РОССИЙСКАЯ АКАДЕМИЯ НАУК

Физический институт имени П.Н.Лебедева

Ф И А Н

ПРЕПРИНТ

7

О.Д. ЧЕРНАВСКАЯ, Д.С. ЧЕРНАВСКИЙ,
А.П. НИКИТИН

# О ВОЗМОЖНОМ МЕХАНИЗМЕ «ИНТУИТИВНОГО» И «ЛОГИЧЕСКОГО» В НЕЙРОКОМПЬЮТИНГЕ

МОСКВА 2009

Аннотация

Предлагается принцип устройства нейрокомпьютера, позволяющий разрешить некоторые проблемы современного нейрокомпьютинга и описать эффекты интуиции, творчества, а также ряд других явлений, природа которых до сих пор не изучена. Основная идея заключается в том, что функции **обучения** и **сохранения информации** являются **комплементарными** и должны выполняться двумя разными подсистемами, причем подсистема, способная обучаться, обязана содержать «**шум**», т.е. **случайный** элемент. Обсуждается вопрос о том, насколько искусственный интеллект может имитировать человеческое мышление. Предлагается ряд методов, позволяющих имитировать в нейрокомпьютере эффекты, ранее считавшиеся присущими только живому организму (юмор, сон, и т.д.).

## On the possible mechanism of “intuitive” and “logical” in the neural computing

**O.D. Chernavskaya, D.S. Chernavskii, A.P. Nikitin**

The concept of the neural computer arrangement is proposed which allows to solve some problems of modern neural computing and to describe the effects of intuition, creation, as well as a set of effects still not understood sufficiently. The concept is based on the idea, that the activity of knowledge acquisition and of information store are complementary and should be performed by two different subsystems, with the one able to learn has to involve the “noise”, i.e., an occasional element. The problem is discussed to what extent the artificial intellect could emulate the living one. A set of possible methods are suggested in order to emulate the effects which traditionally were prescribed to the human body only (the sense of humor, the sleep and dreams, etc.)

## 1. Введение

Проблема имитации интеллекта живого (человека) искусственными методами, т.е. в электронном приборе или компьютерной программе, представляет чрезвычайный интерес и разрабатывается уже несколько десятилетий. Разумеется, эта проблема лежит на стыке многих наук — медицины, биологии, математики, электронных технологий, информатики и т.д., включая философию, — и решение ее требует интеграции всех этих разнородных и разно-языковых знаний. Особую роль при этом играют те научные направления, которые сами по себе представляют интеграцию традиционно выделяемых наук — теория распознавания, теория самоорганизующихся систем, синергетика.

На этом пути сделано уже очень много.

В рамках медицины и биологии изучены функции и цели отдельных участков головного и спинного мозга животных и человека (см., например [1], [2] и ссылки *ibid*).

На языке синергетики сформулированы (см. [3]) основные цели и задачи мышления — распознавание, прогнозирование, принятие решения; предложена модель целеполагания и т.д. Отметим, однако, что четкое определение процесса мышления, несмотря на многочисленные и упорные исследования (см., например, [5] и ссылки *ibid*), пока предложено не было.

В рамках нейроинформатики и математики предложены модели восприятия, запоминания, кодирования и передачи информации: например, процессор Хопфилда [6], процессор Гроссберга [7], карты Кохонена [8], и т.д. (см. также [8] и ссылки *ibid*)

Подчеркнем, что прорыв в данной области произошел благодаря концепции *обучения* компьютера вместо *программирования*, т.е. разработки детерминированных алгоритмов. Нейрокомпьютер, в отличие от обычного, воспринимает *информацию*, а не *команды*, и обрабатывает ее сам.

Компьютеры, построенные по этому принципу, оказались чрезвычайно успешными (существенно более, чем компьютеры в обычном понимании, даже сверхмощные) в решении определенных задач. Как правило, это задачи типа распознавания образа, сжатия и передачи информации и т.д. Каждое из этих направлений получило широкое прикладное применение и развивалось дальше во многом независимо от общей проблемы искусственного интеллекта. Можно сказать, что в какой-то степени реализована старая китайская сказка про братьев Ли: один из них умел перешагивать стены, второй — выпивать моря, и т.д., однако одного, усредненного брата, который умеет всего понемногу, среди них не было.

Таким образом, задача создания (и понимания) искусственного интеллекта все еще далека от окончательного решения. В частности, остается

еще масса очень интересных и *непонятых* проблем, причем как в живом интеллекте, так и в искусственном.

1). В человеческом мышлении принято разделять «логическое» и «интуитивное». Но что есть «интуиция» и что есть «логика»? Может ли нейрокомпьютер обладать интуицией?

2). Что есть «автопилот», а именно способность человека выполнять определенные действия, требующие анализа и прогнозирования ситуации (т.е. обратной связи) *как бы* бессознательно, не включая логико-аналитический аппарат? Например, умение ходить развивается очень рано и практически бессознательно, но умение водить машину в начале требует значительных умственных усилий, и, казалось бы, переключение сознания на какие-то другие задачи, невозможно, однако каждый опытный водитель знает, что именно так и происходит.

3). Что есть **сон**? Может ли «спать» нейрокомпьютер? К чему это может приводить?

4). Что есть «творчество», т.е. генерация новой информации? На каком этапе и где это происходит?

5). Что есть «чувство юмора»? Может ли «робот» им обладать?

6). Как можно интерпретировать понятия, которыми оперируют психотерапевты – подсознание, сверх-Я и т.д. ?

В данной работе предлагается принципиальная схема устройства нейрокомпьютера, способного имитировать явления, перечисленные выше. Мы также попытаемся ответить на вопрос о том, как можно модернизировать современные нейрокомпьютеры, чтобы максимально приблизить искусственный интеллект к живому, и до какой степени это возможно.

Основная концепция заключается в следующем:

- способность **обучаться** и способность **сохранять информацию** являются *комплементарными* (т.е. взаимодополняющими); эти функции должны быть *разделены* между двумя «устройствами»;
- для обучения принципиально необходимо наличие **шума**, т.е. *случайного, хаотического элемента*, присутствующего в «устройстве» обучения; для сохранения информации (без искажений) шум не нужен и даже вреден.

При условии выполнения этих двух условий все перечисленные эффекты, как будет показано ниже, получают простое и естественное объяснение.

Подчеркнем, что нашей конечной целью является компьютерно-ориентированная симуляция тех эффектов, о которых шла речь. Тем не менее мы будем постоянно обращаться к человеческому интеллекту и образу мышления как к эталону, опираясь при этом на главным образом на здравый смысл и повседневные наблюдения и лишь в минимальной степени на

специальные знания современной нейрофизиологии [1,2]. Это дает важное преимущество: можно «ставить эксперименты» с собственным мыслительным процессом и проверять выдвинутые гипотезы «на себе». Кроме того, существуют многочисленные, широко распространенные убеждения и предубеждения; ниже мы обсудим некоторые из них.

Формулируя цель данной работы кратко, можно сказать, что мы предлагаем, не вдаваясь в излишние подробности и технические детали, принципиальную схему того самого «сводного брата», который не обладает сверхъестественными возможностями в какой-либо одной сфере деятельности (и, таким образом, практическое применение вряд ли имеет), но обладает наиболее «человеческим лицом». Эта задача очевидно относится к разряду фундаментальных.

**2. Современные представления о принципах мышления и обучения.**

Сформулируем современные представления о принципиальной схеме работы человеческого мозга и, соответственно, нейрокомпьютера. Поскольку этой теме посвящена обширная литература (см., например, [1,2,3,4, 9] и ссылки *ibid)*, то сделаем это в сжатой форме.

Кора головного мозга представляет собой сложный фрактальный объект, напоминающий ореховую скорлупу. Она содержит активные элементы — *нейроны,* связанные между собой волокнами — *аксонами* и *дендритами* (выходными и входными связями). Передача сигнала между нейронами регулируется нейротрансмиттерами, химический состав которых определяет состояние системы (т.е., в конечном итоге отвечает за эмоции). Нейроны делятся на перцепционные, внутренние и эффекторные. Их специфика понятна из названий; ниже мы будем говорить только об имитации деятельности внутренних нейронов, тонкости восприятия информации (перцепционные нейроны) и передачи команд к действиям (эффекторные нейроны) не будут обсуждаться.

В свое время велась активная дискуссия о том, что есть нейрон: сравнительно простой бистабильный объект или самостоятельный «компьютер»? Победила первая точка зрения, и упор сейчас делается на *сложный нелинейный* характер связей между нейронами.

По мере *обучения (*смысл этого понятия обсудим чуть позже*)* нейроны активизируются («зажигаются») и, в конце концов, каждому воспринятому мозгом образу соответствует некая цепочка или карта из «горящих» нейронов. Естественно предположить, что для записи *схожих* предметов задействуются отдельные фрагменты цепочек, соответствующих уже запомненным похожим образам; это можно назвать «*принципом экономии зажженных нейронов*».

На последующих этапах обработки информации эти образы (карты) могут сворачиваться в более компактную форму, т.е. каким-то образом *кодируются* (например, по принципу процессора Гроссберга [7]).

Крайне важным является то обстоятельство, что связи между нейронами устроены таким образом, что они *укрепляются* по мере активизации. Иными словами, чем чаще или дольше человек видит какой-то образ (или вспоминает его), тем более четко этот образ «записывается», тем прочнее связи в цепочке, ему соответствующей.

Простая математическая модель подобной системы описывается уравнениями типа Ланжевена и имеет следующий вид (см. [4]):

$$\frac{du_i}{dt} = \varphi(u_i) + \sum_{i \neq j} \psi_{ij} u_j + ... + g(t) \cdot \xi(t), \qquad (1)$$

здесь $u_i$ — переменная, соответствующая состоянию *i-го* нейрона; функция $\varphi(u_i)$ описывает свойства нейрона, $\psi_{ij}$ описывает силы связи между нейронами. В простейшем случае нейрон — бистабильный элемент, например:

$$\varphi(u_i) = u_i - u_i^3 .$$

В стационарном состоянии $u_i = \pm 1$ (+1 соответствует «горящей лампочке», –1 — пассивное состояние»). Функции связи $\psi_{ij}(t)$ формируются в процессе обучения и в этом смысле зависят от времени, но в *обученном* нейрокомпьютере они практически постоянны. Уравнения для них гораздо сложнее и должны учитывать эффект мультипликативного усиления при активизации.

В известной модели этого типа, модели Хопфилда (см. [3–5]), связи заданы простейшим способом (симметричны, положительны, причем каждый нейрон стремится переключить соседний в «свое» состояние); в результате такой процессор умеет очень мало (например, только запоминать и различать буквы). Здесь мы не будем вдаваться в проблему конкретизации уравнений для связей. Наша задача — продемонстрировать на качественном уровне те *возможности,* которые содержит подобная модель.

Данная модель обладает одним значимым свойством — она может быть представлена в потенциальной форме:

$$\frac{du_i}{dt} = \frac{\partial \Omega(u_1, u_2, ....u_N)}{\partial u_i}; \; i = 1, 2, \ldots . N, \qquad (2)$$

где $N$ — число элементов системы; $\Omega(u)$ --- потенциал, который равен

$$\Omega = -\left[ \sum \left( \frac{1}{2} u_i^2 - \frac{1}{4} u_i^4 \right) + \sum_{i=j} \psi_{ij} u_i u_j \right] \qquad (3)$$

Эти уравнения имеют очень важную (в иллюстративном смысле) механическую аналогию: они описывают движение шарика единичной массы в потенциальном поле $\Omega$ в многомерном фазовом пространстве. Эта аналогия

позволяет использовать хорошо известные понятия и образы механики применительно к мышлению. На этом языке потенциальная функция, возникшая в результате обучения, представляет собой сложный ландшафт с лунками, каналами и водоразделами. Само обучение может рассматриваться как *формирование этого ландшафта*, причем, чем активнее работа с образом, тем глубже соответствующий ему канал.

Подчеркнем, что работа обученного компьютера (с уже сложившимся ландшафтом) в большой степени детерминирована — из канала шарик выйти уже не может. Эта проблема может быть решена в рамках процессора Больцмана, который соответствует включению эффекта *шума* — последний член в правой части уравнения (1) содержит случайную функцию $\xi(t)$, распределенную, например, по Гауссу; уровень (амплитуда) этого шума может, вообще говоря, зависеть от времени как $g(t)$. Это соответствует картине «шарик прыгает по ландшафту из канала в канал», причем то, *сколько времени* он проводит в *данном* канале и *как часто* перепрыгивает из канала в канал, управляется соотношением между частотой шума, его амплитудой и глубиной канала. Такое структурное решение убирает детерминированность, но порождает проблемы другого рода (каким образом такой процессор может прийти хоть к какому-то решению?). Ниже мы вернемся к этой проблеме.

Обсудим более подробно смысл слова «обучение». Именно этот процесс отличает нейрокомпьютер от обычного компьютера, работающего по заданной программе. В современных нейрокомпьютерах и схема обучения, и сама архитектура, т.е. способ компоновки нейронов, очень сильно зависят от рода задач, которые данное устройство призвано решать (см., например, [9] и ссылки *ibid*). Здесь мы обсудим только общие принципы обучения, применимые к возможно более широкому кругу задач.

Прежде всего, следует отличать «обучение образу» (или так называемому декларативному или фактологическому знанию) и «обучение действию» (т.е. так называемому «прескрипционному», предписывающему знанию). В первом случае это восприятие и запись (запоминание) некоторой внешней картины, образа; во втором — решение некоторой (ментальной!) задачи, результатом которого является какое-то действие.

Пока мы будем говорить только о первом варианте, «обучение образу».

Декларативные/фактологические знания подразделяются на *опытные* (называемые еще *эпизодические*), т.е. приобретенные самостоятельно, путем проб и ошибок, и *внедренные* (или *семантические*), т.е. то, чему «учат в школе»[1]. Последние передаются уже в закодированном виде и запоминаются (записываются) механически.

[1] Термины декларативное/прескрипционное, эпизодическое/ семантическое знание приняты в нейропсихологии, см. Элханон Голдберг

Как приобретаются опытные знания? Упрощенная схема процесса приобретения знаний может быть представлена следующим образом.

Для человека, точнее, ребенка, **первичное обучение** есть *отражение* в мозгу того, что он видит, т.е. *построение картины мира.* Сначала она *размыта*, затем, по мере изучения (т.е. тоже отражения) отдельных фрагментов, картина становится более *четкой*. Кроме того, она становится *детализированной,* т.е. построенной из этих фрагментов как детский конструктор. Тогда, попав в несколько другую обстановку (например, другую комнату) ребенок уже различает предметы, похожие на то, что ему знакомо и ассоциирует их друг с другом.

Если он видит что-то похожее на то, что видел раньше, этот предмет запоминается (как самостоятельный предмет) *легче* (так называемый «overlearning effect» [1]).

Постепенно из отдельных образов формируются *понятия*: «дом», «машина» и т.д.

Поясним: если человеку надо записать какую-то информацию, смысла которой он не понимает, он берет *неважно какой* кусок бумаги и записывает; только тогда, когда эта информация приобретает для него какой-то смысл, т.е. связь с другими фрагментами информации, возникает желание собрать чем-то связанные между собой бумажки в *одном* (тоже, в сущности, не важно каком) месте, в компактной форме (*сложить в папку*). Папки, относящиеся к одной тематике, объединяют в *директории.* Таким образом, мы описали простейшую схему иерархии записи, принятую в обычных компьютерах. Но, вообще говоря, принцип архивирования в человеческом мозге может быть иным.

Нейрокомпьютер обучается так же: те образы, которые воспринимают внешние рецепторы, преобразуется и записывается в виде цепочек/карт *вообще говоря случайно*. При первых предъявлениях тот образ, который записывается, может быть *искажен* и/или не точен, *размыт*. После нескольких предъявлений образ становится четким, а связи в такой цепочке укрепляются — она становится куском «ландшафта» (одним из каналов). На более высоком уровне цепочки, сходные по каким-либо признакам, объединяются в *«понятия*».

Описанный процесс в теории распознавания носит название «обучение без учителя» и наиболее близок к процессу первичного обучения ребенка. В нейрокомпьютинге этот метод применяется для достижения наиболее компактного способа записи («сжатие данных») для последующей их передачи.

Существует еще и «обучение с учителем»: результат (решение, эталон) известен, конечная цель обучения есть воспроизведение образа наиболее близко к эталону. В нейрокомпьютинге такая схема обучения применяется в рамках задачи распознавания образа. Поскольку наша задача — представить

себе наиболее общий принцип записи информации, мы далее будем иметь в виду оба этих способа обучения.

В рамках такой картины становится понятным и даже тривиальным вопрос о том, что есть *ассоциации.* Если один и тот же нейрон задействован в двух цепочках/картах, соответствующих разным образам, эти образы ассоциативно связаны.

Что есть в этой картине *память* и где она хранится?[2] В сущности, память — это и есть тот самый ландшафт, который был построен в результате всей жизнедеятельности человека или нейрокомпьютера (обучение, работа, специальный тренинг и т.д.). Память «включается», если по каким-то причинам (звук, запах и т.д.) активизируется пусть даже один нейрон, принадлежащий какому-то образу; тогда активируется вся цепочка, а соответствующий образ «всплывает в памяти».

Интересен и противоположный процесс — что значит *«забыть»*? Естественно предположить, что те связи, которые *давно не активизировались*, слабеют («каналы мелеют»)[3]. Как правило, их удается восстановить, привлекая близкие образы (т.е. «проходя по побочным каналам») и/или *историю* их формирования (чтобы вспомнить, куда что-то положил, надо представить себе, как это делал). Возможен и другой механизм забывания — по мере последующего обучения нейроны канала вовлекаются в другие образы, т.е. данный канал в «ландшафте» пересекается множеством других, его очертания размываются и идентификация затрудняется[4]. В этом случае информацию можно восстановить только при повторе (ситуация «дежа вю» [10]).

Отметим, что сказанное относится только к «опытным» (приобретенным или эпизодическим) знаниям; внедренные знания забываются, если однажды забылись, бесповоротно.

Из сказанного становится также ясно, что «внедренные» знания носят вторичный характер и становятся осмысленными только при соотнесении с «опытом». Человек, который никогда не видел пальмы, но прочел ее описание, вполне может понять, что это, если он видел другие деревья; в противном случае он может запомнить это описания только механически, как набор бессмысленных знаков (выучить «наизусть как стишок»). Без соотнесения эти знания скорее носят именно модульный характер.

---

[2] Этот вопрос был крайне актуальным в нейрофизиологии середины XX века: тогда считалось, что память носит «модульный» характер, т.е. отдельные образы записываются компактно и связи друг с другом не имеют; сейчас возобладала противоположная точка зрения, парадигма «распределенной памяти» [1], которой придерживаемся и мы.

[3] В нейрофизиологии он носит название «trace decay», постепенное ослабление следов [10]

[4] В нейрофизиологии он называется «торможением следов вследствие побочного интерферирующего воздействия» [10]

Человек, далекий от биологии, воспринимает фразу из учебника биологии «Генотип только тогда проявляется в фенотипе, когда рецессивная аллель гомозиготна» исключительно как заклинание, но вполне может ее запомнить как, скажем, музыкальную фразу.

Обсудим несколько подробнее, что значит «*понятие*», т.е. то, каким образом отдельные схожие образы свертываются в один обобщающий. В нейрокомпьютинге рассматривается процессор Гроссберга [7], соответствующий принципу *автолокализации* образа, т.е. записи достаточно полной информации в одном (в пределе) нейроне (так называемый «портрет бабушки»). Формирование понятия «собака» традиционно рассматривают именно как автолокализацию образов Жучек, Баронов, Рексов … в одном слове «собака». Однако в когнитологии (науке о сознании) «собака» представляется как набор нескольких общих черт (и соответственно, нейронных «кусков») вышеупомянутых конкретных собак (см. [2]). На языке «ландшафта» формирование понятий естественно представить как систему *мостов и развязок* над естественным ландшафтом, которая может быть сколь угодно сложной (многоуровневой). Этот образ гораздо ближе к «когнитологическому» понятию; в дальнейшем мы будем использовать именно его. Но необходимо подчеркнуть, что проблема *кодирования* отнюдь не тривиальна и требует самостоятельного исследования.

Что есть собственно «*мышление*»? Пока речь по сути дела шла только об обучении; что же происходит, когда уже обученный человек или робот начинают мыслить? Как реализуются главные задачи мышления?

Очевидно что, когда речь идет о решении какой-либо задачи, процесс должен иметь начало и конец. На языке ландшафта это значит, что шарик должен *найти путь* в какую-то лунку и там *остаться.* Первая часть задачи предполагает неявно возможность *поиска*, т.е. случайных попыток в шумовом поле; вторая, напротив, неявно предполагает, что шум закончился, наступила ясность. Каким образом это противоречие может быть разрешено?

Эти вопросы мы обсудим ниже.

Кроме того, мы не касались вопроса о том, как происходит обучение *действию?* Что есть автопилот? Как можно интерпретировать понятия «логического» и «интуитивного» в мышлении? Мы полагаем, что естественное объяснение эти проблемы получают лишь в том случае, если принять другую, несколько более сложную, схему ментального процессора, где функции ***обучения*** *и* ***запоминания разделены****.*

## 3. Принципиальная схема устройства нейропроцессора с разделенными функциями

Как видно из предыдущего изложения, обучение, особенно, первичное, неразрывно связано со случайным элементом, шумом, благодаря которому устанавливаются первичные связи в цепочке. Но для *запоминания образа*, т.е. для того, чтобы данная цепочка стала, после нескольких повторных предъявлений, устойчивым куском ландшафта*, шум вреден*, так как приводит к порождению близких, но *разных* цепочек связей.

Принимая во внимание эти соображения, мы предлагаем следующую принципиальную схему. Существуют две системы описанного выше типа, строение которых одинаково, но одна из них умеет **обучаться**, другая — **сохранять** плоды обучения первой. Для удобства и определенности первую систему будем называть «правым процессором (ПП)», а вторую — «левым процессором, ЛП» (имея, конечно, в виду аллюзию с полушариями головного мозга).

Принципиальную роль здесь играет **шум,** т.е. случайные флуктуации состояния нейронов, которые неизбежно существуют в любом живом (или даже просто *реальном*) организме. Для того чтобы система умела *учиться,* необходимо, чтобы уровень шума был достаточно высок по сравнению с силой связей; чтобы умела *работать,* т.е. сохранять, реализовывать и передавать уже полученные знания, шум должен быть минимален. При этом обе системы «очень внимательно смотрят друг на друга», т.е. *могут непрерывно обмениваться информацией*.

В человеческом мозге число связей полушарий друг с другом, *corpus colossum*, на два порядка больше, чем рецепторов, связывающих мозг с внешним миром [1,9].

Такую схему можно реализовать, рассматривая две системы «сопряженных» переменных

$$\frac{du_i}{dt} = \varphi(u_i) + \sum_{i \neq j} \psi_{ij} u_j + g(t) \cdot \xi(t) + \sum \chi^R_{ij}(t) v_j(t), \qquad (4)$$

$$\frac{dv_i}{dt} = \varphi(v_i) + \sum_{i \neq j} \psi^L_{ij} v_j + \sum \chi^L_{ij}(t) u_j(t), \qquad (5)$$

где переменные $u$ и $v$ относятся к ПП и ЛП, соответственно. Связи в первом приближении считаем одинаковыми для двух подсистем, т.е. $\psi^R_{ij} = \psi^L_{ij} = \psi_{ij}$ Они, разумеется, зависят от времени, т.е. меняются в процессе обучения; но в уже обученных системах они почти постоянны (по модулю появления новых связей при дообучении или творчестве, эти вопросы обсуждаются в следующем разделе).

По сравнению с процессором Больцмана (член, соответствующий шуму), эти уравнения непременно содержат перекрестные члены, отвечающие за «внимание к другому процессору», $\chi^R_{ij}(t)$ и $\chi^L_{ij}(t)$; их зависимость от времени

нетривиальна и изменяется в процессе решения задачи. Их характер мы пока не конкретизируем; более подробное обсуждение см. в следующем разделе.

Кроме того, должна существовать еще *как минимум одна* структура, связанная с обеими выше названными и принимающая *решение*, какая из 2-х систем «права». В частности, она ответственна за *действия*, т.е. запускает двигательный аппарат. В человеческом организме (см. [1]) существует *мозжечок,* достаточно *«старая»* (в эволюционном смысле) структура. Говоря о движениях, мы будем апеллировать именно к ней, не конкретизируя, как именно принимается решение о данном действии — эта проблема выходит за рамки данной работы.

Функционирует такая система следующим образом.

**При первичном обучении** ЛП всегда является ***ведомым***, а ПП — ***ведущим.***

***Обучение образу = отражение***

Когда системе предъявляется какой-то объект/образ, он случайным образом записывается в обеих системах, причем эта запись может *искажать* объект и/или зафиксировать его *не четко*. При этом цепочка, созданная в ПП, *не прочна* по сравнению с уровнем шума и *не запоминается, стряхивается из-за шума*. При вторичном предъявлении в ПП, опять же благодаря шуму, создается несколько другая цепочка[5]; в ЛП записываются *обе*. После нескольких предъявлений все цепочки в ЛП, будучи наложены друг на друга, формируют некий «*усредненный» образ* (возможно, с потерей неких нюансов), достаточно сильный даже для того, чтобы противостоять шуму в ПП. Тогда он передается уже из ЛП в ПП и записывается (=запоминается) там так же[6]. В качестве первого приближения предположим, что эти записи, т.е. связи, формирующие тот самый ландшафт (картина, обсуждавшаяся применительно к одному процессору), в обеих системах остаются одинаковыми (имея, однако, в виду, что это предположение не принципиально и может быть подвергнуто ревизии). Все сказанное можно отнести и к внедренным, «семантическим» знаниям, которые записываются в закодированном виде.

***Обучение неизвестному действию***

Прежде всего заметим, что знания о действиях не могут (при всем желании) быть *внедренными,* они всегда *опытные* (эпизодические) — в противном случае не было бы уникальных *мастеров* в какой бы то ни было отрасли человеческой деятельности — спорте, изготовлении скрипок, балете,

---

[5] Совсем другой она быть не может благодаря принципу «экономии зажженных нейронов», см.выше

[6] Возможно, однако, что какие-то из тех слабых цепочек, которые возникают в ПП при однократном предъявлении предмета и соответствуют *нюансам* могут иногда «всплывать» в памяти как некое трудно уловимое воспоминание. Заметим, что здесь уже роли меняются, и ведомым становится ПП.

и т.д. Так что имитировать на уровне нейрокомпьютера имеет смысл только обучение методом «попыток». Кроме того, здесь «обучение с учителем» и «без учителя» отличаются не принципиально: учитель может только помочь советом, но не передать свой *внутренний код*.

Итак, пусть ставится задача воспроизвести какое-либо действие. Внешний сигнал или задача из обучающего множества поступает на рецепторы обеих систем. Первая система (ПП) срабатывает в точности так, как было описано выше, т.е. включение некоего нейрона приводит к спонтанному образованию некоей цепочки/карты активных нейронов, соответствующей предъявленному образу. Вторая система, ЛП **повторяет** действия первой до тех пор, пока не наступает *конфликт в рамках поставленной задачи*.

Под конфликтом мы будем понимать ощутимое несоответствие полученных результатов ожидаемым. Само понятие требует более пристального внимания, сейчас же мы будем апеллировать к интуитивному пониманию этого термина. Если, например, речь идет о человеке, который учится ходить, конфликт — это падение.

В случае наступления *конфликта* все связи, возникшие в ПП, «*сбрасываются*», т.е. память о них теряется. Этот эффект достигается именно за счет того, что ПП находится в случайном шумовом поле, которое «трясет» нейроны, а связи, которые могут между ними возникнуть, *не сильнее этого шумового воздействия*. В ЛП эти связи сохраняются (поскольку шума здесь практически нет), а та связь, которая привела к «неправильному шагу», т.е. конфликту, *блокируется* (на языке ландшафта это означает возникновение в этом месте *барьера)*.

Тогда, при второй попытке решения поставленной задачи, в ЛП строится та же цепочка, что и при первой попытке, вплоть до последнего шага. Что делать дальше ЛП не знает, а смотрит на ПП. В ПП цепочка может выстраиваться совершенно *так же*, как при первом шаге (1), а может и *по-другому* (2).

(1) Неверный шаг приводит к блокированию еще одной связи в ЛП, верный — к продолжению и укреплению этой цепочки в ЛП и блокированию следующей связи, приводящей к конфликту.

(2) ЛП *повторяет* новую попытку ПП, несмотря на то, что «проторенная» цепочка с укрепленными связями уже существует. Это казалось бы парадоксальное поведение было бы невозможно для одной, самостоятельной, системы нейронов, но естественно в рамках предлагаемой схемы: как уже говорилось, *на этапе обучения ЛП является ведомым, ПП — ведущим*. Это означает, что на этом этапе ЛП непременно «отражает» то, что происходит в ПП, запоминая при этом ту промежуточную информацию, которая в ПП «стряхивается», чтобы не мешать поиску. Шаг, приводящий к конфликту, разумеется, блокируется.

На каком-то этапе ПП находит решение задачи, т.е. активирует ту (*возможно, не единственную*) цепочку нейронов, которая приводит к желаемому результату. Эта цепочка, разумеется, «прописывается» в ЛП. После того, как эта цепочка укрепилась, благодаря повторным активациям, настолько, что превысила ***порог*****,** который отвечает уровню шума в ПП, она прописывается (как и в случае с обучением образу, см. выше) и в ПП, и таким образом запоминается уже в обеих подсистемах. С этого момента *управление* передается ЛП, где данная цепочка укрепляется дальше.

Что при этом происходит с предыдущими, (возможно, неудачными) попытками? По-видимому, они сохраняются в виде «бледных» (неглубоких) каналов для того, чтобы в нештатной ситуации (например, сломанная нога) можно было вспомнить, а как еще можно пытаться, пусть не красиво, медленно, но ходить.

***Если обе системы уже обучены***

Внешний сигнал (образ, задача) поступают одновременно на обе подсистемы. В ЛП он активирует уже существующую, *заведомо надпороговую* цепочку нейронов и далее идет по наиболее «сильным» цепочкам вплоть до выполнения задачи. В том случае, когда эта цепочка становится столь сильной, что превышает некоторый второй, более высокий порог, она автоматически передается в структуру, отвечающую непосредственно за движение (например, в «мозжечок») и запускается *независимо от того, что происходит в ЛП и ПП*, при поступлении сигнала, коррелированного с данной цепочкой[7].

Это и есть ***автопилот***, т.е. не ***поиск решения***, а ***запуск*** уже существующей и отлаженной ***программы***. Ее действие уже не связано напрямую с процессами, происходящими *ни в ЛП, ни в ПП*. Однако важно отметить, что при наступлении некоторой *опасной* или *сомнительной* ситуации ***ЛП*** немедленно активирует соответствующую цепочку, и правильность действий уже проверяется *по шагам*, т.е. *осознанно* в терминах устойчивых канонических образов: *программа* подвергается *ревизии*.

Что при этом должно происходить в ПП и ЛП?

Каждый опытный водитель знает, что когда включается «автопилот», ЛП и ПП ведут себя *почти* так, как если бы данная внешняя задача (в обсуждаемом случае — управление автомобилем) *снята*, и переключаются на совершенно другие, внутренние проблемы. Слово «почти» означает, что «боковое зрение» и «боковое сознание» тем не менее следят за процессом и при необходимости немедленно включается весь арсенал накопленных когнитивных (т.е. сознательных) умений.

«Боковое сознание» естественно связать с активностью ПП, где могут включаться и переключаться цепочки, связанные с происходящим процессом

[7] Коррелированный, например, как в случае эффекта «правая нога в пол» у **пассажира**, имеющего опыт вождения машины

лишь отдаленно, т.е. происходит «свободное блуждание по близким и дальним связям».

В случае *нештатной*, *критической* ситуации, когда проверенная и даже переданная автопилоту цепочка приводит (или может привести) к конфликту, т.е. катастрофе, управление автоматически передается именно ПП (за неимением лучшего), и, если ПП имело достаточно возможностей «свободного поиска», оно может найти совершенно неожиданное решение и среагировать адекватно. В противном случае, т.е. если система будет упорно следовать выработанной *программе*, возникает эффект «робот упорно топает в пропасть».

Все сказанное можно отнести не только к таким «примитивным» умениям, как хождение, управление автомобилем и т.д. Любые часто повторяющиеся ситуации=задачи приводят к выработке неких *стереотипов поведения*[8].

*Как* вырабатываются эти стереотипы, т.е. каким образом уже достаточно обученная, *опытная* система решает некую *новую* задачу? Иными словами, как происходит процесс «мышления»? Именно этот вопрос мы рассмотрим ниже.

**4. Мышление: интуиция, логика, творчество.**

Как мы уже отмечали, четкого определения понятия «мышление» нет. Эта проблема обсуждалась очень давно и активно (см., например, [3–5] и ссылки *ibid*). Были сформулированы основные цели и задачи мышления – распознавание, прогноз развития ситуации, принятие решения и т.п. Очевидно, эти задачи имеют *разные уровни иерархии*. Поясним.

Известно, какие именно отделы мозга, обоих его полушарий, отвечают за определенные функции [1]: задние отделы — зрение, височные доли — речь и т.д. Существуют еще и *кортикальные лобные доли* (гораздо более «молодая» в эволюционном смысле структура, достаточное развитие имеет только у человека), роль которых долгое время представлялась загадочной, поскольку ни за какие *конкретные* функции (зрение, речь, движение и т.д.) они не отвечают. Лишь в последние десятилетия было показано, что их роль та же, что у *дирижера оркестра*.

*«*Префронтальная кора играет центральную роль в формировании целей и задач, затем в разработке планов действий, требующихся для достижения этих целей. Она выбирает когнитивные умения, требующиеся для воплощения планов, координирует эти умения и применяет их в правильном порядке. Наконец, лобная часть коры головного мозга ответственна за оценивание

[8] «Люди склонны обучаться путем приобретения ситуационно-специфических умственных шаблонов или аттракторов» — цитируется по [1]

наших действий как успеха или неудачи относительно наших намерений.» (цитируется по [1])

*Как именно* реализуются все эти функции «дирижера» — *постановка* задачи, *оценка* результатов, и принятие *решений* — вопрос открытый и требующий отдельной *самостоятельной модели*. Пока мы можем делать только некие предположения.

Здесь мы не будем вдаваться в анализ постановки и формулировки конкретных задач, иерархию задач и т.д.; но заметим, что на каком-то этапе процесс решения *любой* задачи представляет собой процесс ***оперирования образами.*** Попытаемся представить, ***как*** это может происходить.

*Каким образом* в предлагаемой схеме происходит акт *мышления*, соответствующий *решению* какой-либо *поставленной задачи*? *Решением* любой задачи тоже должен быть некий образ, отвечающий заданным параметрам[9]. Он может быть *сконструирован* или *выбран* из уже имеющихся. Для поиска этого образа необходимо (см. например, [12]): *прогнозирование искомого*, его ключевых характеристик; необходимы критерии для оценки будущего решения. При «обучении с учителем» такая проблема не возникает, оценивать решение — функция учителя (сравнение с данным эталоном). При «обучении без учителя» образ «идеального решения», эталон, должен быть сконструирован самостоятельно, исходя из накопленного опыта, с применением некоторых фундаментальных критериев.

Ими могут быть:

- *критерий простоты* — можно предположить, что более короткая «цепочка», более простая карта предпочтительнее более длинной, более сложной;
- *критерий красоты* — можно предположить, что предпочтительна та цепочка/ карта, которая наименее противоречит существующему комплексу цепочек/ карт («личности нейрокомпьютера»), т.е. минимально искажает иные существующие цепочки/ карты.

По-видимому, оба критерия соответствуют минимуму энергозатрат на сохранение данной цепочки/карты, т.е. принципу «экономии зажженных нейронов». Этот принцип представляется актуальным как для человека, так и для нейрокомпьютера.

Однако в процессе решения задачи сами критерии могут подвергаться ревизии[10].

---

[9] Возможно, что решением может быть и *алгоритм, т.е. цепочка определенным способом связанных образов;* это, в сущности, частный случай *образа* в более широком понимании этого слова

[10] Пример: пишем статью. Образ «идеальной статьи» говорит, что статья должна быть не короче 0.4, но и не длиннее 1.0 печатного листа. Также в ней должна быть хотя бы одна мысль, кажущаяся умной. Текст должен быть иерархичен, т.е. структурирован и

В картине двух связанных подсистем, «ландшафт» которых в первом приближении мы будем считать одинаковым, мышление выглядит следующим образом. Искомое «идеальное» решение формируется в ЛП (и передается в ПП), но, во-первых, *не четко*, а во-вторых, *путь* к нему *не известен*.

Тогда в ПП происходит следующее: шарик прыгает случайно *из одного канала в другой* (как правило, достаточно близкий), «включая» тем самым образы, не связанные между собой непосредственно (дальние связи). Эти образы проверяются на предмет близости к идеальному образу (как? — см. ниже); рано или поздно «включается» образ, уже существующий в виде канала в данном ландшафте и достаточно близкий к «идеальному».

В ЛП: шарик катится по существующим каналам, связанным ассоциативно (ближние связи), по направлению, условно говоря, к «идеальному» образу**.** При этом естественно задействовать привычные, «глубокие», часто используемые каналы (т.е. *алгоритмы* или *аттракторы*).

Пример: чтобы описать развитие какого-то процесса во времени надо написать уравнение типа Ланжевена, включив в правую часть члены типа «источника», «взаимодействия» и «стока», решить его известными способами и сравнить полученное решение с известными характеристиками изучаемого процесса; если результаты не удовлетворительные, изменить правую часть уравнения, снова его решить и т.д.

Таким образом, рано или поздно шарик докатится до существующего канала, *достаточно близкого* к «идеальному» образу, и задача будет решена.

В описанном процессе результат деятельности обеих подсистем один и тот же (точнее, результаты могут быть *различны*, но в равной степени *удовлетворительные*), и действуют они вроде бы независимо друг от друга. На самом деле, это не так, и требуется участие обеих подсистем. Прежде всего, необходимо различать ***опытных*** (ландшафт которого достаточно разнообразен, хотя каналы не глубоки) и ***хорошо обученных*** (с развитой системой глубоких каналов-*аттракторов*, которые соответствуют неким *стереотипам*, *алгоритмам*) ***субъектов***.

Возможны варианты.

1). Образ– «решение» быстрее находится в ПП (для ***опытного*** субъекта как правило так). Тогда этот образ передается в ЛП, где, во-первых, *проверяется правильность* решения (т.е. проводится *детальное сравнение* найденного образа с эталоном), и, во-вторых, решается уже *задача Дирихле*, т.е. поиск пути между известными начальными и конечными точками.

2). Поиск решения в ЛП, представляющий собой, по сути дела, задачу Коши (поиск пути с известными начальными данными), сталкивается с

---

подчинен основной мысли. Обилие умных мыслей приводит к хаотичности текста, что нежелательно. Впрочем, получается то, что получается.

трудностями или неопределенностями (на пути шарика возникает разветвление канала, т.е. локальное состояние *безразличного* или *неустойчивого равновесия*) — тогда выбирается путь, «предлагаемый», т.е. опробованный, в ПП, даже если он неверен. Неудача или конфликт возвращает шарик в точку неопределенности; возможно, что поиск приходится начинать с самого начала.

3). Образ-«решение» быстрее находится в ЛП (для ***хорошо обученного*** субъекта); тогда роль ПП действительно нулевая.

Все эти варианты в равной мере правдоподобны; наиболее же реалистично смешение их, т.е. постоянное сравнение результатов поиска в двух подсистемах и постоянная корректировка выбора пути (ПП и ЛП «советуются друг с другом»). Но реалистичен и вариант решения задачи, при крайней необходимости, силами только одной (любой) подсистемы; однако, согласно изложенному, этот путь дольше и результат может быть иным. Вопрос о том, какая из систем – ПП или ЛП – играет главную роль, как мы видим, носит скорее схоластический характер. Важно отметить лишь то, что в ЛП путь шарика всегда можно *проследить (и проговорить, т.е. рассказать*), используя канонические образы, взятые из арсенала обучающего множества.

Что есть в данной схеме логика, а что интуиция?

Что есть **интуиция** — интуитивно понятно. Это *неаргументированное* решение, «прямое усмотрение истины», правильность которого, согласно Канту, проверяется только «*внутренним удовлетворением*» (что может имитировать это чувство — критерий «красоты»?).

В предложенной схеме: процессы, происходящие в ПП, носят достаточно случайный характер благодаря шуму. Здесь непрерывно возникают и разрушаются побочные связи, перебрасывающие сигнал («шарик») из одной цепочки в соседние, но *ассоциативно связанные* (возможно, не напрямую). На языке ландшафта: шарик прыгает по ландшафту, не следуя системе каналов, при этом ему проще перепрыгнуть в тот канал, который отделен от данного сравнительно низким барьером. Таким образом, в ПП происходит свободное блуждание по близким понятиям; это блуждание, в силу своей *случайности*, не *запоминается*, поскольку *не повторяется*, и, таким образом, *не становится новыми каналами*, частью ландшафта (в противном случае само понятие ландшафта стало бы слишком «расплывчатым» и потеряло смысл).

Важно отметить, что если система не обучена, или обучена плохо, это блуждание просто активизирует случайным образом некоторые нейроны, не приводя к чему-либо содержательному. Чем лучше система обучена, тем больше цепей, связанных с различными образами в ней запомнено, так что блуждание происходит именно по ассоциативным цепям и может приводить к правильному «ответу».

Описанный процесс естественно считать имитацией *интуитивного* мышления. Как видно, это интуитивное мышление тем более эффективно, чем больше «жизненный опыт», т.е. чем *шире* обучающее множество и чем более богат и *разнообразен* ландшафт.

Как можно *запомнить* путь, приведший к «озарению», и так ли он случаен? Здесь мы касаемся (забегая вперед) крайне важной и актуальной (в частности, для медицинской диагностики) проблемы *перевода интуитивного знания в логическое* (см. например, [4]). Эта проблема далека от окончательного решения. В рамках нашей схемы решение проблемы можно представить как попытку ЛП *вспомнить* тот путь, который прописан в ПП «штрихом», и *восстановить* или *аппроксимировать* промежутки кусками *канонических каналов*[11].

Что есть **логика** и «**логическое мышление**»? Мы брали это слово в кавычки, поскольку в «научном» и «обыденном» понимании этого слова присутствуют явные расхождения.

В научном мире четкого определения логики нет, но приняты следующие аксиомы.

**«Логическое мышление есть вынесение суждения на основе конкретного *однозначного* алгоритма. Оно не индивидуально, поскольку при заданных начальных условиях все, кто его используют, должны прийти к одинаковым выводам. Логику составляет набор принятых аксиом. На каждый вопрос, сформулированный в рамках принятой аксиоматики, должен быть однозначный ответ. Знания в форме логических алгоритмов можно передавать и сохранять в веках.»**

В таком понимании логическое мышление — лишь часть, причем весьма малая того, что подразумевается в обычной (не научной) жизни, когда говорят «давайте мыслить логически». В жизни *обычного* человека задачи, имеющие однозначное решение, встречаются только на уроках математики и физики (научная деятельность представляет некоторое исключение, хотя и не абсолютное).

Передача знаний не требует, вообще говоря, строгого кода помимо языка. Не говоря о гуманитарных знаниях, которые тоже передаются *вербально*, не будучи при этом строго алгоритмизованными, ***любая*** информация может быть передана вербально, на языке слов и на уровне образов. Даже уравнения Ньютона и Шредингера можно «рассказать».

С другой стороны, алгоритм, в обыденном понимании этого слова, совсем не обязательно приводит к *единому* результату; простейший пример —

[11] Поясним: опытный врач может поставить диагноз только по походке и состоянию волос пациента, при этом *обосновать* этот диагноз либо не удается вообще, либо удается в процессе подготовки лекции для студентов. Для этого надо *вспомнить*, о чем (возможно, совсем постороннем) думал, когда осматривал похожих пациентов.

кулинарный рецепт. Это безусловно алгоритм, т.е. записанная кодом (словами) последовательность действий, но результат может быть непредсказуем, хотя вероятность достижения желаемого результата при следовании алгоритму заведомо выше, чем при случайном поиске решения (произвольном смешивании ингредиентов кулинарного рецепта).

Еще одно замечание: передача знаний, *необходимая и неизбежная* в жизни человека, передача от родителей к детям, ***никогда*** не происходит «логически» в научном понимании этого слова, а тем не менее таки происходит. Подчеркнем, что этот эффект не связан с тем, что процесс самообучения ребенка неизбежно включает элемент случайности, а с тем, что сам родитель не способен сформулировать «жизненный алгоритм» на языке *однозначного кода*.

**Мы предлагаем**: ввести термин «***образно-последовательное***» мышление, или «***образная логика***» для обозначения того процесса, который происходит в ЛП.

Он естественно будет включать в себя, как предельный случай, логическое мышление в старом смысле слова. Однако он включает и менее строгие *алгоритмы*, т.е. цепочки *связанных друг с другом ассоциативно* образов, приводящие к решению какой-либо задачи.

Важно подчеркнуть, что эти алгоритмы *индивидуальны.* Какой именно путь был использован в ЛП для решения задачи, какие каналы задействованы, зависит сугубо от *личности* человека (нейрокомпьютера) и ее *состояния* в данный конкретный момент. Р. Пенроуз [5] отмечал, что мысль может быть ***не вербальна****:* когда человек думает, в его голове прокручиваются *собственные*, возможно, визуальные, возможно, геометрические или иные образы; для публикации же своих размышлений необходимо *приложить усилия* для вербализации хода своей мысли.

Кроме того, эти алгоритмы *не единственны:* к одному и тому же «образу–решению» можно подойти по *разным каналам*, соответствующим *разным образным рядам*. На этом основано *искусство перевода,* способность к *коммуникации*, т.е. передачи информации в данном обществе: независимо от того, *как именно* человек (компьютер) пришел к какому-нибудь решению, он должен построить цепочку каналов из *общепринятых* в данном обществе образов и *вербализовать* ее. Более того, *выбор слов* для вербализации также должен быть ориентирован на определенное сообщество и связан именно с тем образным рядом, который наиболее адекватен, адаптирован к ситуации[12].

[12] С.А. Чернавский (мой дед – О.Ч.) говорил: «не говорите, что человек дурак, это грубо, скажите, что у него трансцендентно-авторитарное мышление, смысл тот же, но приличия соблюдены»

Если цепочка образов сложилась таким образом, что каждый последующий образ *естественно* (т.е. *привычно* для какого-либо, достаточно широкого, круга людей) связан с предыдущим, она может быть вербализована и опубликована. Именно такой характер имеет «логическое мышление» в более широком, обыденном понимании.

Как можно интерпретировать **творчество**?

Четкого определения этого понятия также не существует. Принято считать, что творчество есть создание чего-то *принципиально нового* (однако, необходимо конкретизировать, для *кого именно* нового). В терминах синергетики [3,4]: «творчество есть *генерация новой ценной информации*». В рамках предлагаемой схемы естественно определить творчество как *строительство* или *сотворение ландшафта*. Все эти определения не противоречат друг другу; последнее, на наш взгляд, наиболее конструктивно, поскольку определяет одновременно и *цель*, и *механизм* творчества.

В нашей схеме в строительстве ландшафта участвуют обе подсистемы, но ведущую роль — *прокладывание «каналов»* — играет ПП, а *углубление, запоминание и шлифовка* их — задача ЛП.

С этой точки зрения и первичное обучение, и обучение вообще являются творчеством. Известное утверждение «любой ребенок талантлив» становится очевидным, если его перефразировать: «любой ребенок креативен». Он строит свой, сначала маленький ландшафт, который часто может отличаться от общепринятого, но тем не менее иметь смысл (3-х-летний ребенок говорит «я *уронился*», что забавно и интересно, талантливо); потом эти отличия забываются и «усредняются» внешними воздействиями.

Но и на более поздних этапах любое обучение чему-либо новому — процесс творческий, если новое знание естественно *встраивается* в уже существующий ландшафт, становится его *продолжением*, архитектурной деталью, а не запоминается механически, как побочный *пристроенный* модуль.

Следующий этап творчества — это разработка собственных новых (для себя) *алгоритмов*, моделей, которые связывают различные образы в *ситуации*. На языке ландшафта это можно представить себе как *асфальтирование важных дорог* и *наведение мостов*, т.е. установление причинно-следственных связей. При этом сигнал о потребности в них идет из ЛП, но в самом акте «создания мостов» необходимо участие ПП.

Искусство возникает как продолжение процесса создания внутреннего ландшафта во внешний мир — это в каком-то смысле ***обучение наоборот***, *потребность в гармонии* между внутренним миром, содержащим в себе прекрасные воображаемые картины (мелодии, статуи и т.п.), и внешним миром, где это все необходимо создать. Такая потребность возникает далеко не у всех людей. Точнее, потребность достраивать собственный ландшафт,

отражающий внешний мир, образами/ситуациями, не существующими во внешнем мире, но похожими на реальные (имеющими право на существование), называется ***воображением*** и присутствует практически у каждого, но талантом и способностью воплощать, реализовывать свои внутренние образы наделены единицы.

Повторим еще раз, что принципиальным моментом как для строительства, так и для достраивания ландшафта является шум в ПП, случайное самовозбуждение нейронов. Ранее мы предполагали (для простоты изложения), что уровень шума в ПП постоянен. Однако, вероятно более естественным является иное предположение: шум в каких-то ситуациях слабый, в каких-то других ситуациях усиливается, и тогда *потребность и способность* к *достраиванию* внутреннего ландшафта увеличивается. Такие состояния естественно отождествить с «приступами вдохновения»[13]. Отметим, что из изложенного следует, что «приступы вдохновения» у нейрокомпьютера можно вызывать искусственно, повышая уровень шума в ПП.

Обсудим отдельно **научное творчество**.

Начнем с утверждения: 90% научной работы творчеством не является. Вообще научная работа по определению связана с *алгоритмами*, т.е. закономерностями, повторяющимися ситуациями. Большинство научных исследований связано или со сбором информации и систематизированием ее на основе *известных алгоритмов*, или с применением опять же *известных алгоритмов* к проблемам, слегка отличающимся от уже известных. Все перечисленное относится к сфере деятельности ЛП.

Творчество начинается с разработки *новых научных алгоритмов,* т.е. при *поиске* новых закономерностей, связывающих (*как мосты*) известные явления, а *поиск*, как не раз отмечалось, невозможен без шума, случайного элемента.

Возможно 2 варианта.

1). Ранее полагалось, что шум в ЛП пренебрежимо мал. Возможно, однако, и другое предположение: шум в ЛП (по-прежнему, гораздо меньший, чем в ПП) включается при *осознанном* желании решить определенную задачу и может проявляться внешне в виде нервозности (как сказано выше, [4]). Тогда новый алгоритм формируется в ЛП, поскольку шум позволяет в какой-то момент «перескочить» из одной канонической цепочки в другую, тем самым *проложив новый канал, создав новую связь,* которую человечество еще не увидело, и которую не учат в школе. Ньютона должно было ударить яблоко, чтобы он увидел связь силы, массы и ускорения, называемую вторым законом Ньютона. На нашем языке это означает, что связи, возникшие в его ЛП, встряхнулись и выявилась одна конкретная новая.

---

[13] Фраза из одного научного доклада «Творческие люди пребывают в состоянии вдохновения различной степени тяжести»

2). ***Общественная потребность*** в научном творчестве возникает, когда две устоявшиеся и многократно проверенные научные дисциплины входят в конфликт. Тогда между ними возникает *большой барьер*, для преодоления которого нужны *значительные усилия* и «незамыленный», более широкий взгляд (т.е. объединение всех доступных обучающих множеств)[14]. Здесь более вероятно, что решение находится все же в ПП, поскольку объединение далеких областей знания, «дальние связи» - функция именно ПП.

Важно подчеркнуть, что необходимым элементом в любом варианте является *шум*, который порождает ***перемешивающий слой*** из известных, «канонических» цепочек, позволяющий проводить поиск нового решения и исключить априорную детерминированность. Но поскольку все решения в ПП находятся не «шагами» (т.е. кусочками привычных связей, канонических каналов), а «прыжками», ЛП должно-таки «встряхнуться», чтобы, зная, *куда* надо попасть, проложить туда необходимые каналы.

[14] Пример разрешения научного противоречия: проблема возрастания энтропии в квантовой механике. С одной стороны, энтропия реальных систем должна возрастать; с другой --- теорема фон Неймана, согласно которой она тождественно равна 0. Решение было найдено авторами данной работы [13]; оно потребовало осознания необходимости и роли *ансамбля* в тех случаях, когда система *не устойчива*. Именно опыт изучения *неустойчивых* систем позволил разрешить противоречие путем введения более точного определения энтропии, связанного с очередностью усреднения, что в рамках собственно квантовой механики не нужно. Отметим, что новое утверждение в этом случае, как это и требуется для развития науки, не *отрицает*, а *обобщает* старое.

## 5. Чем и насколько «робот»[15] отличается от человека?

В рамках предложенной схемы можно ответить на вопрос, в чем заключаются отличия искусственного интеллекта от живого.

Широко распространено мнение, что «у робота не может быть интуиции».

В рамках предложенной схемы двух связанных подсистем это абсолютно не так. Мы видели, что процессы, происходящие в ПП, неизбежно имитируют именно *интуитивное* мышление, иначе «робот» ничему не научится. Другое дело, что без ЛП он не сможет внятно объяснить, чему он научился.

Следующая парадигма: у робота не может быть чувства юмора. Вообще говоря, это тоже не так. Робот не умеет смеяться (хотя можно заложить аудио-имитацию смеха в определенных ситуациях), но может шутить. Было высказано [14] предположение, что смех есть реакция человека на ситуацию (высказывание) *неожиданную*, но тем не менее имеющую смысл на уровне вторичных или побочных ассоциаций. Например: «дед Макар сидел на крылечке и крутил козью ножку. Коза недовольно блеяла». В этом случае ожидаемый логически и уже возникающий в мозгу образ папиросы-самокрутки необходимо *стряхнуть* и заменить образом козы, стоящей рядом с крылечком и недовольно брыкающейся. Для этого возникает *волна вибрации*, перебрасывающая «сигнал» из одной цепочки в другую (связанную с первой словом *коза*), которая и называется смехом. Отметим, что такое чувство юмора требует достаточно широкого обучающего множества, т.е. по крайней мере знания того, что есть папироса и что есть коза («для того, чтобы робот понимал шутки про зарплату, он должен ее получать»).

Заметим, что в рамках обсуждаемой гипотезы становится понятно, почему шутка, повторенная дважды, смеха не вызывает: ситуация становится ожидаемой. С другой стороны, *пережитая* смешная ситуация при воспоминании по-прежнему вызывает смех, поскольку он есть часть воспоминания.

Можно ли этот эффект имитировать? Разумеется, можно включить в программу нейрокомпьютера функцию «смех» (как в презентациях Power Point), но, по-видимому, ее можно привязать только к деятельности ЛП. Будучи снабжен такой функцией, ПП, который непрерывно перебирает цепочки, относящиеся к разным задачам и образам, будет практически постоянно хохотать.

Можно предположить, что число «органов чувств» робота меньше, чем у человека. Это тоже не так: видео- и аудио- ряды элементарно

[15] Здесь и далее слово «робот» употребляется как синоним «нейрокомпьютер», исключительно для краткости.

оцифровываются, несколько сложнее с тактильными и одорологическими воздействиями (ощущения и запах). Но в принципе и их можно закодировать и превратить в определенные сигналы.

Далее: может ли робот испытывать боль? Канонически считалось, что это свойство присуще только живым организмам и к точным наукам — физике, химии и т.д. — отношения иметь не может[16]. Так ли это? Определенное внешнее тактильное воздействие (удар о землю) можно имитировать как *конфликт определенного рода = боль,* и система будет искать пути, исключающие такое воздействие.

Могут ли у робота быть эмоции? Этот вопрос более сложный. Имитировать гнев, ярость, недовольство можно так же, как и смех (включить опцию «грязные ругательства» и скоррелировать ее с конфликтом), как и радость (опция «радостные крики», когда найден правильный и «красивый» ответ).

Может ли при этом нейрокомпьютер *чувствовать*? Вопрос более тонкий. На первый взгляд ответ отрицательный, но может быть в будущем понятие «чувство» будет формализовано и, соответственно, появится возможность его имитировать.

Одно из основных отличий, которое (пока?) не может быть преодолено даже на уровне имитации: роботу не может быть «хорошо» или «плохо». Возможно потому, что мы сами (пока?) не понимаем, что это такое. Человек не может объяснить, т.е. сформулировать на языке определенных алгоритмов или даже образов, почему ему в данный момент хорошо, просто хорошо и все. Если будет найдена *формула любви*, возможно, и нейрокомпьютеры приобретут совсем человеческий облик.

Так или иначе, те, вроде бы принципиальные, отличия, которые традиционно считались непреодолимыми, могут быть устранены (по крайней мере на уровне имитации или аналогии).

Что же действительно непреодолимо?

*Робот думает, когда получает внешний сигнал и/или задачу, человек же думает всегда, даже когда спит.*

Действительно, человека, в отличие от робота, невозможно «отключить от сети». Он непрерывно обдумывает какие-то задачи и проблемы, поставленные самому себе, даже если его никто к этому не принуждает. Даже во сне, когда внешние раздражители не воспринимаются, человеческий организм живет, поэтому посылает какие-то сигналы в мозг. В это время ПП освобождается от конкретных задач и получает возможность «разобраться в себе», т.е. свободно «гулять» по существующим ассоциативным цепочкам и порождать новые. Если такие цепочки оказываются почему-либо достаточно

[16] Классический пример *нонсенса*: «Тело массы М падает с высоты Н. Телу больно, тело этого не любит».

сильными, они проникают в ЛП в виде *снов.* Почему человеку приснился тот или иной сон — связано с конкретной личностью, историей, эмоционально окрашенной памятью и т.д. Они всегда и несомненно имеют смысл и несут большую информацию о деятельности и состоянии ПП, надо ее только понять.

Можно ли имитировать такое состояние нейрокомпьютера? В принципе любой компьютер (даже не нейро-) обязательно снабжен системой авто-диагностики и авто-усовершенствования, которая работает и в «ждущем режиме». Можно ли считать это состояние сном? Вряд ли: такая функция запрограммирована, т.е. связана с четким известным алгоритмом, а это совсем не то, что происходит во сне.

Можно, не отключая компьютер от сети и убрав внешние «целевые» воздействия, наложить, например, *слабый случайный шум* (ниже порога восприятия ЛП). Тогда в ПП робота, так же, как и у человека, будут генерироваться случайные процессы блуждания по связям, т.е. по его памяти, и порождаться новые, «невыученные» связи. Этот процесс может улучшить интуицию робота, что может иметь интересные практические приложения. Однако подчеркнем, что сон человека — функция его *личности,* а не только формальной памяти (отражение обучающего множества). Даже если абстрагироваться от вопроса «что есть *личность* робота» и считать, что она существует и представляет собой просто его «ландшафт», т.е. массив «каналов», обученных цепочек, то предлагаемый способ имитации сна все же не несет информации о состоянии «организма» робота, хотя и способствует дообучению.

Таким образом, можно заключить, что мы не можем имитировать только то, чего не понимаем.

### 6. Заключительные замечания

Таким образом, мы показали, что предложенная схема позволяет реализовать те эффекты, которые обсуждались во введении, а именно :

- интуиция и логика;
- автопилот;
- «вещие сны»;
- творчество, в частности, научное.

Мы видели, что для этого *принципиальным моментом* является наличие именно *двух связанных* подобных подсистем, отличающихся тем, что в одной из них уровень *шума* достаточно высок для обеспечения эффекта «перемешивающего слоя», в другой же, напротив, шум минимален. Именно при соблюдении этого условия система способна и к самообучению, и к сохранению как приобретенных, так и полученных извне знаний.

Вообще роль *шума*, случайного элемента, заслуживает специального обсуждения. «Шум» стал всерьез рассматриваться как самостоятельное

явление только в середине-конце XX века; до этого (да зачастую и сейчас) полагали, что шум — лишь досадная и неизбежная помеха, требующая значительных усилий для ее минимизации. И только в последние десятилетия стало понятно, что шум может играть созидательную роль, порождая «перемешивающий слой» в сложных, развивающихся системах и тем самым обеспечивая собственно развитие.

В данной работе мы рассмотрели только простейшую ситуацию: постоянный по амплитуде и частоте шум в ПП и минимальный, пренебрежимый шум в ЛП. На самом деле спектр возможностей на этом пути очень широк: варьируя амплитуду и частоту шума в обеих подсистемах, можно моделировать интересные эффекты, вызываемые резонансными явлениями, синхронные и циклические процессы и т.д. Эти задачи мы оставляем на будущее.

Ранее (см. например [3,4]) предполагалось, что процесс обучения и запоминания может быть реализован в рамках *одной* системы, процессора Больцмана, для чего необходимо:

1) наличие шума в начальной стадии (для возможности случайного поиска);

2) постепенное снижение уровня шума, так, что в конечной стадии принимается единственное решение, дальнейших «колебаний» нет, и это решение запоминается.

Но возникает ряд вопросов, главные из которых: что, как и почему включает этот шум, а затем снижает его уровень? Высказывалось предположение [4], что аналогом этого эффекта у человека является *нервозность,* возникающая перед принятием *важного* решения, но почему она уменьшается? Более того, при решении следующей аналогичной задачи начальный шум вновь создаст непредсказуемый «перемешивающий слой», так что все предыдущие усилия потеряют смысл («жизнь ничему не научила»).

Если задача не решена с первой попытки (т.е. в 90% случаев), как сохраняется информация о ней и других неудачных попытках? В одной системе — никак, т.е. поиск решения каждый раз идет «вслепую». В системе, состоящей из двух частей, «шумной» и «тихой», эти вопросы решены. Кроме того, в природе полушария все-таки два, значит это зачем-то нужно.

С другой стороны, человеческий организм, как известно, содержит чрезвычайно много степеней защиты и *резервного функционирования.* Это значит, что *при необходимости* (например, инсульт) любое из полушарий может худо-бедно работать за двоих, т.е. включать/выключать шум по мере поступления задач и, таким образом, справляться и с поиском, и с запоминанием информации. Но получается это с трудом. Таким образом, концепция двух подсистем *не отрицает* того, что предполагалось ранее, но *включает* в себя как вариант, реализуемый при определенных условиях,

ч*астный случай* — таким образом, общее правило научного творчества, о котором говорилось выше, в рамках нашей концепции соблюдается. В общем случае участие двух систем и «разделение труда» между ними представляется более естественным.

Косвенным аргументом в пользу предложенной выше схемы может служить также тот факт, что природа, создавшая два почти одинаковых полушария, создала также мужчину и женщину. Зачем? Почему мозг не един, равно как почему не создать единого Человека, способного к самовоспроизводству?

В середине 1960-х В.Геодакян [15] высказал гипотезу, которая представляется нам очень естественной. Распределения мужчин и женщин по любому содержательному признаку имеют разный вид. Если у женщин это, как правило, гауссово распределение вокруг некоторого среднего, то распределение мужчин гораздо шире, причем «хвосты» более тяжелые (типа Парето, см., например, [16]), т.е. «экстремалов»-мужчин гораздо больше. В зависимости от меняющихся внешних условий выживает та часть мужчин, которые имеют отклонения от «нормы» (среднего) в нужную сторону. Важно, что прогнозировать изменения внешних условий на уровне биосферы практически невозможно, поэтому множество мужчин, приспосабливающихся к этим условиям должно быть максимально широким, иметь оба «хвоста». Распределение женщин «подстраивается» под мужские мутации, т.е., оставаясь гауссовым, смещается в сторону модифицированного распределения мужчин, сохраняя при этом нужные мутации.

Таким образом, мужчины и женщины на глобальном уровне имеют разные функции: задача мужчины — *поиск*, т.е. *обучение* новым условиям, задача женщины — *сохранение ценной информации*. В нашей же схеме ситуация аналогична: роль мужчин выполняет ПП, роль женщин – ЛП.

Широко распространено *суеверие*, что ПП (интуиция) лучше развито у женщин, а ЛП («логика») — у мужчин, что, казалось бы, противоречит только что сказанному. На наш взгляд, эта проблема принадлежит скорее к разряду социальных. Мы полагаем, что интуиция развивается независимо от полового признака, но зависит от рода деятельности (обучающего множества), которое действительно зависит от пола в силу сложившихся в обществе традиций.

Иными словами, ассоциативный ряд у женщин и мужчин настолько разный, что те образы, которыми мыслят и оперируют женщины, не понятны мужчинам, и наоборот. Однако если прибегнуть к услугам *переводчика*, т.е. человека, в ассоциативном арсенале (ландшафте) которого имеются и те, и другие образные ряды, взаимопонимание вполне возможно, и соотношение интуитивного и логического в подходе к решению проблем может оказаться чисто индивидуальным.

В любом случае, социальный аспект мышления заслуживает отдельного исследования. Прежде всего, это относится к акту *вербализации* полученных решений, необходимому для применения данного решения в *данном обществе*.

Ранее мы неоднократно употребляли термины «рассказать», «проговорить» и т.д., не останавливаясь специально на механизмах, им соответствующих, а эта проблема отнюдь не тривиальна.

Насколько искусственный интеллект («робот») может быть приближен к живому, обсуждалось в п.5. Было показано, что интуиция не только возможна, но и имманентно неизбежна в предложенной схеме. Непреодолимое (пока?) отличие заключается в том, что ассоциативные связи, возникающие в процессе обучения (ландшафт) человека имеют эмоциональную окраску: какие-то картины обучающего множества ему приятны, какие-то вызывают тоску.

Что есть тоска? Мы сами не знаем.

Еще один аспект этой проблемы, обсуждавшийся в п.5 — сон. Это очень важный момент жизнедеятельности человека может, в принципе, быть привнесен в нейрокомпьютер. Это состояние, когда отключены внешние воздействия, позволяет ПП свободно бродить по ассоциациям и генерировать новые ассоциативные связи, которые «штрихом» прописываются в ЛП как *возможные.* Отличие же заключается в том, что человек во время сна воспринимает слабые сигналы *собственного организма,* или, иными словами, собственной *личности*, поэтому информация, сгенерированная в ПП и проникающая в ЛП в виде *снов*, имеет большую ценность и смысл, чем то, что может породить «спящий» робот.

Подводя итог, можно сказать, что имитация живого интеллекта искусственным невозможна только в тех случаях, когда мы *не можем четко сформулировать, что именно имитируем.*

Самое существенное отличие заключается в том, что *робот думает, только когда ему ставят какую-либо задачу извне; человек думает всегда (даже если ему за это не платят).* Возможно ли преодолеть это отличие и необходимо ли это? Данная проблема выходит за рамки предлагаемого исследования.

Разумеется, в рамках обсуждаемой схемы остаются нерешенными многие вопросы. Первый и главный из них — проблема принятия решения, т.е. выбор одного из двух предлагаемых ЛП и ПП (если они различны). Оно не должно быть случайным (иначе объект будет страдать дебилизмом), но в тоже время не может быть полностью детерминированным (объект «тупой робот»). Процесс принятия решения, или, более широко, *организация* мозговой деятельности (см. [1]), требует специальной самостоятельной модели. Пока мы можем делать только некоторые качественные предположения.

Еще одна проблема, примыкающая к первой — формализация фактора *внимания.* Интуитивно понятно, что это такое; понятен и тот факт, что роль его в процессе мышления (и живого, и искусственного) должна быть велика, но интерпретировать этот фактор *только* на используемом нами языке *ландшафта* не удается, этот фактор также относится к сфере *организации.*

Кроме того, мы не конкретизировали вид связей $\psi^R_{ij}$ и $\psi^L_{ij}$ («ландшафта»), считая их в первом приближении одинаковыми. Вообще говоря, это не так; более того, в ЛП должен быть реализован и кодирующий процесс, переводящий конкретные *образы* в абстрактные *понятия;* процессор Гроссберга [7] может быть и не единственным способом кодировки.

Вообще проблемы кодировки требуют специального исследования, которое сейчас находится в стадии активной разработки. Не конкретизировался также и вид «перекрестных» связей $\chi^R_{ij}(t)$ и $\chi^L_{ij}(t)$, качественные эффекты «перекрестного взаимодействия» двух подсистем обсуждались лишь вербально. Более строгий подход также требует специального исследования возможного вида перекрестных связей (по-видимому, сопряженного с исследованием вариаций амплитуды и частоты шумов в ПП и ЛП, см. выше).

Имеются и другие аспекты, требующие внимания и специального анализа; работа над ними продолжается.

В заключение следует сказать, что роль и функции двух полушарий человеческого мозга исследовались и обсуждались давно, и когнитология, наука о сознании, развивается бурно. Накоплено много интересных фактов и концепций.

Так, известно, что в детстве критично поражение ПП, в старости – наоборот, ЛП. Мудрость и рационализм традиционно относят к функциям ЛП, в то время, как интуицию и творческое начало — к ПП. Обсуждались также указания на то, что в ПП информация обрабатывается *параллельно*, а в ЛП – *последовательно* («задача Коши», [9]). В [1,2], на основе многолетнего клинического анализа функций дисфункций мозговой активности человека, было убедительно показано, что ПП отвечает за обработку *новой* (для данного индивидуума) информации, в то время как знакомая, «*рутинная*», информация обрабатывается в ПП (концепция «новизна – рутина»), что чрезвычайно близко к нашей концепции «учиться – учить»[17].

Мы намеренно не обсуждали ранее эти соображения, поскольку шли «с другой стороны», а именно, от универсального с точки зрения *физики* явления: системы с «шумом» обеспечивают генерацию новой (для данной системы) информации (т.е. дают возможность *учиться, записывать* новую информацию), в то время как для *сохранения* информации и *оперирования* ею

[17] справедливости ради отметим, что эта концепция стала известна авторам *после* того, как настоящая работа была фактически завершена

«шум» не нужен и даже вреден (*ценность* информации резко снижается)[3]. Но нетрудно видеть, что наша концепция двух связанных подсистем квазислучайной («шумной»)» (ПП) и квазидетерминированной (ЛП), по крайней мере, ***не противоречит*** тому, что известно о деятельности полушарий мозга человека; в некоторых случаях (например, интерпретация *механизмов* интуиции и «логики») альтернативные концепции отсутствуют (по крайней мере, нам не известны).

Подчеркнем, что мы не претендуем на адекватное *моделирование человеческого мышления* (хотя, конечно, держим это в уме). Мы лишь предлагаем *схему компьютерного устройства*, в рамках которой можно реализовать те эффекты, которые нас интересовали с самого начала, а именно *интуицию* робота, *автопилот, сознание/подсознание, творчество* и, возможно, ряд других, аналоги которых мы наблюдаем в реальной жизни.

Авторы приносят искреннюю благодарность С. А. Шумскому за плодотворные дискуссии, которые, собственно, и привлекли наш интерес к данной проблеме.

Работа выполнена при поддержке гранта РГНФ № 07-03-00658а

**Список литературы.**

1. Голдберг Э. Управляющий разум, *Москва, Смысл,* 2003.

2. Голдберг Э. Парадокс мудрости. *Москва, УРСС,* 2005.

3. Чернавский Д.С. «Синергетика и информация». *Москва., УРСС*, 2004

4. Чернавский Д.С., Карп В.П., Родштат И.В., Никитин А.П., Чернавская Н.М. «Распознавание. Аутодиагностика. Мышление». (Синергетика и наука о человеке) – *Москва: Радиотехника*, 272 с., 2004.

5. Пенроуз Р. «Новый ум короля», *Москва, УРСС*, 687с., 2005; «Тени разума*», Москва-Ижевск, Институт компьютерных исследований*, 688с., 2005.

6. Hopfield J.J. «Neural networks and physical systems with emergent collective computational abilities». *Proc.Natl.Acad.Sci.*, 79, pp. 2554-2558, 1982

7.Grossberg S. (ed.) «The adaptive brain I: Cognition, learning, reinforcement and rhythm»; «The adaptive brain II: Vision, speech, language and motor control». *North-Holland, Amsterdam*, 1987.

8. Kohonen T. «Self-organization and associative memory». *Springer-Verlag, New York*, 1984

9. Ежов А.А., Шумский С.А. «Нейрокомпьютинг и его применения в экономике и бизнесе», *курс лекций прочитанный в МИФИ 2006.*

10. Лурия А.Р. Нейрофизиология памяти. *Москва, Педагогика*, 1976.

11. Чернавский Д.С., Никитин А.П., Чернавская О.Д., Щепетов Д.С. «О свойствах динамических окон в кубическом отображении». *Краткие сообщения по Физике,* №2, с. 1, 2007.

12. Карп В.П., Никитин А.П. «Интуитивное и логическое в задачах распознавания и принятия решений» *Эпистемология и философия науки,* №3, 2005.

13. Чернавский Д.С., Чернавская О.Д. «О проблеме необратимости в квантовой механике», *Краткие Сообщения по Физике,* т.5, с.1, 1999.

14. Суслов И.М. «Компьютерная модель чувства юмора» *Биофизика*, т. 37, в. 2, стр. 318-324, 1992.

15. Геодакян В.А. «Эволюционная логика дифференциации полов» *Природа*, №1, с. 70-80, 1983; *ДАН*, т. 269, №12, с.477-482, 1983.

16. Чернавский Д.С., Никитин А.П., Чернавская О.Д. «О возникновении распределения Парето в нелинейных динамических системах» *Биофизика*, т. 53, №2, с. 351-358, 2008