 3б, нижний блок) есть *своя условная семантическая информация*. Она не является единственно правильной и, соответственно, не нуждается в доказательстве: система *интуитивно* восприняла «увиденное» именно так, а логического пути, однозначной процедуры, в этом случае нет.

После завершения описанной процедуры устанавливаются связи нейрона-символа с его нейронами-прародителями согласно уравнению (6), и активация конкретного символа приводит к активации его образа, и наоборот. Параметры α_k самих нейронов-символов, как и в предыдущем случае, модифицируются: $\alpha_k \rightarrow \alpha_k(\{\Psi_{ik}\})$.

4.3. Формирование символов более высокого уровня иерархии («символ класса»)

Следуя тому же алгоритму, можно представить и процесс формирования символов более высокого уровня иерархии, в частности, символа класса. В силу принципа компактности записи информации, схожие в чем-то объекты (принадлежащие к одному *классу*) записываются цепочками нейронов, имеющими общие составляющие.

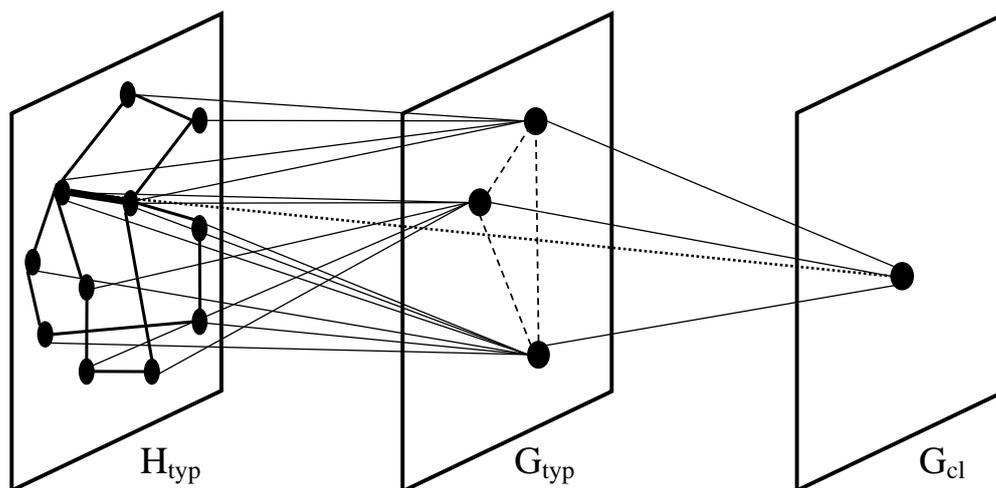


Рис. 4. Схема формирования символа класса.

На рис. 4 показана пластина типичных образов $H_{\text{тип}}$, на которой выделены три образа; их символы представлены на пластине типичных образов $G_{\text{тип}}$. Те связи, которые являются общими для всех образов, наиболее сильны. После превышения некоторого порога $\Omega_{\text{th}}^{\text{cl}}$ (большего, чем порог образования символа, см. рис. 1а) эта «цепочка общности» (она не является образом чего-либо, а соответствует именно *набору общих признаков*) передается на пластину G_{cl} следующего уровня иерархии, где по описанному выше алгоритму, превращается в **символ класса**. Отметим, что *иерархия символов* построена по принципу: чем выше уровень символа, тем с меньшим числом нейронов-признаков он связан, и тем больше множество образов, к нему относящихся (что соответствует принципу иерархии объектов природы по К. Линнею).

Как символ класса влияет на образную подсистему? По построению: активация символа класса после обучения связей непосредственно активирует лишь нейроны, которые служат общими признаками данного класса. Включает ли активация этих общих признаков какие-то конкретные образы, и если да, то какие, и в какой последовательности?

По-видимому, здесь более уместно говорить не о непосредственной активации кого-либо, а о **внимании** ко всему классу («огибающая» всех образов, имеющих эти общие нейроны-признаки). Это можно реализовать **параметрически**: если допустить, что порог активации β_i *нужных* нейронов Хопфилда (тех, которые являются общими признаками данного класса) может зависеть от активности символа этого класса $G_{\{i\}}$, т.е. $\beta_i \rightarrow \beta_i(G_{\{i\}})$, то все эти образы переводятся в *ждущий режим* — они еще не активны, но малейшее дополнительное воздействие приводит к их активации. Другими словами, эти образы получают *«преимущественное право» на активацию*, а это и есть то, что называется **вниманием**.

Таким образом, после обучения соответствующих межпластинных связей

и настройки параметров уравнений, активация символа класса будет переводить в *ждущий режим* все образы, относящиеся к данному классу.

Символы образа и символы класса в данной схеме не связаны между собой. Однако можно ввести связь «символ(образа)↔символ(класса)», Ξ_{vk} , через нейроны-«прародители» на пластине Хопфилда:

$$\frac{d\Xi_{vk}(t)}{dt} \propto \frac{1}{\tau_{\Xi}} G_v \left[\sum_i^M (H_i + 1) \right] G_k, \quad (7)$$

где суммирование проводится по общим нейронам-признакам для данного класса. Подчеркнем, что так определенная связь основана именно на связи *образов*.

4.4. Связи символ↔символ.

До сих пор обсуждались процессы, в основе которых лежат нейроны процессоров Хопфилда (*образные* нейроны). Именно эти нейроны порождают (при соблюдении определенных условий) формирование символов, участвуют в формировании связей «символ ↔ образ» и «символ образа ↔ символ класса». Однако естественно и даже необходимо предположить наличие связей «символ↔символ», *не опосредованных образами*.

Иными словами, должен формироваться тип связей, принципиально отличающийся от всех остальных тем, что никак не апеллирует к *образным* нейронам, т.е. внешним и *реальным* впечатлениям, «поданным» на систему. Связи нового типа строятся исключительно на основе той *картины мира*, которая сложилась внутри системы в процессе ее самоорганизации. Такие связи можно назвать *абстрактными*, и именно они позволяют вводить — через связь с другими символами — *абстрактные понятия*, не имеющие конкретного материального (образного) содержания (например, «совесть», «бесконечность», «понятие» и т.п.). Анализ таких абстрактных связей непосредственно связан с проблемой перехода символ→слово и требует специального исследования.

Первым шагом к решению проблемы естественно предложить алгоритм формирования связи символов одного уровня иерархии Y_{vk} , через общий для них символ более высокого уровня иерархии (*символ класса*):

$$\frac{dY_{vk}(t)}{dt} \propto \frac{1}{\tau_Y} G^{\sigma}_v [G^{\sigma+1}] G^{\sigma}_k, \quad (8)$$

на рис. 4 эти связи изображены пунктиром. Характерное время обучения τ_Y в этом случае, по-видимому, должно быть больше, чем для «непосредственного» обучения связей (6); впрочем, эта проблема требует самостоятельного исследования. Последнее утверждение относится и к самой структуре связей «символ↔символ», и к проблеме формирования элемента «символ-слово» (СС).

4.5. Обратный процесс: реконструкция образов через символы

В предыдущих разделах обсуждалось, как формируется символ на основе уже существующих образов. Возможен ли обратный процесс — возникновение образа вследствие активации символа или набора символов? Вопрос не праздный, поскольку именно так может осуществляться внешнее семантическое влияние на систему, если *символ* имеет статус *слова*, т.е. носителя общего в некотором сообществе смысла.

Простейший случай представляет собой путь: слово-символ типичного образа → декомпозиция → типичный образ (например, «нейрон Билла Клинтона», см. [14] и ссылки *ibid*). Тогда активация символа (в данном случае, *называние слова*) приводит к активации его образа. Однако не каждый образ может иметь свой собственный символ, тем более, отнюдь не каждый «внутренний» символ имеет соответствующее ему *слово*.

Символы более высокого уровня иерархии (также как и их *слова*) связаны с множеством образов. Следовательно, *конкретный образ* (даже если он не имеет своего символа-слова) может быть активирован *набором слов-символов* высокого уровня иерархии. Иначе говоря, конкретный образ можно представить как *пересечение множеств*, соответствующих данным символам, а также *произведению символов* в уравнении для образных нейронов (см. ниже, последний член в уравнении (10)), а также понятию «описать словами».

Всё сказанное верно в том случае, если данный образ *присутствует в библиотеке образов* — записан на пластине типичных образов или хотя бы в «размытом» множестве. Что делать, если такого *образа в библиотеке нет*? В этом случае образ может быть *воспринят извне*, через символьную систему, но для этого требуется *воображение*, т.е. реконструкция «предлагаемых обстоятельств» с использованием элементов, присутствующих в собственной образной библиотеке. *Механизм воображения* — очень интересная проблема и предмет для отдельного специального исследования.

Таким образом, мы видим, что символьные нейроны (как внутренние символы, так и символы-слова) и их связи (как между собой, так и с образными нейронами) весьма разнообразны и сами формируют развитую инфраструктуру, которую можно назвать *полем символьной (семантической) информации*.

5. Система уравнений, описывающая процесс мышления

Всё, сказанное выше, можно суммировать в форме системы связанных уравнений для двух образных и одной (с учетом ее собственной инфраструктуры) символьной подсистем:

$$\begin{aligned} \frac{dH^0_i(t)}{dt} = \frac{1}{\tau_i} [H^0_i + \beta_i(\{G^\sigma_{\{i\}}\})((H^0_i)^2 - 1) - (H^0_i)^3] + \sum_{j \neq i}^n \Omega^0_{ij} H^0_j + \\ Z(t)\xi_i(t) + \sum_j^n \Lambda_{ij}^{LR}(t) H_j^{typ} + \sum_k \Psi_{ik} G_{\{i\}}^k + \prod_\sigma G_{\{i\}}^\sigma \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{dH^{typ}_i(t)}{dt} = \frac{1}{\tau_i} [H^{typ}_i + \beta_i(\{G^\sigma_{\{i\}}\})((H^{typ}_i)^2 - 1) - (H^{typ}_i)^3] + \\ \sum_{j \neq i}^n \Omega^{typ}_{ij} H^{typ}_j + \sum_j^n \Lambda_{ij}^{RL}(t) H_j^o + \sum_k \Psi_{ik} G_{\{i\}}^k + \prod_\sigma G_{\{i\}}^\sigma \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{dG^\sigma_k(t)}{dt} = \frac{1}{\tau_k} \{ -(\alpha_i - 1)G_k + \alpha_k G_k^2 - G_k^3 \} - \sum_{j \neq k} \Gamma_{jk}(t) G_k^\sigma G^\sigma_j + \\ \sum_i \Psi_{ki} H_i + \sum_l \Xi_{kl} G_l^{\sigma+1} + \sum_l Y_{lk} G_l^\sigma \end{aligned} \quad (11)$$

Здесь $H^0_i(t)$, $H^{typ}_i(t)$ и $G^\sigma_k(t)$ — переменные, описывающие состояние *двух* образных и *одной* символьной (включая все ее уровни иерархии) подсистем, соответственно. Такое представление подчеркивает тот факт, что в данной схеме символьная подсистема связана как с типичными образами, так и с их *прообразами* в «размытом» множестве.

Отметим, что уравнения для образных подсистем фактически совпадают (по модулю учета влияния символов) с теми, которые были предложены в [2] для описания «*интуитивного*» (правополушарного — **ПП**) и «*логического*» (левополушарного — **ЛП**) мышления. Иными словами, обе работы не противоречат друг другу, если множество *первичных* образов отождествить с **ПП**, а множество *типичных образов* — с **ЛП**. Собственно, принципы построения этих подсистем и разделения их функций действительно совпадают в обеих работах; система уравнений (9–11) представляет собой следующий шаг в моделировании процесса мышления.

В свете изложенного выше (и фактически развивая точку зрения, высказанную в [2]), можно сформулировать следующее положение: в рассматриваемой

мой системе процессоров **чисто интуитивными** следует считать те процессы, в которых вовлечены нейроны (и их связи) «первичных» образов (присутствующих в «размытом» множестве и *отсутствующих* в множестве типичных образов) — т.е. та информация, которая записана в системе на ее *внутреннем* языке, но не получила *семантического* наполнения и не была преобразована в символьную форму (так как свои символы имеют только типичные образы).

С другой стороны, **чисто логической** формой мышления, согласно самому происхождению слова «логика» ($\lambda\acute{o}\gamma\omicron\varsigma$ = слово) следует считать те процессы, которые проходят *только с участием символов*, причем не всех, а только тех, которые получили *статус слова* (или могут быть описаны, через свои типичные образы, *другими словами* — см. п. 4.4.). Подобные процессы в описанной нами схеме вовлекают только нейроны *типичных образов* (и, соответственно, их символы), т.е. «левое полушарие», и никак не апеллируют к множеству первичных образов («правое полушарие»).

Из сказанного ясно, что в чистом виде обе формы мышления практически не существуют: сама форма уравнений подразумевает взаимодействие всех элементов со всеми, и выделение упомянутых связей (при разрыве остальных) выглядит неестественным и ненужным. В принципе, оно возможно, но скорее всего, в связи с определенными патологиями мыслительной системы. Гораздо более естественно выглядит «диалог» между ПП, ЛП и символьным полем. Более подробно эту проблему мы обсудим в Приложении.

В приведенном виде система уравнений (9–11) содержит все возможные взаимодействия типа образ \leftrightarrow образ, символ \leftrightarrow образ и символ \leftrightarrow символ. Уравнения обучения внутрисластинных связей пластин Хопфилда (Ω) и Гроссберга (Γ) приведены ранее, причем связи Ω^0 обучаются согласно «альтернативному варианту» (4), а Ω^{yp} обучаются «по Хопфилду» (5). Взаимодействие образ \leftrightarrow символ обеспечивают связи Ψ , обученные согласно (6). Связи символа с «соседями» (Y) и символами более высокого уровня иерархии (Ξ) заданы уравнениями (8) и (7), соответственно.

Межпластинные связи Λ между образными подсистемами не должны быть симметричны ($\Lambda^{\text{RL}} \neq \Lambda^{\text{LR}}$), их назначение меняется на различных этапах формирования системы (это обстоятельство подчеркнуто выделением явной зависимости $\Lambda(t)$). На этапе образования *типичного* множества из *первичного* работают *прямые* связи, их назначение — передавать (копировать) образы, осуществляя переход образ \rightarrow типичный образ ($\Lambda_{ii}^{\text{RL}} = 1$, $\Lambda_{ij}^{\text{LR}} = 0$). Далее роль связей образ \leftrightarrow образ — обеспечивать постоянный «диалог» между двумя образными подсистемами. Следуя аналогии с полушариями, использованной в [2], межпластинные связи Λ естественно соотнести с *corpus colossum*. Остальные связи, меняющиеся в процессе обучения, остаются практически постоянными для *обученной* системы.

Подчеркнем, что данная система уравнений не является алгоритмом ре-

шения какой-либо конкретной задачи. Скорее, она представляет собой аналог «производящего функционала», или *язык*, пригодный для описания алгоритмов. Число подсистем можно увеличивать, выделяя специфические подструктуры (пластины и их связи) как в образных, так и в символьной подсистемах.

Решение такой системы уравнений (буде оно когда-либо может быть получено), т.е. траектории $H_i(t)$ и $G_k(t)$, по сути представляет собой «ход мысли». Однако необходимо помнить, что вид уравнений универсален, а сами они (т.е. *обученные связи*) строго *индивидуальны* для некоторой конкретной мыслительной системы. Следовательно, *результат решения* такой системы уравнений представляет скорее когнитивный интерес. Более плодотворным представляется *анализ* возможных решений и их классов, т.е. именно *алгоритмов*, создаваемых самой системой. Эта задача выходит за рамки данного исследования и будет решаться отдельно.

6. Заключение

Анализ роли понятий *образ* и *символ* в моделировании процесса мышления средствами нейрокомпьютинга приводит к ряду важных выводов.

1. Образные подсистемы играют ведущую роль в *записи, хранении* и даже сжатии информации, причем принципиально необходимо наличие *двух подсистем* – пластины *первичных («размытых») образов* и пластины *типичных образов*. Понятие *типичный образ* формируется после и на основании сформированного *размытого множества*, и связано с частичной потерей информации («забыть можно только то, что знал»). Возможность мыслить *только «по Хопфилду»* не противоречит основам динамической теории информации.

2. Формирование символьной подсистемы – поля условной информации – есть следующий принципиально важный шаг, поскольку *символьная информация является семантической*. Символы формируются *только после и на основе* формирования типичных образов. Роль символа не только (и не столько) в том, чтобы сжимать (кодировать) информацию; его истинное назначение в том, чтобы *обеспечивать взаимодействие* пластин.

3. На следующем шаге — от внутренней семантической информации к общесловной, т.е. к *языку общения* — *символы обеспечивают коммуникацию* с другими мыслящими системами по семантическому каналу.

4. Параметрическое влияние символьной подсистемы на образную обеспечивает парадигму *внимания*: в данной схеме внимание есть перевод в *ждущий режим* (понижение порогов активации) нейронов тех образов, которые связаны с активированным символом.

5. Взаимодействие образного и символьного полей представляет собой ключевой элемент в процессе мышления — по сути, это и есть процесс мышления. Предложенная система уравнений для описания взаимодействия образных и символьной подсистем должна описывать «ход мысли» в индивидуальной си-

стеме.

Обсудим специально вопрос о том, насколько формальный нейрон «не похож» на реальный. В предлагаемой схеме, построенной из двух типов формальных нейронов, воспроизводятся (в результате их *нелинейного взаимодействия*) многие эффекты, которые, казалось бы, свидетельствуют в пользу концепции нейрона как сложного организма. Так, «нейрон Билла Клинтона» (см. [14] и ссылки *ibid*) в нашей концепции есть элемент символьной подсистемы, который стал таковым в результате *обучения и настройки* системы. Еще один пример конкретной модификации самих нейронов типа Хопфилда в результате формирования символа класса – появление параметрической зависимости от символа класса, т.е. модификации параметров β . Также модифицируются и параметры α нейронов типа Гроссберга в результате формирования связей символ \leftrightarrow образ.

Вообще специализация есть не изначальное, но приобретаемое свойство как для «символьных», так и для «образных» нейронов, но точно то же утверждение справедливо и в отношении реальных нейронов (см. [5] и ссылки *ibid*). Однако остается область, которую не удастся формализовать адекватно в рамках предложенной схемы, — это *эмоции*, а их роль в процессе мышления (*эмоциональный выбор*) несомненна и очевидна. Эта проблема требует приближения модели формального нейрона к реальному.

В заключение перечислим те вопросы, которые остались не исследованными в должной мере; ответы на них сами по себе могут служить «ключами» для модификации модели и более глубокого понимания процесса мышления. Это:

1). С чем соотнести *порог* Ω_{th}^{typ} перехода первичного образа в типичный и порог Ω_{th}^{cl} образования символа класса? Какую роль при этом может играть *амплитуда шума*?

2). *Когда* появляется потребность в символе, т.е. что является *сигналом* к началу процесса перехода $TO \rightarrow C$? Сразу после формирования TO или должна быть некая задержка?

3). Каковы *характерные времена* обучения связей? С какими параметрами системы их следует соотносить?

Все эти (и связанные с ними) вопросы требуют пристального внимания и продолжения работы в данном направлении.

Авторы выражают искреннюю благодарность Ю.И. Александрову за интересные и плодотворные дискуссии о проблеме формального и реального нейрона.

ПРИЛОЖЕНИЕ 1. МЕХАНИЗМЫ ИНТУИЦИИ И ЛОГИКИ

Понятия *интуитивное* и *логическое*, равно как *сознание* и *подсознание*, до сих пор, несмотря на их популярность и распространенность, остаются дискуссионными, и механизм этих типов мышления остается в центре внимания и интереса. Наиболее формализовано выглядит понятие «логика»: его этимология связана с греческим λόγος=слово, мысль, т.е. в буквальном переводе логическое мышление есть просто мышление вербализованное. Однако в современном мире определение слова «логика» во-первых, неоднозначно², т.е. четкого таки нет, во-вторых, включает отнюдь не *только словесное* содержание:

Л1. Научная дисциплина, изучающая способы доказательств и опровержений.

Л2. Внутренняя закономерность, присущая явлениям природы, общества.

Л3. Правильный, разумный ход рассуждений, умозаключений.

*Л4. Способы достижения истины в процессе познания опосредованным путём, не из чувственного опыта, а из знаний, полученных ранее, наука о способах получения **выводного знания**.*

Из приведенных определений видно, что слово нагружено целым спектром функций, где понятие вербализации (как в варианте 2) иногда даже и не требуется. Мы будем иметь в виду все эти определения, не пытаясь вывести «правильное».

По поводу интуиции ситуация схожая:

И1. Неосознанное чувство, позволяющее постигать суть чего-либо и подсказывающее правильное поведение, решение.

И2. Способность непосредственного постижения истины без обоснования с помощью доказательств.

И3. Прямое усмотрение истины, единственным критерием правильности признающая только внутреннее удовлетворение (Иммануил Кант).

Отметим, что определения интуиции даются «от противного», т.е. указывается, чего *нет* в интуитивном мышлении. Исключение составляет лишь определение Канта, которое дает хоть какой-то (а точнее, чисто *эмоциональный*) критерий правильности интуитивного решения.

Та же ситуация имеет место с понятием **сознание**.

С1. Процесс отражения действительности мозгом человека, включающий все формы психической деятельности и обуславливающий целенаправленную деятельность человека.

С2. Восприятие и понимание окружающего, свойственные человеку; ум, разум. Способность осмысленно воспринимать окружающее.

С3. Ясное понимание чего-либо.

С4. Мысль о чем-либо, чувство, ощущение чего-либо.

И, наконец, также неоднозначно понятие **бессознательное**.

Б1. Совокупность ... процессов, в отношении которых отсутствует субъективный контроль.

*Б2. Все, что **не** является объектом осознания.*

² даже если абстрагироваться от самостоятельно специфицированных понятий, таких как формальная, дедуктивная, индуктивная, классическая, математическая и т.д. логики

Здесь также понятие «сознание» перегружено (слишком много определений по сути своей то же, что и отсутствие определения), а его антоним определяется от противоположного.

Моделирование процесса мышления, основанное на нейрокомпьютинге и ДТИ, дает возможность подойти к этой проблеме несколько иным путем, а именно, — не давать определение рассматриваемым понятиям, а попытаться представить себе, какие *механизмы* в рассматриваемой мыслительной системе могут выполнять заявленные в приведенных выше определениях функции. Обсудим, как соотносятся понятия *интуитивного* и *логического* мышления с рассмотренными функциями образных и символьных нейронов.

В рамках представленной нами картины мыслительной системы можно явно выделить 4 элемента, в равной степени необходимых для функционирования всей системы и при этом имеющих четко различные функции. Это:

1. «серые» нейроны (*первичный образ, O*) и *слабые* («серые») связи Ω^0 , возникающие в результате пусть даже однократного воздействия;

2. *типичные образы (TO)* и *сильные* («черные») связи Ω^{typ} , возникающие в результате длительного обучения системы как реакция на объект в разных его проявлениях;

3. *символы* типичных образов (*C*) и смысловые (*семантические*) связи Ψ между символами и *TO*, $C \leftrightarrow TO$ — условная информация, возникающая в результате внутреннего кодирования и *осознания* факта однозначной связи образа с реальным объектом;

4. символы-слова (*CC*), имеющие не прямую (через образные нейроны), а *семантическую* связь и с внутренними символами, и с внешними понятиями-словами, и *абстрактные Y* связи $C \leftrightarrow CC$, не опосредованные образными нейронами-прародителями, а возникшие в результате настройки всей системы.

Кроме того, существует (независимо от них) еще и пятый элемент — *шум*, или случайное воздействие, которое (в силу случайности) не может быть иллюстрировано, но присутствует в общей системе уравнений (9)–(11). Специфическую роль шума мы обсудим чуть ниже, а его природу, как мы неоднократно упоминали (в частности, в [2,3]), естественно связать с *эмоциональным* полем.

Основные элементы не независимы и не равноправны. Порядок их формирования в данной системе не произволен, а определен именно так, как они представлены: *TO* возникает после и на основе формирования множества первичных образов, *C* формируется на основе *TO*, *CC* основан на связях $C \leftrightarrow CC$. Переход от каждого предыдущего уровня к последующему представляет собой принципиально важный шаг, приводящий к *новому типу информации*, поскольку на каждом переходе система генерирует новый *тип связей*: слабые → сильные → семантические → абстрактные.

Важно подчеркнуть, что каждый переход сопровождается *потерей* части

накопленной информации — точнее, она остается на предыдущем уровне и не попадает далее. Такую информацию можно назвать *внутренней*, или *скрытой* (*служебной* для данной системы) — она не переходит на уровень вербализации, следовательно, не может быть передана другим мыслительным системам («опубликована»). В отличие от нее, *внешней* можно назвать информацию, *осознанную* в общепринятом смысле, т.е. понятную не только себе, но и другим — вербализованную информацию. Последняя возникает только на финальном, четвертом уровне, который завершает формирование системы. Таким образом, *сознанием* можно назвать *процесс последовательной активации символов и образов (=ход мысли)*, который может быть *вербализован*.

Подчеркнем, что, вербализация (или *сознание*) еще не является гарантией правильности или *логичности* процесса мышления (ср. с определениями Л2, Л3). Последнее подразумевает возможность *доказательства*, т.е. активацию не только *общепонятных* слов, но и *общеизвестных* и *общепринятых* причинно-следственных связей. Такие связи не имеют специального обозначения в предлагаемой модели, поскольку возникают (*выводятся*) не в результате самоорганизации индивидуальной мыслительной системы (т.е. обобщения индивидуального опыта), а основаны на обобщенном опыте *всего социума* и, в конечном итоге, приобретают статус *законов природы*. Таким образом, к *логическому* мышлению естественно отнести оперирование *вербализованными понятиями и абстрактными связями*, причем лишь теми, которые считаются *установленными* в данном социуме.

С другой стороны, *выводы, основанные на скрытой информации*, естественно интерпретировать как *интуитивное* мышление. *Скрытая* (*внутренняя*) информация сугубо *индивидуальна*, возникает в результате *самоорганизации* данной мыслительной системы и поэтому *не требует доказательств*.

Отметим, что внутренняя информация имеет *разные уровни* скрытости, что существенно влияет на усилия по ее извлечению на уровень *сознания*. Так, роль «*серых*» связей — в том, чтобы хранить казалось бы ненужную образную информацию, которая потом, при каких-то непредвиденных обстоятельствах, может оказаться важной. «*Серые*» связи в принципе не имеют отношения к процессу формирования символов (хотя бы даже внутренних, служебных), следовательно, эта информация сама по себе не может быть не только вербализована, но даже и «вытащена на поверхность», т.е. активирована с помощью какого-либо символа. Иными словами, она не может быть *подконтрольна* даже внутри данной индивидуальной системы — т.е., по определению Б1, представляет собой *бессознательное* или *подсознание*. В этой связи становится понятна и важна роль *шума*, который может *активировать* нужную «серую» цепочку и тем самым вывести ее на уровень *сознания* («вдруг увидеть внутренним взором»). Последний эффект естественно ассоциировать с *озарением* (*insight*). Поскольку в данной схеме этот акт принципиально связан с шумом, указать

его источник, тем более доказать правильность, невозможно. Согласно Канту, акт озарения сугубо интуитивен.

Следующий уровень представлен процессами с участием типичных образов и внутренних символов. При необходимости каждый такой процесс может быть вербализован при помощи абстрактных связей $C \leftrightarrow CC$. Однако при переходе $C \rightarrow CC$ от семантической информации к вербализованной остается множество символов, для которых стандартных слов не существует — это некие цельные «картинки», описание которых требует декомпозиции и поэлементного анализа. Иными словами, один внутренний символ может быть описан при помощи многих слов, причем чем ярче и конкретнее «внутренняя картинка», тем больше слов необходимо для ее описания. Однако этот акт требует не *озарения*, а *подбора* нужных слов, что уже граничит с *логическим* мышлением.

Все сказанное может быть проиллюстрировано на примере задач, связанных с определением сходства и/или различия каких-либо объектов. Особенно интересны ситуации, когда требуется найти что-то общее у очевидно различных объектов, или, наоборот, найти тонкие отличия у двух или более очень похожих. Такие задачи решаются на уровне образных подсистем *автоматически*: сходство определяется общими нейронами, участвующими в записи образов, различие — разными, и система это *знает*.

Отметим, что в пластине *ТО* задача может решаться быстро, но не точно: тонкости сходства или отличия могут проявляться только во множестве первичных образов (*О*), т.е. на уровне «серых» связей.

Однако это знание не может быть *формализовано* (относительно человека можно сказать, что оно существует на уровне *ощущений*), пока оно не *осознано*. В данном случае *осознание* означает, что нескольким конкретным нейронам (в задачах, о которых идет речь, их число невелико), отвечающим за сходство/различие образов, необходимо придать статус «семантических признаков» — а именно, найти или сформировать соответствующие им *символы*, пусть хотя бы внутренние. Иными словами, *служебное* знание должно быть переведено в семантическое. Следующий шаг — вербализация этого знания — требует уже активной работы именно в символьном поле (формирования абстрактных связей данного внутреннего символа с символами-словами).

Жизненная аналогия: на вопрос «одинаково или нет?» быстро отвечаешь «да» или «нет», а на следующий вопрос «почему?» — долго мучаешься. Когда (если) ответ удастся *сформулировать*, т.е. не только *вербализовать*, но и *объяснить* — это уже *логическое* мышление, и именно в этом заключается его преимущество.

Решение, найденное в образной системе, *абсолютно верно* для данной системы (индивида), но может быть *объективно ошибочно*, так как способ записи (активация определенных образных нейронов) *сугубо индивидуален*. Таким образом, это решение по определению является чисто интуитивным: оно основано на *индивидуальном опыте*, т.е. *картине мира*, записанной в данной си-

стеме, и не требует доказательств. Однако оно *может быть* доказано, если оно *осознано* (выражено через внутренние символы) и, затем, *вербализовано*. Последнее по сути есть метод перевода интуитивного знания в «логическое».

Таким образом, выделение специфически интуитивных и логических процессов в мышлении представляется вопросом терминологии. Принципиально же противопоставление «внутренней» (*скрытой*) и «внешней» (*общепонятной*) информации, причем часто внутренняя информация может быть переоформлена в виде комбинации символов-слов, т.е. опубликована (в теории распознавания этот процесс носит название «формализация экспертного знания»).

Из всего сказанного можно заключить, что естественный процесс мышления должен включать операции со всеми элементами, независимо от их принадлежности к информационным полям различной природы.

В заключение коснемся проблемы, которая мучила авторов со времен написания работы [2]: чем отличается *интуиция* (*insight*) от *мудрости*? Последняя обычно ассоциируется с почтенным возрастом, первая не имеет привязки к возрасту, но имеет гендерный оттенок (впрочем, последнее спорно).

Принято считать, что интуиция присуща в основном женщинам, а логическое мышление — прерогатива мужского пола. Однако это утверждение сомнительно и может быть верно лишь «в среднем».

Сейчас мы можем попытаться ответить на этот вопрос.

Само понятие *инсайта* подразумевает тот факт, что система *знает* (на уровне подсознания) верное решение, но не может его активировать: *серые* цепочки нейронов не имеют связей с символами. Само по себе *знание*, т.е. количество вообще всех, в том числе «серых», образных связей, с возрастом увеличивается, следовательно, вероятность *скрытого* знания также повышается с возрастом. Однако проблема в том, чтобы это знание вывести в область *сознания*, иными словами, найти «мост» к нужному «серому» образу.

Мы видим следующий механизм, требующий длительной и интенсивной эксплуатации системы, т.е. активной работы и развития инфраструктуры символьного (равно как и образного) полей. Развитая инфраструктура *абстрактных* и *семантических* связей эквивалентна той самой системе «мостов и развязок» *над* образным полем («ландшафтом»), которая обсуждалась (вербально, на уровне предположения) в работе [2] и которая способна помочь «проложить путь» к нужному образу. В некотором смысле она способна *заменить шум* по скорости достижения нужного образа, причем в случае мудрости этот путь *не случаен*. Отличие мудрости от инсайта именно в том, что весь путь к цели может быть *осознан и вербализован* — и в этом преимущество мудрости. Подчеркнем, однако, что говоря о мудрости, мы имеем в виду *все семантические и абстрактные* связи, что гораздо больше, чем *только логическое* мышление.

Таким образом, согласно этой логике мудрость связывается именно с *большим опытом* (и/или почтенным возрастом), а гендерного влияния ни на мудрость, ни на интуицию не просматривается.

Литература

1. Чернавский Д.С., Карп В.П., Родитат И.В., Никитин А.П., Чернавская Н.М. Распознавание. Аутодиагностика. Мышление. – М.: Радиотехника, 2003.
2. Чернавская О.Д., Никитин А.П., Чернавский Д.С. Концепция интуитивного и логического в нейрокомпьютинге // Биофизика. – 2009. – т.54. – № 6. – с. 1103.
3. Чернавская О.Д., Чернавский Д.С., Карп В.П., Никитин А.П., Рожило Я.А. Процесс мышления в контексте динамической теории информации. Часть I: основные цели и задачи мышления. / Препринт ФИАН, 2011. – №10. – 20 с.
4. Чернавский Д.С. Синергетика и информация: Динамическая теория информации. – М.: Наука, 2001.
5. Александров Ю.И., Анохин К.В., Безденежных Б.Н., Гарина Н.С., Греченко Т.Н., Латанов А.В., Палихова Т.А., Савельев С.В., Соколов Е.Н., Тушмалова Н.А., Филиппов В.А., Черноризов А.М. Нейрон. Обработка сигналов. Пластичность Моделирование. Фундаментальное руководство. – Тюмень: Изд. ТГУ, 2008.
6. *Fitz Hugh R.* Impulses and physiological states in theoretical models of nerve membrane // *Biophys. J.*– 1961.– v.1.– p.445; *Nagumo J., Arimoto S., Yashukawa S.*//*Proc.IRE.*– 1962.– v.50.–p.2062
7. Шамис А.С. Пути моделирования мышления. – М.: КомКнига, 2006.
8. *Hopfield J.J.* Neural networks and physical systems with emergent collective computational abilities // *PNAS* – 1982 – v.79. – p. 2554.
9. *McCulloch W.S., Pitts W.* A Logical Calculus of the Ideas Immanent in Nervous Activity // *Bulletin of Mathematical Biophysics.* – 1943. – v.5. – p.115.
10. *Muller B., Reinhardt J.* Neural networks. – Springer Verlag, 1990.
11. *Grossberg S.* Studies of Mind and Brain. – Boston: Riedel, 1982; The adaptive brain. – Amsterdam: Elsevier/North-Holland, 1987.
12. Чернавский Д.С., Карп В.П., Васильев А.А., Чернавская О.Д. Математическая модель процессора локализации образа / Препринт ФИАН №9. – 2011. – 17 с.
13. Бонгард М.М. Проблема узнавания. М.: – Наука, 1967 – 320 с.
14. Яхно В. Г., Полевая С.А., Парин С.Б. Базовая архитектура системы, описывающей нейробиологические механизмы осознания сенсорных сигналов. // Когнитивные исследования: Сборник научных трудов: Вып. 4 / под ред. Ю.И. Александрова, В.Д.Соколова, — М. Изд-во «Институт психологии РАН», 2010, с. 273.