БИОФИЗИКА, 2016, том 61, вып. 1, с. 185–200

ДИСКУССИИ

УДК 577.38

ЕСТЕСТВЕННО-КОНСТРУКТИВИСТСКИЙ ПОДХОД К МОДЕЛИРОВАНИЮ МЫШЛЕНИЯ

© 2016 г. О.Д. Чернавская, Д.С. Чернавский

Физический институт им. П.Н.Лебедева РАН, 119991, Москва, Ленинский просп., 53

E-mail: olgadmitcher@gmail.com

Поступила в редакцию 10.11.15 г.

Рассмотрен естественно-конструктивистский подход к моделированию когнитивных процессов, основанный на динамической теории информации, технике нелинейных дифференциальных уравнений и концепции «динамического формального нейрона». Представлен вариант архитектуры когнитивной системы, разработанный в рамках естественно-конструктивистского подхода. Важной конструктивной особенностью этой архитектуры является разделение всей системы на две подобных подсистемы (аналог правого и левого полушарий мозга). В одной из них происходит генерация информации и обучение, в другой – рецепция и работа с уже известной информацией. Это разделение функций обеспечивается присутствием шума (случайного фактора) в подсистеме генерации; в подсистеме рецепции все процессы происходят последовательно и не случайно. Обсуждается интерпретация понятий: интуиция, логика, сознание и подсознание. Архитектура естественно-конструктивистского подхода сравнивается с другими теоретическими подходами (теорией графов и концепцией «когнитома») и анатомическими данными. Предлагается идея эксперимента, который может подтвердить или опровергнуть основные выводы естественно-конструктивистского подхода.

Ключевые слова: информация, образ, символ, шум, интуиция, архитектура.

Понимание и моделирование процесса мышления привлекает к себе неизменный интерес и представляет яркий пример междисциплинарной проблемы (см. [1,2] и ссылки там же). Действительно, понимание механизмов работы мозга требует знания физики (система материальная), биологии (система живая), а также психологии и философии (система *мыслящая и говорящая*). В последние годы сформировалось специальное научное направление под названием «когнитология» (от лат cognitus – познавать), объединяющая усилия нейрофизиологов, лингвистов, математиков и др. для решения этой проблемы. В рамках самого этого направления существуют разные «подходы к моделированию мышления» (см. [3–9]), в частности робототехника [7], искусственный интеллект [4], биологически-инспирированные когнитивные архитектуры (*BICA*, [5,6]) и т.д., основанные на различных компьютерных технологиях.

После того, как в работах [10–12] была предложена и развита концепция нейропроцессоров, основанная на парадигме обучения, надежды на решение проблемы традиционно связывают с нейрокомпьютингом. Действительно, объединение компьютерной парадигмы с парадигмой *обучения* гораздо ближе к механизмам мышления и человека и животных (условные рефлексы), чем исполнения программы, заданной извне (так работают обычные компьютеры). Именно это направление представляется наиболее адекватным для парадигмы искусственного интеллекта. Однако излишне упрощенное представление о «формальном» нейроне, традиционно используемое в нейрокомпьютинге [13], сужает возможности стандартных нейропроцессоров и инициирует попытки ревизии основных парадигм нейрокомпьютинга (см. [14,15]).

В настоящее время интерес к проблеме моделирования когнитивных процессов и создания искусственного интеллекта переживает несомненный подъем. На наш взгляд, это связано с развитием экспериментальной техники (функциональная магнито-резонансная томография, магнитная и электрическая энцефалография и т.п.), которая позволяет ставить эксперименты на людях. Сейчас стало возможным детектировать активность головного мозга человека с хорошим пространственным и (что особенно важно) временны́м разрешением [16]. Это позволяет проследить динамику активности если

Сокращения: **ПП** – правое полушарие, **ЛП** – левое полушарие, ЕКП – естественно-конструктивистский подход, ДТИ – динамическая теория информации.

не одного нейрона, то определенных зон мозга, функциональное назначение которых известно из анатомических данных. В результате появилась возможность получать *комбинированные* экспериментальные данные – как объективные (измеряемые приборами), так и субъективные, т.е. оценка самим человеком его состояния и (главное) мыслей, т.е. того, о чем он думает в данный момент. Это открывает широкие возможности для проверки модельных предсказаний.

Важно подчеркнуть, что существующие имитационные модели (см., например, [5–8]) направлены на создание систем, способных выполнять определенный круг задач *лучше*, чем человек. При этом приоритет отдается *надежности, эффективности и быстродействию* предлагаемых моделей. Однако в последнее время становится все более ясным, что при моделировании *человеческого* мышления такой подход не вполне адекватен. Действительно, человек, сталкиваясь с различными жизненными ситуациями, должен научиться решать самые разные, в том числе некорректно поставленные, задачи. На первый план выходит проблема «выживания», или *адаптационной способности* к неожиданным и непредсказуемым ситуациям. Поэтому человеческое мышление не *детерминировано,* часто *непредсказуемо* и всегда *индивидуально*. Загадка индивидуальности мышления – один из основных современных «вызовов» (challenge) для моделирования процесса мышления.

Еще одна ключевая проблема в моделировании мышления – имитация и интерпретация логического и интуитивного мышления. К ней примыкает «загадка двух полушарий»: почему головной мозга человека разделен на две очень близкие по строению, но все же несколько различные подсистемы, которые к тому же общаются чрезвычайно активно: связи между полушариями (*corpus callosum*) на порядки многочисленнее, чем связи с окружающим миром. Высказывалась (и завоевала популярность) гипотеза о том, что специализация полушарий действительно существует, причем правое (**ПП**) обучается, а левое (**ЛП**) – оперирует с уже *известной* информацией [17]. С другой стороны, широко распространено мнение, что **ПП** отвечает за интуитивное мышление, а **ЛП –** за логическое (см., например [3,18]); при этом, однако, сами понятия *интуитивного* и *логического* не имеют четкого и однозначного определения. Тем не менее сам факт необходимости двух связанных подсистем (полушарий) является одновременно и загадкой, и подсказкой.

Большинство теоретических моделей строится так, чтобы воспроизвести известные экспериментальные данные. В наших работах [19–22] мы разрабатывали и использовали так называемый «Естественно-конструктивистский» подход (ЕКП), основанный на динамической теории информации (ДТИ, [23]), технике нелинейных динамических дифференциальных уравнений и нейрокомпьютинге (см. [24,25] и ссылки там же). Однако последняя составляющая нами понимается в смысле, отличном от стандартного подхода, в частности, к понятию «формальный нейрон». Само название ЕКП отражает тот факт, что мы идем не от эксперимента, а «конструируем» когнитивную систему при помощи тех инструментов, которые необходимы для выполнения функций мышления.

Следует подчеркнуть, что единого определения процесса мышления не существует. В рамках ЕКП в работе [19] были исследованы основные принципы процесса мышления с позиций ДТИ и предложено определение мышления путем перечисления тех функций, которые оно должно выполнять. В таком подходе мышление есть *самоорганизованный процесс записи (восприятия), сохранения (запоминания), кодирования, обработки, генерации, а также распространения «своей» информации.*

В данной работе мы представляем вариант архитектуры когнитивной системы, разработанный в рамках ЕКП, и сравниваем его с представлениями, принятыми в других теоретических подходах – теории графов (нейросетей) [26] и концепции *когнитома* [27]. Кроме того, предлагается идея постановки эксперимента по проверке предсказаний, сделанных на основе предлагаемой ЕКП-архитектуры.

ОСНОВЫ ЕСТЕСТВЕННО-КОНСТРУКТИВИСТСКОГО ПОДХОДА

Концепция динамического формального нейрона. Традиционный нейрокомпьютинг основан на концепции формального нейрона как дискретного сумматора сигналов, который активируется, если сигнал превышает некоторый порог. С другой стороны, в нейрофизиологии нейрон – гораздо более сложная система. Наиболее релевантной (см. [15]) остается модель Хочкина–Хаксли [28] и ее упрощенный вариант – модель ФицХью–Нагумо [29,30]. В рамках ЕКП мы используем континуальное представление формального нейрона, которое представляет собой предельный случай модели ФицХью–Нагумо. Здесь *динамический* формальный нейрон (см. [21,22]) есть *бистабильный* элемент, имеющий два устойчивых стационарных состояния. Он описывается динамическим дифференциальным уравнением вида

$$\frac{dH(t)}{dt} = \frac{1}{\tau_H}[H - \beta(H^2 - 1) - H^3] \equiv \Im_H(H,\beta), \quad (1)$$

где $H(t)$ – переменная, описывающая состояние нейрона; τ_H – характерное время его активации; β – параметр, определяющий порог возбуждения (регулирует степень готовности данного нейрона к активации, т.е. адресное *внимание*). Здесь стационарные состояния переменных H равны +1 (активное) и –1 (пассивное), как в процессоре Хопфилда [11]. Для удобства дальнейшего изложения мы обозначим всю конструкцию, регулирующую поведение одного «хопфилдовского» нейрона через $\Im_H(H,\beta)$.

Основные элементы динамической теории информации применительно к мыслительным системам. Само понятие «*информация*» сравнительно ново, о нем заговорили в середине XX века. В научной литературе можно найти большое количество различных определений этого понятия. Наиболее ясный и *конструктивный* (что важно именно для моделирования мышления) характер имеет определение, предложенное Г. Кастлером [31]: *Информация есть запомненный выбор одного варианта из множества возможных и равноправных*.

Это определение не противоречит остальным, но в отличие от них дает представление о том, *как* информация возникает. Выбор может быть сделан в результате двух различных процессов – *рецепции* и *генерации* информации.

Рецепция информации есть выбор, предопределенный (навязанный) извне, для чего используется термин «обучение с учителем».

Генерация информации – свободный, т.е. не предопределенный извне, *случайный* выбор. Важно отметить, что генерация информации возможна *только в том случае*, если система находится в состоянии «перемешивающего слоя» [32], или «*джокера*» [33], т.е. когда эта система (возможно, в относительно короткий промежуток времени) ведет себя квазихаотично или подвергается *случайному воздействию* (обычно для этого используется термин «*шум*»).

В зависимости от того, *кем* делается выбор, возникают:

– *Объективная*, или *безусловная* информация – выбор, «сделанный» Природой, который отражается в устройстве внешнего (по отношению к мыслящему субъекту) материального мира, т.е. фактически законы физики. Эта информация не генерируется, а рецепцируется из окружающей среды либо непосредственно, либо с помощью приборов.

– *Условная* информация – выбор, сделанный *коллективом субъектов* в результате их *взаимодействия*: общения, борьбы, договоренности, условности. Примерами могут служить код (в частности, генетический), алфавит, язык и т. п. Важно, что данный выбор не обязан (а часто и не может) быть *наилучшим* (варианты часто *a priori* равноправны), но он должен быть сделан и принят в данном сообществе.

Подчеркнем, что условная информация играет особую роль в мышлении человека. Дело в том, что сам процесс восприятии и записи объективной внешней информации *субъективен* («смотрят глаза – видит мозг»): эта информация перерабатывается (кодируется нейронами) и становится уже внутренней, индивидуальной и *своей* для данного человека (системы). Таким образом человек *генерирует* свое восприятие мира, и в этом смысле переработанная *объективная* информация превращается в *условную для данной системы*. Иными словами, вся система связей, возникающая в результате обучения, есть результат «договоренности» ансамбля нейронов. Именно эту информацию человек сохраняет, защищает и распространяет.

Распространение информации согласно ДТИ означает способность системы *формулировать* свою условную информацию на общепринятом *языке,* а также *понимать* семантическое содержание символьной информации, поступающей извне. Язык (членораздельная *речь*) играет чрезвычайно важную роль в процессе мышления человека. В середине–конце прошлого века был очень популярным вопрос «как мозг делает мысль?». Сейчас становится понятным, что он это делает при помощи языка (речи): в экспериментах [34] было показано, что вспоминая (и даже воображая) какую-то ситуацию, человек *проговаривает* (хотя бы «про себя») свои мысли, следовательно, формулируя (формируя) их. Таким образом, *вербализация*, т.е. овладение языком, является обязательной функцией когнитивной системы высокого уровня.

Наконец, подчеркнем важный вывод из ДТИ: процессы *сохранения* (запоминания, рецепции) и *создания* (генерации) *новой* информации *дуальны*, т.е. *альтернативны* (в частности, при возникновении новой старая информация может пострадать). Отсюда следует, что для выполнения обеих функций необходимо участие *двух подсистем*, дополняющих друг друга: одна, в которой информация возникает (генерируется), и другая, где она сохраняется. В первой должны быть условия, необходимые для свободного (случайного) выбора: *перемешивающий слой* или *шум;* вторая должна быть стабильна.

Необходимые инструменты. В работе [20] детально исследовались «инструменты», т.е. типы процессоров, необходимые для решения перечисленных задач мышления. Под процессором здесь и далее понимаем пластину, населенную n динамическими формальными нейронами.

Было показано, что задачи записи и сохранения образной информации (без ее *осмысления)* решаются наиболее естественно при помощи линейных аддитивных процессоров типа Хопфилда (распределенная память). Здесь реальному внешнему образу, рецептируемому системой, соответствует некоторый набор M активированых нейронов, распределенных (както) по всей пластине. Тогда образы, имеющие в своем наборе *общие* нейроны, связаны *ассоциативно*. Такой процессор может быть описан уравнениями вида:

$$\frac{dH_i(t)}{dt} = \frac{1}{\tau_H}\left\{\Im_H(H_i,\beta_i) + \sum_{i\neq j}^{n}\Omega_{ij}H_j\right\}, \qquad (2)$$

где функция $\Im_H(H_i,\beta_i)$, определяющая внутреннюю динамику одного нейрона, определена в (1), Ω_{ij} –матрица внутрипластинных связей; $i,j = 1\ldots n$. Связи между нейронами, составляющими данный образ, модифицируются в процессе обучения. Закон обучения связей, предложенный самим Дж. Хопфилдом в [10], отвечает принципу «*отсечения лишнего*». Все связи изначально одинаковые (*сильные)*; в процессе обучения «нужные» связи *не меняются*, а связи нейронов образа со всеми другими нейронами пластины *вымирают* по закону

$$\Omega_{ij}^{\mathrm{Hopf}}(t) = \Omega_0\{1 - \frac{1}{2\tau^{\Omega}}\int_0^t [1 - H_i(t')H_j(t')]\zeta(t')dt'\}, \qquad (3)$$

где Ω_0 и τ^{Ω} – параметры обучения, функция $\zeta(t)$ обеспечивает эффект плавного вымораживания «посторонних» связей. Так достигается эффект «очищения образа»: если на такую пластину подается набор нейронов, не совсем соответствующий данному образу, под воздействием соседей «лишние» нейроны гаснут, а недостающие активируются. (Именно в этом заключается основное преимущество «распределенной» памяти.) Этот процессор был придуман как *инструмент распознавания* уже выученных образов.

Можно, однако, предложить и другой вариант обучения связей, известный в нейрофизиологии как «правило Хебба» [35]:

$$\Omega_{ij}^{\mathrm{Hebb}}(t) = \frac{\Omega_0}{4\tau^{\Omega}}\int_0^t [H_i(t') + 1][H_j(t') + 1]\zeta(t')dt', \qquad (4)$$

где Ω_0, как и в уравнении (3), – характерная величина этих связей; τ^{Ω} – характерное время обучения; функция $\zeta(t)$ здесь обеспечивает эффект «насыщения» (связи не могут усиливаться до бесконечности). Здесь связи между *активными* нейронами, изначально слабые, усиливаются («*чернеют*») в процессе обучения, формируя новый образ, а связи с неактивированными нейронами ($H_i = -1$) остаются слабыми. Этот процесс отвечает именно записи нового образа (новому обучению). Обратим специальное внимание на тот факт, что если какой-либо объект предъявляется системе в течение *только* короткого промежутка времени, то связи в соответствующем ему образе остаются относительно слабыми («серыми»).

Отсюда следует естественный вывод: для *записи новой* информации необходимо использовать процессор Хопфилда со связями, обучаемыми «по Хеббу» (4), а для *сохранения* выученных образов тот же процессор, но обучаемый «по Хопфилду» (3).

Для *кодирования* образной информации, т.е. конвертирования *образа* (набор M нейронов) в *символ* (один нейрон на более высоком уровне иерархии) используется *процессор локализации*, т.е. процессор «типа Гроссберга» [11] с нелинейным (конкурентным) взаимодействием. Такой процессор может быть описан уравнениями вида:

$$\frac{dG_k(t)}{dt} = \frac{1}{\tau_G}\{[-(\alpha_k - 1)G_k + \alpha_k G_k^2 - G_k^3] - \qquad (5)$$
$$- \sum_{l\neq k}^{n}\Gamma_{kl}G_kG_l\} \equiv \frac{1}{\tau_G}\left\{\Im_G(G_k,\alpha_k) - \sum_{i\neq k}^{n}\Gamma_{kl}G_kG_l\right\},$$

где G_k – переменные формальных динамических нейронов типа Гроссберга; $k = 1\ldots n$. Для удобства дальнейшего представления уравнение записано так, что стационарные состояния нейронов: активное $G = +1$ (активное) и $G = 0$ (пассивное). Параметры τ_G – характерное время активации и α_k – порог активации (регулирует *конкурентоспособность* данного нейрона). Здесь, по аналогии с (1), введена функция, регулирующая поведение уединенного «символьного» нейрона $\Im_G(G_k, \alpha_k)$.

Процесс формирования символа можно представить следующим образом. Вначале на процессор $\boldsymbol{G}$ подается *образ,* т.е. активируется

тот же набор M нейронов, что и на «образной» пластине $\boldsymbol{H}$. Эффект выбора одного нейрона из M активных достигается путем изменения связей Γ по закону:

$$\frac{d\Gamma_{kl}(t)}{dt} = -\frac{1}{\tau^{\Gamma}}\{G_k G_l (G_k - G_l)\}, \qquad (6)$$

где τ^{Γ} – характерное время выбора победителя. Исследования модели в работах [21,36] показали, что в симметричном случае α_k $(t = 0) = \alpha$ и $\Gamma_{lk} = \Gamma_{kl} = \Gamma(t = 0) = \Gamma_0$, процесс выбора символа *неустойчив*. Это значит, что *малейшее* (случайное!) преимущество одного из активных нейронов провоцирует (в результате их нелинейного взаимодействия) его экспансию и подавление остальных. Таким образом, реализуется парадигма Кохонена [12] – «победитель получает все». Важно подчеркнуть, что *какой именно* нейрон станет символом данного образа, заранее предсказать нельзя, это решает сама пластина в процессе выбора символа. Именно так обеспечивается *индивидуальность* искусственной системы. В этом смысле процесс формирования символа представляет собой яркий пример возникновения *условной* информации в данной системе.

После того, как данный G-нейрон получил статус символа и сформировал межпластинные связи со своим образом (см. ниже), он выводится из конкурентной борьбы за право стать символом какого-либо другого образа. Это достигается путем *параметрической* модификации нейрона-символа: $\alpha_k \to \alpha_k(f(\{H_i\}))$. Фактически на временном масштабе $t >> \tau^{\Gamma}$ нейрон-символ прекращает свое конкурентное взаимодействие с соседями, но приобретает возможность участвовать в *кооперативном* взаимодействии с другими символами («свободные» G-нейроны могут только конкурировать).

Следует отметить еще один важный момент. Кодирование (формирование символа) означает одновременно *осмысление* поступившей извне информации. Сам факт возникновения символа означает, что система восприняла данный набор из M активных нейронов на процессоре $\boldsymbol{H}$ как описание *одного реального объекта* и присвоила ему свой символ («имя»). Поэтому межпластинные связи символа с его образом мы далее называем *семантическими*.

АРХИТЕКТУРА КОГНИТИВНОЙ СИСТЕМЫ

Система уравнений для описания взаимодействия нейронов*. Согласно *общим* принципам ДТИ вся система должна состоять из двух подсистем – для генерации и рецепции информации. По причинам, которые будут понятны далее, будем называть эти подсистемы **ПП** (правая подсистема) и **ЛП** (левая подсистема), а соответствующие переменные будут (как правило) снабжены индексами «**П**» и «**Л**». Каждая из подсистем должна содержать одинаковый набор нейропроцессоров. Число самих процессоров (уровней иерархии) N не конкретизируется и не лимитируется: они могут возникать «по мере необходимости». Внутрипластинные и межпластинные связи обеспечивают взаимодействие нейронов на данной пластине и пластин между собой. Такая система может быть описана уравнениями вида:

$$\frac{dH_i^0(t)}{dt} = \frac{1}{\tau_H}[\Im(H_i^0, \beta_i(G_i^{\Pi,\sigma})) + \sum_{i \neq j}^{n} \Omega_{ij}^0 H_j^0 + \sum_k \Psi_{ik}^{\Pi,0} G_k^{\Pi,1}] + Z(t)\xi_i(t) + \Lambda^{\Л \to \П}(t) \cdot H_i^{\text{typ}}, \qquad (7)$$

$$\frac{dG_k^{\Pi,\sigma}(t)}{dt} = \frac{1}{\tau_G}[\Im_G(G_k^{\Pi,\sigma}, \alpha_k(\{\Psi_{ki}^{\Pi,(\sigma-1)}\})) - \sum_{i \neq k}^{n} \Gamma_{kl} G_k^{\Pi,\sigma} G_l^{\Pi,\sigma} + \sum_{\nu} \sum_{m}^{n} \Xi_{km}^{\Pi,\sigma} G_m^{\Pi,(\sigma-\nu)} + \sum_{m}^{n} \Psi_{km}^{\Pi,(\sigma-1)} G_m^{\Pi,(\sigma-1)} + \sum_{i \neq k}^{n} \Phi_{kl}^{\Pi,\sigma} G_l^{\Pi,\sigma} + \sum_{m}^{n} \Psi_{km}^{\Pi,(\sigma+1)} G_m^{\Pi,(\sigma+1)}] + Z(t)\xi_k(t) + \Lambda^{\Л \to \П} G_k^{\Л,\sigma}, \qquad (8)$$

$$\frac{dH_i^{\text{typ}}(t)}{dt} = \frac{1}{\tau_H}[\Im_H(H_i^{\text{typ}}, \beta_i(G_i^{\Л,\sigma})) + \sum_{i \neq j}^{n} \Omega_{ij}^{\text{typ}} H_j^{\text{typ}} + \sum_k^n \Psi_{ik}^{\Л,0} G_k^{\Л,1}] + \Lambda^{\П \to \Л} H_i^0, \qquad (9)$$

* Схема когнитивной системы, построенной на этих принципах, разрабатывалась и обсуждалась в работах [21,22]. Здесь приводятся кратко основные позиции.

$$\frac{dG_k^{Л,\sigma}(t)}{dt} = \frac{1}{\tau_G}[\Im_G(G_k^{Л,\sigma}, \alpha_k(\{\Psi_{ki}{}^{Л,(\sigma-1)}\})) + \sum_{v}\sum_{m}^{n}\Xi_{km}^{Л,\sigma}G_m^{Л,(\sigma-\nu)} + \sum_{i}^{n}\Psi_{ki}^{Л,(\sigma+1)}G_i^{Л,(\sigma-1)} + \sum_{k\neq l}^{n}\Phi_{kl}^{Л,\sigma}G_l^{Л,\sigma} + \sum_{m}^{n}\Psi_{km}^{Л,(\sigma+1)}G_m^{Л,(\sigma+1)}] + \Lambda^{П\to Л}G_k^{П,\sigma}, \quad (10)$$

где динамические переменные $H_i(t)$ и $G_k^{\sigma}(t)$ относятся к «образным» и «символьным» *динамическим формальным нейронам* (их внутренняя динамика задается функциями $\Im_H$ и $\Im_G$, определенными в уравнениях (1) и (5) соответственно) и изменяются в пределах $-1 \le H \le 1$ и $0 \le G \le 1$; $i, k = 1,\ldots n$, где n – число нейронов на пластине. Индекс σ определяет *уровень иерархии* символьной пластины: $\sigma = 1\ldots \boldsymbol{N}$, где $\boldsymbol{N}$ – полное число пластин в системе. Для записи иногда удобно считать, что образные пластины $\boldsymbol{H}$ можно рассматривать как пластины $\boldsymbol{G}$ нулевого уровня $\sigma = 0$: $\boldsymbol{H} \equiv \boldsymbol{G}^0$.

Уравнения (7), (8) относятся к подсистеме **ПП**, а уравнения (9), (10) – к **ЛП**. Они отличаются наличием случайного компонента (*шума*) $Z(t)\xi_i(t)$ в (7), (8), где $Z(t)$ – амплитуда шума, $0 < \xi_i(t) < 1$ – случайная функция, определяемая (например) по методу Монте-Карло. Кроме того, конкурентные связи Γ отсутствуют в уравнениях для $G^{Л,\sigma}$, поскольку процесс *выбора* символа, требующий участия шума, происходит только в **ПП**; символы в **ЛП** появляются благодаря переносу из **ПП** при помощи связей $\Lambda^{П\to Л}$ (последний член в уравнении (10)).

Связи в **ПП** обучаются «по Хеббу» [35], т.е. аналогично уравнению (4) для внутрипластинных образных связей: $\Omega^0 \equiv \Omega^{Hebb}$. Кооперативные связи нейронов-символов $\Phi^{П,\sigma}$ обучаются по тому же принципу:

$$\frac{d\Phi_{lk}^{П,\sigma}(t)}{dt} \propto \frac{\Phi_0}{\tau^{\Phi}}G_l^{П,\sigma}G_k^{П,\sigma}, \quad (11)$$

где Φ_0 – характерная величина; τ^{Φ} – характерное время обучения. Эти связи обеспечивают возможность создавать «обобщенный» образ, или *образ-из-символов* на любом σ-*м* уровне иерархии. Это – устойчивая общность символов, не соответствующая какому-то реально «увиденному» объекту, а возникающая в самой системе благодаря тому, что эти символы по каким-то причинам (либо ассоциативно, либо вынужденно) активируются одновременно. Именно эти связи позволяют вводить на высоких уровнях иерархии *символы-понятия*, не связанные с определенными реальными образами и представляющие собой «абстрактное знание».

Межпластинные связи Ψ обучаются также по правилу Хебба. Они обеспечивают *семантическое содержание* символа, связывая каждый символ на уровне σ с его образом на предыдущем уровне σ–1:

$$\frac{d\Psi_{ki}^{П,(\sigma-1)}(t)}{dt} \propto \frac{\Psi_0}{\tau^{\Psi}}G_k^{П,\sigma}G_i^{П,(\sigma-1)}, \quad (12)$$

и символ-участник нового образа с новым (старшим по иерархии) символом:

$$\frac{d\Psi_{km}^{П,(\sigma+1)}(t)}{dt} \propto \frac{\Psi_0}{\tau^{\Psi}}G_k^{П,\sigma}G_m^{П,(\sigma+1)}. \quad (13)$$

Такие связи мы называем *семантическими*; параметры Ψ_0 и τ^{Ψ} определяют характерную величину и время формирования этих связей. Благодаря ним символ на уровне σ приобретает, образно говоря, «ноги и руки»: связи $\Psi_{ki}^{\sigma-1}$ создают базу («опору») символа на предыдущем уровне, а связями $\Psi_{ki}^{\sigma+1}$ он «дотягивается» до новых (m-х) символов на следующем уровне. Этот алгоритм порождает квазифрактальную структуру системы, поскольку повторяется на всех уровнях иерархии. Заметим, что символы высокого уровня абстракции реже компонуются в образы, так как имеют слишком разный смысл. Поэтому структура на больших масштабах приобретает характер «дерева» (иерархическая геометрия).

Рассмотрим отдельно специальный вид взаимодействия, который порождает «вынужденные» *ассоциативно-иерархические* связи. Символы на разных (произвольных) уровнях иерархии связаны, если они имеют K общих нейронов-прародителей на низшем уровне, т.е. на образных пластинах:

$$\frac{d\Xi_{mk}^{П,\sigma}(t)}{dt} \propto \frac{\Xi_0}{\tau^{\Xi}}G_k^{П,\sigma}[\sum_{i}^{K}(H_i^0+1)]G_m^{П,(\sigma-\nu)}; \quad \nu = 0,\sigma, \quad (14)$$

где параметры Ξ_0 и τ^{Ξ} по-прежнему характерная величина и время формирования. Подчеркнем, что так определенная связь основана на *объективной* общности объектов, поэтому их символы ассоциативно связаны. Эти связи нарушают рисунок фрактальности, поскольку они имеют топологию типа «дерева»: чем выше уровень символа, тем с меньшим числом K

нейронов-признаков он связан и тем больше множество образов (в том числе, обобщенных), к нему относящихся.

Наконец, еще один результат развития и обучения системы – параметрическая модификация динамических формальных нейронов, которые реально участвуют в формировании информационной структуры всей когнитивной системы. Один из механизмов параметрической модификации – воздействие символов высокого уровня иерархии на соответствующие им образные нейроны: $\beta_i \to \alpha_i(G^{\sigma}_{\{i\}})$. Прежде всего, это касается так называемых *символов класса,* т.е. символов, порожденных не *образом* объекта, а набором *общих признаков* какого-либо класса объектов. Активация такого символа не может возбудить все «свои» образы, но переводит их в «ждущий режим» за счет понижения порога активации общих образных нейронов β_i. Так эти образы получают *преимущественное право* на активацию. Иными словами, символ класса обеспечивает *внимание* ко всем образам, к нему относящимся (в определенном смысле это – некоторая *гиперповерхность*).

Все сказанное имеет прямое отношение и к параметрам α^{σ}_k символьных нейронов: k-й нейрон на пластине $\boldsymbol{G}^{\sigma}$, становясь участником нового «обобщенного» образа, играет роль *образного* нейрона для всех тех старших по иерархии символов $G^{\sigma+\nu}_{\{k\}}$, к которым он имеет отношение, так что $\alpha^{\sigma}_k \to \alpha^{\sigma}_k(G^{\sigma+\nu}_{\{k\}})$. Кроме того, как обсуждалось выше, после формирования семантического содержания символа, т.е. связей с его образом $\Psi^{(\sigma-1)}_{ik}$, происходит модификация нейронов-символов $\alpha^{\sigma}_k \to \alpha^{\sigma-1}_k(\{\Psi^{(\sigma-1)}_{ik}\})$, которая выводит его из *конкурентных* взаимодействий и включает *кооперативные*. Этот фактор обеспечивает сложную многоуровневую активность нейронов-символов и оставляет «за кадром» те G-нейроны, которые символами не стали.

Таким образом, полная модификация нейрона-символа, отражающая «историю» всех его взаимоотношений с другими нейронами («опыт»), может быть выражена в форме $\alpha^{\sigma}_k \to \alpha^{\sigma}_k(\{\Psi^{(\sigma-1)}_{ik}\},\ G^{\sigma+\nu}_{\{k\}})$.

Для подсистемы **ЛП** образные связи $\Omega^{\text{typ}} \equiv \Omega^{\text{Hopf}}$ определяются уравнением (3). Все остальные связи $\Psi^{\text{Л},\sigma}$, $\Xi^{\text{Л},\sigma}$, $\Phi^{\text{Л},\sigma}$ имеют тот же смысл, что и в **ПП**, но обучаются «по Хопфилду», т.е. аналогично уравнению (3). Параметрическая модификация тех нейронов, которые принимают участие во взаимодействиях любого характера (т.е. приобретают некоторый «*опыт*») происходит так же, как и в **ПП**.

Взаимодействие двух подсистем обеспечивают связи $\Lambda(t)$. Они не обучаются, а *включаются* в зависимости от стадии обучения (или решения задачи) для передачи активности в соответствующую подсистему. Все процессы, требующие генерации новой информации, – формирование нового образа или символа – происходят в **ПП**, где присутствует шум**.** Затем результат этого процесса передается в **ЛП** при помощи прямых межподсистемных связей: $\Lambda^{\text{П}\to\text{Л}} = 1$ (при этом $\Lambda^{\text{Л}\to\text{П}} = 0$). Обратные связи $\Lambda^{\text{Л}\to\text{П}}$ должны включаться в уже обученной системе в том случае, если поступающая извне информация оказывается неизвестной системе, т.е. *новой*. Тогда система должна пройти фазу дообучения, главную роль в которой играет **ПП**.

Заметим, что уравнения (7)–(10) с первого взгляда кажутся сложными и громоздкими, однако они имеют простой и ясный смысл. Эти уравнения представляют собой скорее *язык*, при помощи которого можно описать характер и схему возможных взаимодействий. Опираясь на них, можно делать выводы об общих принципах строения архитектуры когнитивной системы.

Элементарный акт обучения. Выделим «элементарный акт» обучения, или общий элемент, на котором строится архитектура когнитивной системы (см. рис. 1). Этот процесс происходит по принципу «почернения связей» и проходит в два этапа. На первом этапе (рис. 1а) *образ* формируется на пластине уровня $(\sigma-1)$ вплоть до того, как кооперативные связи (Ω^0 в случае $\sigma = 1$ или $\Phi^{\text{П},\sigma}$ на других уровнях) становятся, в соответствии с уравнением (4), достаточно сильными («черными»). Тогда он передается прямыми (необученными) межпластинными связями ψ на пластину $G^{\text{П},\sigma}$ и одновременно, при помощи прямых межподсистемных связей Λ, на пластину того же уровня $G^{\text{Л},(\sigma-1)}$ в **ЛП**. На следующем этапе (рис. 2б) новый символ формируется вместе с его *семантическими* межпластинными связями $\Psi^{\text{П},(\sigma-1)}$. Снова по принципу почернения связей $\Psi^{\text{П},(\sigma-1)}$ должны стать достаточно «черными», и только после этого новый символ передается в **ЛП**. Здесь связи $\Psi^{\text{Л},(\sigma-1)}$ обучаются по правилу Хопфилда (3), после чего процесс завершен.

Схема архитектуры когнитивной системы. Схема архитектуры, разработанной в рамках ЕКП в работах [21,22], приведена на рис. 2. Она состоит из двух (подобных) подсистем: **ПП** (правая подсистема), содержащая шум, и

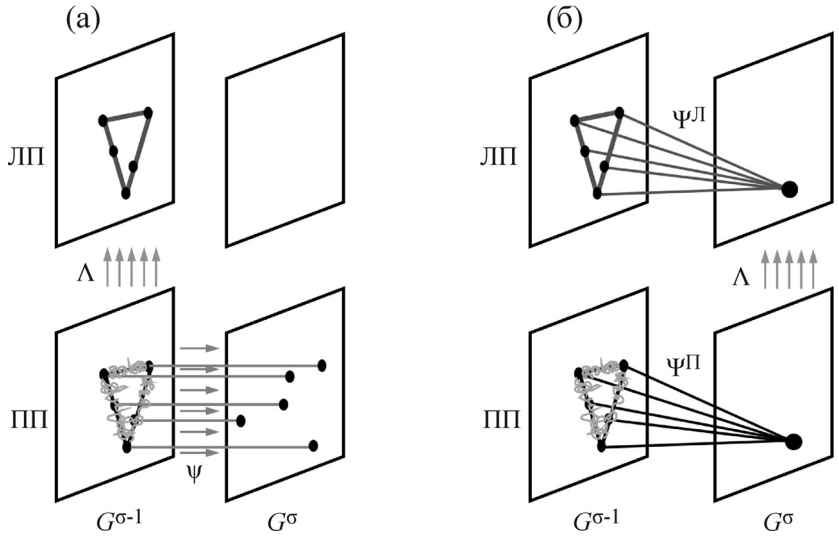

Рис. 1. Иллюстрация «элементарного акта» формирования символа следующего уровня: (а) – после формирования образа на пластине $G^{\sigma-1}$ он передается (при помощи прямых межпластинных связей ψ) на пластину следующего уровня G^{σ} в **ПП** и, одновременно (межподсистемными связями Λ) на пластину того же уровня $G^{\sigma-1}$ **ЛП**; (б) – после того, как *семантические* межпластинные связи между выбранным символом и его «образом» $\Psi^{\Pi,(\sigma-1)}$ обучатся до достаточно «черного» состояния, символ передается в **ЛП**, где семантические связи $\Psi^{Л,(\sigma-1)}$ обучаются по Хопфилду.

ЛП (левая подсистема), свободная от шума. Термины выбраны так, что условно эти подсистемы можно соотнести с полушариями головного мозга, а связи между ними Λ – с *corpus callosum*. При этом наличие шума в **ПП** обеспечивает генерацию, т.е. производство *новой* информации и *обучение*, а **ЛП** отвечает за *рецепцию* и обработку *уже известной* (выученной) информации. Такая специализация следует только из принципов ДТИ, и тот факт, что наш (теоретический) вывод совпадает с выводами практикующего психолога Э. Голдберга [17], является приятным сюрпризом и косвенным подтверждением адекватности ЕКП.

Все связи в **ПП** обучаются по правилу Хебба [35]: изначально слабые, связи усиливаются

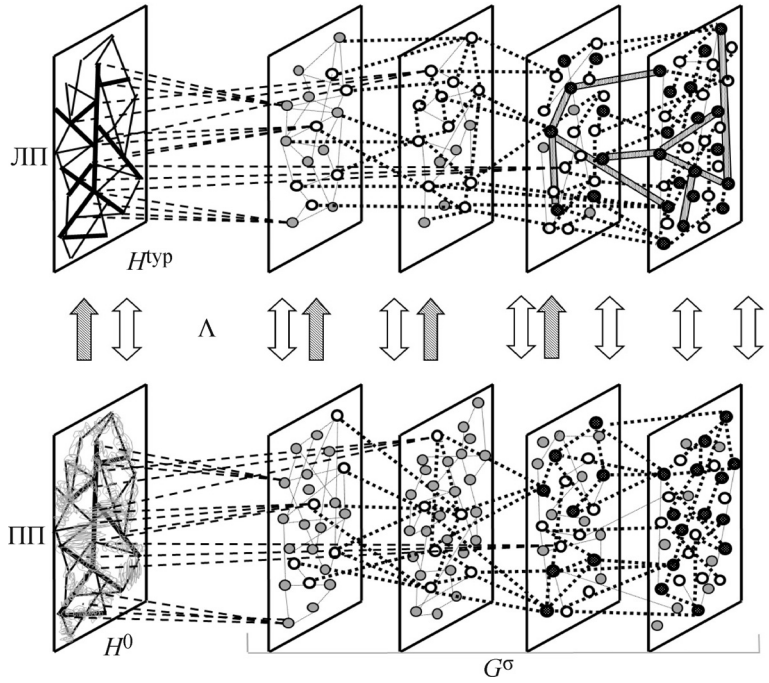

Рис. 2. Схема ЕКП-архитектуры когнитивной системы (пояснения в тексте).

(«чернеют») в процессе обучения вплоть до некоторого порогового значения, после чего соответствующий объект передается в **ЛП**. В **ЛП**, напротив, связи обучаются согласно оригинальному принципу Хопфилда [10] «отсечение лишнего»: все информативные связи, изначально сильные («черные»), не меняются в процессе обучения, а связи между нейронами образа и «посторонними» нейронами пластины постепенно вымирают. Иными словами, в **ПП** происходит *выбор,* а в **ЛП** – *отбор*.

Вся система представляет собой сложную многоуровневую блочно-иерархическую структуру, которая *сама* развивается (на рисунке – слева направо), путем самоорганизации по принципу «почернения связей», т.е. на каждом уровне повторяется (реплицируется) элементарный акт, представленный на рис. 1. Заметим, что данное представление (геометрия) архитектуры выбрано только из соображения удобства презентации: пластины (процессоры) могут располагаться не параллельно друг другу, как на рис. 2, а последовательно, вдоль некоторой поверхности (но тогда межпластинные связи трудно изображать). Новые уровни появляются «по мере необходимости», т.е. когда на предыдущем уровне формируется новый образ. В физике такой принцип построения системы называется термином «скейлинг», или *фрактальная структура.*

На низшем уровне ($\sigma = 0$) находятся пластины типа Хопфилда, содержащие образную информацию. При этом пластина $\boldsymbol{H}^0$ в **ПП** содержит *всю* образную информацию, поступающую в данную систему через «органы чувств», т.е. рецепторы. Внутрипластинные связи варьируются от «серых» (слабых) до «черных» (сильных). Функция этой пластины – запись *новых* образов (т.е. обучение). Роль «серых» связей мы обсудим ниже.

Пластина $\boldsymbol{H}^{\text{тур}}$ в **ЛП** содержит информацию, *отобранную* для хранения (запоминания). Она «заполняется» по мере обучения (с учителем, роль которого играет пластина $\boldsymbol{H}^0$) теми образами, связи которых достаточно почернели (*типичные* образы); все другие (серые) связи отмирают. Именно она играет главную роль в распознавании уже знакомых образов.

На следующем уровне ($\sigma = 1$) формируются (в **ПП**) *символы типичных образов,* которые несут семантическую нагрузку, т.е. *осознание* того факта, что данная цепочка активных нейронов описывает *один реальный* объект. Семантическое содержание (смысл) такого символа – его *декомпозиция* (при помощи семантических межпластинных связей Ψ) в свой образ, соответствующий этому реальному объекту. Только после формирования достаточно «черных» связей Ψ такой символ передается в **ЛП**.

На этом же уровне происходит *начальная вербализация*, т.е. возникают слова как *наименования* «выученных» объектов. Эти наименования придумываются в **ПП**, т.е. выбираются произвольно и индивидуально, понятно только самой системе. Если при этом **ЛП** получает извне (от «Учителя») информацию об *общепринятом* наименовании данного объекта, «внутреннее» имя вытесняется (после некоторого конфликта) общепринятым (такой процесс рассматривался подробно в работе [20]). Именно так происходит обучение речи у детей.

Также на этом же уровне символы кооперируются и создают (в **ПП**) *обобщенные образы* («образ-из-символов»), которые получают свой символ на следующем уровне ($\sigma + 1$). Эти образы достаточно примитивны, поскольку привязаны к конкретным объектам. Однако даже на этом уровне поэт может создать образ из примитивных слов, вместе рождающих яркую и запоминающуюся картину («*ночь, улица, фонарь, аптека*, …»).

На последующих уровнях ($\sigma > 1$) процесс повторяется, при этом степень «абстракции» вновь создаваемых образов нарастает. Это значит, что новый обобщенный образ трудно привязать к какому-либо объекту и объяснить на образном уровне (т.е. на примерах).

На верхних уровнях иерархии ($\sigma >> 1$) возникает *абстрактная информация* – инфраструктура символов и связей между ними, не опосредованных образами, т.е. нейронами-прародителями пластин Хопфилда. Здесь рождаются *символы-понятия*, не связанные с какими-либо образами (например, *совесть*, *бесконечность, красота, число* и т.д.). Такая информация возникает в обученной системе как результат взаимодействия всех пластин (не «чувственное», а «выводное» знание). Именно такая информация может быть *вербализована,* т.е. выражена в *символьной* форме при помощи *условленного* в данном обществе *языка*. При этом сам язык, т.е. правила связи символов-слов (грамматика и синтаксис), воспринимается (рецептируется) извне непосредственно в **ЛП**, где и хранится. Именно эти уровни обеспечивают возможность *коммуникации* с аналогичными системами, т.е. возможность передавать свою условную информацию («объяснять словами») и воспринимать семантическое содержание символьной (*вербальной*) информации, поступающей извне. Кроме того, здесь **ЛП** получает возможность воспринимать *новую* информацию не только от **ПП**,

но и извне, в символьной форме, от внешнего «Учителя». Такие знания в психологии называют «семантическими», в отличие от «эпизодических», которые система (**ПП**) получает в процессе обретения собственного индивидуального опыта.

Таким образом, вся система *растет* от нижних образных уровней информации через *семантическую* информацию (понятную лишь данной индивидуальной системе) к верхним уровням абстрактной информации, которая уже может быть вербализована и *распространена* (понята) в данном сообществе. В этом процессе на каждом этапе формирования нового уровня происходит одно и то же: связи возникают в **ПП** и отрабатываются там до состояния *черных*; только после этого новый символ передается в **ЛП**. При этом часть информации (записанная «серыми» связями) теряется, точнее, не переходит на следующий уровень, а остается на предыдущем в роли *служебной* или *скрытой* информации данной индивидуальной системы.

Интерпретация понятий интуиции, подсознания, сознания и логики. Если под *интуицией* понимать случайное, спонтанное, неаргументированное решение, «прямое усмотрение истины» по Канту [37] без всякого рассуждения и доказательства, – очевидно, ее источником является **ПП** (более точно – шум). Для нее типично *неосознанность пути*, ведущего к ее результату. Если под *логикой* понимать все причинно-следственные непрерывные цепочки связей, – к ней относятся все процессы, происходящие в **ЛП**. В этом смысле сохраняются выводы, сделанные в нашей ранней работе [38], где символьная инфраструктура еще не рассматривалась.

Однако если подходить к этим понятиям более строго, чисто *логическим мышлением* следует считать, по Далю, «правильное доказуемое рассуждение». Отсюда сразу следует, что к нему относятся только *вербализованные* (следовательно, доказательные и общепонятные) рассуждения, причем термин «правильное» означает, что они должны быть основаны на *общепринятых аксиомах*. Тогда между «чистой логикой» и «чистой интуицией» должны существовать другие, промежуточные, алгоритмы мышления.

Та же ситуация возникает с понятиями *сознания* и *подсознания*: если определить *сознание* как «способность осмысленно воспринимать окружающее, ясное понимание чего-либо» (см. *Философский словарь*), то оно возникает только *после вербализации*.

Подсознание определяют как «совокупность процессов, в отношении которых отсутствует *субъективный контроль*» (определение из Википедии), т.е. оно должно опираться на накопленную (хаотично) информацию, которая *не имеет символов*, следовательно, не может быть активировано извне, в символьной форме (т.е. словами).

Опираясь на сказанное выше, можно говорить об интуиции, логике, сознании и подсознании более содержательно.

Предлагаемая архитектура содержит много ($N >> 1$) уровней. Низшие уровни представляют собой *служебную*, или *внутреннюю индивидуальную* информацию данной системы, «вещь в себе». Только вербализованная информация, которая появляется на верхних уровнях иерархии, является *осознанной* в общепринятом смысле (не только индивидуально). На вопрос «как мозг делает мысль?» мы теперь можем попытаться ответить. Поскольку речь есть *последовательный ряд символов*, именно она и формирует (выделяет) то, что называется «мысль» из всего набора паттернов мозговой активности. По образному выражению Т.В. Черниговской «язык – это то, как наш мозг говорит с нами» [40]. Тогда «*сознание*» можно определить как способность системы выстраивать когнитивную активность в последовательный контекстный ряд при помощи речи, таким образом оценивая свое состояние. Главную роль в этом процессе играет ЛП**.**

Выше было показано, что при переходе от каждого предыдущего уровня к последующему часть информации теряется, точнее переходит в разряд «скрытой» (служебной) для данной системы. Рассмотрим ее подробнее.

Наиболее глубокий уровень скрытой информации представляют слабые «серые» связи на образных пластинах – их роль состоит в том, чтобы хранить «случайную» (т.е. ту самую «хаотично накопленную») информацию, которая когда-то может оказаться важной. Эта информация не переходит ни в **ЛП**, ни на уровень G^1, поэтому не может ассоциироваться ни с каким символом, т.е. остается *не осознанной* и *не подконтрольной* системе. Это именно то, что выше было определено как «подсознание». Такая («серая») цепочка может активироваться только благодаря шуму («вдруг увидеть внутренним взором»), что можно интерпретировать как *озарение* (в современной англоязычной терминологии «*moment aha*»).

При переходе от семантической информации к вербализованной остается множество символов, для которых стандартных слов не существует – это некие цельные «картинки», описание которых требует декомпозиции, т.е. один внутренний символ может быть описан при помощи многих слов. Вербализация этой

информации требует не *озарения,* а *подбора* нужных слов. Это всегда возможно, но не всегда просто. На языке теории распознавания этот процесс называется «формализацией экспертного знания».

Таким образом, *скрытая* информация имеет разные уровни глубины, что существенно влияет на усилия по ее извлечению на уровень *сознания*. Выводы, основанные на *скрытой информации*, естественно интерпретировать как *интуитивное* мышление (insight). Отметим, что в данной схеме большинство скрытой информации сосредоточено именно в подсистеме **ПП**.

К *логическому* мышлению, в соответствии с определением Даля, естественно отнести оперирование *вербализованными понятиями и абстрактными связями*, причем лишь теми, которые считаются *установленными* (общепринятыми) в данном социуме. Такие связи присутствуют в **ЛП**, причем только на высоких уровнях иерархии. Сама абстрактная информация имеет собственные уровни и инфраструктуру, которая нарабатывается постепенно, по мере эволюции системы («с годами»). Эту развитую инфраструктуру, объединяющую верхние уровни и **ПП**, и **ЛП**, можно ассоциировать с *мудростью*. Из сказанного ясно, что мудрость больше, чем логика.

Специфика элементов наиболее ярко проявляется при решении задач, связанных с определением сходства/различия объектов. Такие задачи решаются на уровне образных пластин *автоматически*: сходство определяется общими нейронами, различие – разными, и система это *знает*. Однако это знание *не осознано*, пока общие/разные нейроны не выражены через комбинации внутренних символов, тогда *служебно-образное* знание («ощущение») может перейти в *семантическое*. Последующая *вербализация* знания означает выстраивание абстрактных связей внутренних символов со словами. Полученный ответ верен для данной системы (индивида), но может быть ошибочен объективно, поскольку способ записи образной информации индивидуален. Решение, полученное таким образом, интуитивно: оно основано на всем опыте, т.е. «картине мира» индивида. Оно не должно *доказываться* (самой системе это не нужно, она просто знает, что это так). Однако вербализованное решение может быть *объяснено* и *аргументировано*. Если аргументы верны с точки зрения общепринятых аксиом, это уже является доказательством его верности – по сути, это и есть метод «перевода интуитивного знания в логическое».

Решение задач. Рассмотрим кратко, каким образом в приведенной схеме могут решаться основные задачи мышления.

Распознавание. Задачи распознавания решаются в уже обученной системе, обладающей по крайней мере развитыми нижними уровнями $\sigma = 0,1,...$ и ставятся в образных пластинах. Постановка задачи сводится к активации в размытом множестве $\boldsymbol{H}^0$ в **ПП** некоторого набора нейронов (*экзаменуемый* образ), который записывается там «как есть», т.е. вначале слабыми («серыми») связями. Далее этот образ передается прямыми межподсистемными связями на пластину типичных образов $\boldsymbol{H}^{\text{typ}}$ в **ЛП**, после чего возможны варианты.

– Если экзаменуемый объект хорошо известен системе, т.е. полностью *совпадает* с одним из типичных образов, он сразу (быстро!) ассоциируется с соответствующим символом со всеми вытекающими последствиями в смысле положения в иерархии; при этом **ПП** более участия в процессе не принимает.

– Если образ в достаточной степени схож с каким-либо из имеющихся типичных образов (попадает в его «область притяжения»), он также воспринимается как уже известный, обладающий своим символом $G^{\text{Л},1}$. Однако при этом правильность распознавания нуждается в *проверке*, для чего символ передается в **ПП**, где декомпозируется и результат сравнивается с экзаменуемым образом. Таким образом, возникает *петля* по сути аналогичная той, которая предлагается в работах В.Г. Яхно [8]:

$$\begin{array}{c}\boldsymbol{H}^{\text{typ}}(\textbf{ЛП}) \rightarrow \boldsymbol{G}^{\text{Л},1}(\textbf{ЛП}) \\ \uparrow \Lambda^{\text{П}\rightarrow\text{Л}} \quad \downarrow \Lambda^{\text{Л}\rightarrow\text{П}} \\ \boldsymbol{H}^{0}(\textbf{ПП}) \leftarrow \boldsymbol{G}^{\text{П},1}(\textbf{ПП}).\end{array}$$

Здесь роль «гипотез» и «моделей» из [8] играют *символы*, сформированные в самой системе.

Если результат сравнения удовлетворителен (что может быть оценено, например, по «невязке», как и в [8]), то новый объект ассоциируется с уже существующим символом. Если же результат не удовлетворителен, невязка провоцирует повторение, и процедура проходит нескольких итераций. При этом объект в размытом множестве постепенно «чернеет».

– Если запись объекта в **ПП** доходит до стадии «черных связей» ($\Omega^0_{ij} \geq \Omega^{\text{th}}$), а пластина типичных образов так и не идентифицировала его, он становится *новым типичным образом* и записывается на пластине $\boldsymbol{H}^{\text{typ}}$, после чего получает свой *символ*, который связывается с символами более высокого уровня, и.т.д.

Рис. 3. Функциональные зоны неокортекса (взято из [40]).

Из сказанного ясно, что данная система способна обработать и распознать даже *новый* объект, но только с участием размытого множества $\boldsymbol{H}^0$.

Прогноз. Распознавание динамического процесса во времени (прогноз) требует большего развития системы, чем распознавание объекта, а именно участия символов более высокого уровня, нежели символы образов ($\sigma \geq 2$). Динамический процесс распознается по «кадрам», т.е. фиксированное состояние объекта в данный момент времени t_i, соответствует одному образу, который получает собственный символ. Для того, чтобы система *поняла*, что цепочка образов, накопившаяся от момента t_0 до t_n есть, например, динамическая трансформация одного и того же объекта, необходимо, чтобы цепочка соответствующих символов образовала достаточно устойчивые (*черные*) связи так, чтобы сформировался символ более высокого уровня – символ *процесса*. Декомпозиция этого символа дает *образ процесса* – траекторию в фазовом пространстве символов; при этом декомпозиция каждого символа дает соответствующую «картинку». Тогда при предъявлении одной «картинки» система будет ожидать появления следующей. Отметим, что прогноз может оказаться ошибочным, если в какой-то момент извне поступает противоречащая ему информация. Тогда процедура запоминания процесса в **ПП** должна быть проведена заново.

Таким образом, данная система после обучения способна решать задачи распознавания, прогноза событий (по прецедентам), восприятия речи и понимание семантического содержания символов на высоких иерархических уровнях. Это возможно:

– при наличии необходимых (отработанных) алгоритмов и полной информации (логическое мышление, активно **ЛП**);

– при недостатке информации и/или алгоритма (сначала активируется **ЛП**, а недостаток восполняется «воображением», т.е. **ПП**);

– при наличии *противоречивой* информации (разрешение парадоксов, творческое мышление) – активно **ПП**, причем его нижние образные уровни, особенно «размытое множество», т.е. пластина $\boldsymbol{H}^0$.

СОПОСТАВЛЕНИЕ РЕЗУЛЬТАТОВ ЕКП С ДРУГИМИ ТЕОРЕТИЧЕСКИМИ ПОДХОДАМИ И ЭКСПЕРИМЕНТОМ

Анатомические данные. Согласно современным представлениям (см., например, [40]) кору головного мозга можно (условно) разделить на зоны, которые отвечают за зрение, слух, моторную активность, абстрактное мышление, в частности, речь (см. рис. 3).

Этот рисунок представляет собой фактически *зеркальное отражение* нашей схемы, представленной на рис. 2, если расположить процессоры (уровни иерархии) не параллельно друг другу, а последовательно, вдоль некоторой поверхности (что, как уже отмечалось, ничему не противоречит). При этом низшим (образным) уровням схемы соответствуют затылочные доли и так называемая *ассоциативная* кора. Височные

доли (зоны Вернике и Брока) отвечают за *слух* (восприятие слов) и воспроизведение (произнесение слов), но не за саму *речь*. На нашем языке это значит, что отдельные символы становятся словами на некоторых промежуточных уровнях $0 < \sigma < N$. Функцию речи, т.е. связной и осмысленной передачи информации при помощи слов, относят к фронтальным зонам (лобным долям), что в нашей схеме соответствует верхним уровням иерархии (абстрактное мышление). Моторные функции в нашей схеме не рассматривались, но очевидно, что они требуют достаточно развитой системы символов без «глубокой философии», т.е. ответственность за них должны нести средние уровни иерархии, как это и показано на рисунке.

Подчеркнем, что при построении архитектуры мы руководствовались только общими принципами ДТИ, и тот факт, что расположение функциональных зон мозга «повторяет» наши представлении о порядке формирования уровней информации, оказывается опять же приятным сюрпризом.

Другой аспект экспериментальных данных связан со *строением* самой коры. Со времен эксперимента [41], подтвержденного и современными данными, известно, что вся кора головного мозга (независимо от функций) представляет собой один слой, состоящий из *колонок*, расположенных вертикально к поверхности. Колонка возбуждается (или гаснет) целиком, т. е. ведет себя как формальный нейрон. Колонка состоит из шести тонких слоев, отличающихся морфологически. При этом функции слоев и причина именно такого строения неокортекса до конца не ясны.

Формальный нейрон в ЕКП – наиболее естественный кандидат на роль колонки. В рамках ЕКП можно предположить, что слои колонки выполняют функцию проверки и подтверждения поступающего сигнала, подобно пластинам Рекседа [42] в спинном мозге (их роль подробно обсуждалась в [43]). Тогда колонка должна возбуждаться при поступлении двух (или более) сигналов, из которых второй подтверждает (или дополняет) первый. Это свойство биологически оправдано, поскольку обеспечивает «ответственность» колонки за свое поведение: случайный, т.е. не подтвержденный, сигнал не возбуждает колонку. Заметим, что аналогичным свойством обладает иммунная система.

Теория графов. В подходе, основанном на теории графов [26], принимается, что формальный нейрон есть узел графа, связи – ребра графа. Известны *сетевая* топология графа (всеобщая связность) и иерархическая топология типа «дерева». Теория графов позволяет систематизировать эмпирические данные.

Архитектуре когнитивной системы в ЕКП соответствует граф следующей структуры. В общем графе выделены автономные области (блоки), внутри которых узлы и ребра образуют квазисеть. Совокупность блоков образует граф, в котором узлы – блоки и ребра – связи. Такой граф имеет на больших масштабах топологию типа «дерева», что необходимо в иерархической системе. Такая структура не противоречит теории графов, но дополняет ее применительно к когнитивной системе.

Заметим, что с точки зрения теория графов наша архитектура «не красива»: она не симметрична и не отвечает какому-то одному определенному стилю – фрактал, или «дерево», или сеть. Все эти элементы присутствуют в схеме ни рис. 2, причем «удельные веса» этих элементов могут варьироваться в довольно широких пределах в индивидуальных системах. Однако в рамках теории графов нельзя ответить на вопросы, *почему* и *как* образовался тот или иной узел (ребро).

Концепция когнитома. Понятие «когнитома» как системы реализованных степеней свободы, было введено К.В. Анохиным [27]. Она основана на понятии «*кога*», который представляет собой блок, состоящий из двух «гиперплоскостей». На одной имеется образ (набор нейронов), на другой – соответствующий ему «особый элемент», или «элемент опыта». Последний может рассматриваться как «нейрон», но отличающийся от нейронов образа морфологически, биохимически и т.д. «Когнитом» есть иерархически организованный набор «когов», способный решать задачи мышления. Уровни иерархии соответствуют сложности (и содержательности) решаемых задач.

С точки зрения ЕКП «когнитом» отличается от нашей архитектуры лишь терминологически и согласуются, если отождествить понятия: «ког» = «блок формирования символа (обобщенного) образа», «протоког» = «символ класса», «гиперплоскость» = нейропроцессор (пластина), «элемент опыта» = «символ», «когнитом» = вся архитектура когнитивной системы. Отличия между образными и символьными нейронами (морфологические и т.д.) в ЕКП не рассматривается, но вполне допускается. Таким образом, эти подходы также не противоречат, а дополняют друг друга.

Возможности экспериментальной проверки. Экспериментальная техника, позволяющая детектировать активность различных участков головного мозга, бурно развивается. При помощи *инвазивных* методов (вживление электродов в

мозг) были получены весьма интересные данные. Так, было показано, что при приобретении нового опыта (на нашем языке – при формировании нового символа) определенные участки мозга подопытного животного изменяются морфологически (активизируется ген, отвечающий за мутации), причем эти участки индивидуальны [44,45]. Отметим, что эти данные полностью согласуются с нашими представлениями о *параметрической* модификации нейронов, ставших символами какого-либо образа («опыта»). Однако инвазивные методы, широко применяемые в опытах на мышах и других млекопитающих, по отношению к людям практически невозможны.

Неинвазивные методы картографии мозга (функциональная магнито-резонансная томография в сочетании с детекцией кровотока и электроэнцефалографией) также развиваются весьма интенсивно (см. [16] и ссылки там же). Уже сейчас достигнутое разрешение (как пространственное, так и временное) достаточно для того, чтобы определять, какие именно функциональные зоны мозга активны в данный момент, и, более того, проследить *динамику* этой активности.

В рамках ЕКП мы можем предложить идею эксперимента, направленного на картографию *творческого* когнитивного процесса. В качестве «подопытного» приглашается человек творческий с высоким уровнем интеллекта. Ему предлагается сформулировать задачу, одну из тех, которые он должен решать в рамках своей профессиональной деятельности. Предлагать задачу извне не целесообразно, поскольку решение ее безответственно. На основе ЕКП можно предсказать динамику возбуждений участков коры головного мозга «подопытного» в течение решения задачи.

На первом этапе возбуждаются лобные доли (преимущественно **ЛП**). Это – попытка решить задачу «логически».

Второй этап: возбуждаются затылочные и средние части (преимущественно **ПП**). Это – попытка решить задачу «интуитивно», на основе прецедентов. Этот этап наступает, если задача логически не решается.

Третий этап – возбуждаются затылочные зоны **ПП** – попытка решить задачу на основе редких прецедентов, хранящихся в «размытом множестве». Этап наступает, если первые два оканчиваются неудачно.

Четвертый этап – квазихаотическое (в пространстве и времени) возбуждение как лобных, так и затылочных участков обоих полушарий. Этот этап соответствует состоянию «муки творчества».

Пятый этап (*если* он наступает) – ярко вспыхивает вся кора головного мозга. Это значит, что решение принято и сформулировано на общепринятом языке (т.е. настало «озарение»). Если этого не происходит, «муки творчества» продолжаются далее (возможно и после окончания эксперимента).

В этом сценарии важны крупномасштабные возбуждения. При детализации сценария (мелкомасштабная визуализация) должна проявиться индивидуальность когнитивного процесса.

Подчеркнем, что подтверждение (или опровержение) этого сценария является решающим экспериментом проверки адекватности ЕКП. Допустимое разрешение приборов сейчас не позволяет детектировать активность одного нейрона (что представляется наиболее интересным), однако локализовать активность в достаточно малой области, соответствующей в ЕКП одному процессору, вполне возможно.

ЗАКЛЮЧЕНИЕ

Таким образом мы можем заключить, что архитектура когнитивной системы, полученная в рамках ЕКП, не противоречит другим рассмотренным подходам, но в определенном смысле дополняет их. Благодаря выбранной технике динамических дифференциальных уравнений мы можем проследить динамику и мотивацию формирования *символов* (или *графов*, или «*когов*»). Заметим, что предложенная архитектура имеет прямое отношение и к *искусственному интеллекту*: в данной схеме именно **ЛП** играет эту роль, но без **ПП** его формирование и эволюция не понятны.

Выделим ряд ключевых моментов ЕКП, отличающих его от других нейроморфных подходов:

– использование *континуальных* представлений нейропроцессоров;

– разделение всей системы на *две подсистемы* – для генерации и рецепции информации. Условно эти подсистемы можно соотнести с правым и левым церебральными полушариями (**ПП** и **ЛП**), а связи между ними $\Lambda(t)$ – с *corpus callosum*. **ПП** отвечает за обработку *новой* информации, а **ЛП** – за работу с хорошо известной, что полностью согласуется с выводами Э. Голдберга [17];

– учет случайной компоненты («*шума*»), который присутствует только в **ПП**;

– *неустойчивый* характер процесса формирования символа, в результате чего результат

оказывается *непредсказуемым*; именно этот фактор обеспечивает *индивидуальность* искусственной когнитивной системы;

– *самоорганизация* нейронного ансамбля по принципу «почернения связей»;

– *параметрическая модификация* тех нейронов, которые реально участвуют в формировании всей архитектуры (т.е. являются элементами некоего «*опыта*»).

Именно эти особенности позволяют воспроизвести особенности *человеческого* мышления – непредсказуемость, индивидуальность, непрерывное обучение, способность к интуитивному и логическому мышлению и т.п. В других подходах такие проблемы не ставятся вообще.

Следует особо подчеркнуть, что шум (случайная компонента) в рамках ЕКП представляется не неизбежной и досадной помехой (как в радиофизике, проблемах передачи информации и т.п.), а обязательным и *полноправным участником* всех процессов, связанных с генерацией информации. Отметим, что шум (применительно к живым системам – случайный, спонтанный, непредсказуемый поступок) и есть тот самый механизм *выживания*, который в обычных ситуациях мешает действовать столь точно и быстро, как это может робот, но в критической ситуации позволяет *случайно* найти совершенно неожиданный и непредсказуемый выход. Именно этот фактор может сделать искусственную систему «человекоподобной».

Отметим, что представленный вариант архитектуры не является единственно возможным и в рамках ЕКП. Так, предположение о том, что связи в **ЛП** обучаются «по Хопфилду» представляется естественным, но не необходимым условием. Кроме того, данный вариант может (и должен) развиваться и исследоваться далее, поскольку имеется еще много нерешенных вопросов. В частности, мы не касались здесь вопроса о взаимном влиянии *эмоций* и когнитивного процесса. Не рассматривался также *механизм* переключения межподсистемных связей Λ. Эти вопросы заслуживают отдельного исследования. Тем не менее, эти проблемы не влияют на уже сделанные здесь предсказания. Предложенный эксперимент может подтвердить (или опровергнуть) правомерность выводов ЕКП.

Подводя общий итог, можно сказать, что теоретическая когнитология активно развивается и переходит из чисто вербальной формы в математически оснащенную. Существующие теоретические подходы (упомянутые выше) не противоречат, а, скорее, дополняют друг друга. Каждый из них (и все вместе) позволяют интерпретировать существующие экспериментальные данные. Тем не менее теория мышления только сейчас подходит к тому уровню, на котором можно предложить постановку реального целенаправленного («*круциального*») эксперимента. Эта задача – одна из главных и актуальных в плане развития когнитивной науки.

СПИСОК ЛИТЕРАТУРЫ

1. Д. С. Чернавский, Успехи физ. наук **170** (2), 157 (2000).
2. Г. Р. Иваницкий, Успехи физ. наук **180** (4), 337 (2010).
3. А. С. Шамис, *Пути моделирования мышления* (Ком-Книга, М., 2006).
4. А. А. Жданов, *Автономный искусственный интеллект* (Бином. Лаборатория знаний, 2008).
5. J. E. Laird, *The Soar Cognitive Architecture* (MIT Press, 2012).
6. A. Samsonovich, Biol. Inspired Cognitive A*rchitectures,* № 6, 109 (2013).
7. Л. А. Станкевич, в сб. *Пути моделирования мышления,* под ред. В.Г. Редько (УРСС, М., 2014), СС. 262–305.
8. В. Г. Яхно, в *Тр. конф. «Нелинейная динамика в когнитивных исследованиях»* (Нижний Новгород, 2011), сс. 246–249.
9. М. И. Рабинович и М. К. Мюезинолу, Успехи физ. наук **180** (4), 370 (2010).
10. W. S. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).
11. J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
12. S. Grossberg, *Studies of Mind and Brain* (Riedel: Boston, 1982).
13. T. Kohonen, Neural Networks **1** (1), 303 (1988).
14. Е. Е. Витяев, в сб. *Пути моделирования мышления,* под ред. В.Г. Редько (УРСС, М., 2014), сс. 52–168.
15. E. M. Izhikevich and G. M. Edelman, Proc. Natl. Acad. Sci. USA **105** (9), (2008).
16. Л. В. Доронина-Амитонова, И. В. Федотов, А. Б. Федотов и др., Успехи физ. наук **185** (4), 371 (2015).
17. Е. Голдберг, *Парадокс мудрости* (Поколение, М., 2007).
18. V. L. Bianki, Neurosci. Behavi. Physiol. **14** (6), 497 (1984).
19. О. Д. Чернавская, Д. С. Чернавский, В. П. Карп и А. П. Никитин, Сложные системы **1**, 25 (2012).
20. О. Д. Чернавская, Д. С. Чернавский и др., Сложные системы **2,** 47 (2012).
21. О. Д. Чернавская, Д. С. Чернавский и др., в сб. *Пути моделирования мышления,* под ред. В.Г. Редько (УРСС, М., 2014), сс. 29–88.
22. O. D. Chernavskaya, et al., Biol. Inspired Cognitive Architecture **6,** 147 (2013).

23. Д. С. Чернавский, *Синергетика и информация: Динамическая теория информации* (УРСС, М., 2004).

24. А. А. Ежов и С. А. Шумский, *Нейрокомпьютинг и его применения* (МИФИ, М., 2008).

25. S. S. Haykin, *Neural Networks and Learning Machines* (Prentice Hall, 2009).

26. G. M. Edelman and G. Tononi, *A universe of consciousness: how matter becomes imagination* (Basic Books; New York, 2000).

27. К. В. Анохин, *Материалы XVII Всероссийской научно-технической конференции «Нейроинформатика–2015»*. (http://neuroinfo.mephi.ru/conf/Content/Presentations/Anokhin2015.pdf.).

28. A. L. Hodgkin and A. F. Huxley, J. Physiol. **117,** 500 (1952).

29. R. FitzHugh, Biophys. J. **1,** 445 (1961).

30. J. Nagumo, S. Arimoto, and S. Yashizawa, Proc. IRE, **50**, 2062 (1962).

31. H. Quastler, *The emergence of biological organization* (Yale University Press, New Haven, 1964).

32. А. Г. Колупаев и Д. С. Чернавский, Краткие сообщения по физике **1** (2), 12 (1997).

33. С. П. Курдюмов, Г. Г. Малинецкий и Д. С. Чернавский, *Режимы с обострением, джокеры, перемешивающий слой. Новый взгляд на нелинейную динамику* (ИПМ, 2005).

34. V. M. Verkhlyutov, V. L. Ushakov, P. F. Sokolov, and B. M. Velichkovsky, Psychology **7** (4), 4 (2014).

35. D. O. Hebb, *The organization of behavior* (John Wiley & Sons, 1949).

36. Д. С. Щепетов и Д. С. Чернавский, в сб. *Труды конф. «Нелинейная динамика в когнитивных исследованиях»* (Нижний Новгород, 2013), сс. 205–208.

37. И. Кант, *Критика чистого разума* (Мысль, М., 1994).

38. О. Д. Чернавская, А. П. Никитин и Д. С. Чернавский, Биофизика **54** (6), 1103 (2009).

39. Т. В. Черниговская, ЛОГОС **1** (97), 79 (2014).

40. М. И. Беляев, *Мышление. Мозг. Память.* © 2015 (http://milogy.net/mozg01.htm).

41. V. Muncastle, Brain **12**, 7 (1997).

42. B. J. Rexed, Comparative Neurology **96** (3), 415 (1952).

43. Д. С. Чернавский, В. П. Карп, И. В. Родштат и др., *Распознавание. Аутодиагностика. Мышление* (Радиотехника, М., 2003).

44. K. V. Anokhin, R. Mileusnic, I. Y. Shamakina, and S. P. R. Rose, Brain Res. **544** (1), 101 (1991).

45. O. O. Litvin and R. V. Anokhin, Neurosci. Behav. Physiol. **30** (6), 671 (2000).

A Natural-Constructive Approach to Modeling the Cognitive Process

O.D. Chernavskaya and D.S. Chernavskii

Lebedev Physical Institute, Russian Academy of Sciences, Leninskii prosp. 53, Moscow, 19991 Russia

We consider the natural-constructive approach to modeling a cognitive system. This approach is based on the Dynamical Theory of Information, nonlinear differential equation technique, and the concept of “dynamical formal neuron”. The version of an architecture of cognitive system elaborated within the natural-constructive approach is presented. An important constructive feature of this architecture consists in splitting up the whole system into two subsystems, which represent an analogue to the right and left cerebral hemispheres. It is shown that the noise necessarily presented in the right subsystem secures the conditions for generating new information. The interpretation of the concepts of intuition, logic, consciousness and sub-consciousness is discussed. The architecture of the natural-constructive approach is compared to other theoretical approaches (graph theory and the concept of “cognitom”) and anatomy data. The idea of an experiment is proposed to verify the main results of the natural-constructive approach.

Key words: information, image, symbol, noise, intuition, architecture