

# К вопросу о возможности цифрового воссоздания личности (digital immortality)

Ольга Чернавская

Lebedev Physical Institute, Moscow, Russia  
olgadmitcher@gmail.com

**Абстракт.** Идея воспроизведения личности с помощью цифровых (нейросетевых) технологий («цифровое бессмертие») привлекает к себе большое внимание с начала 2000-х по настоящее время. Это действительно возможно? Мы утверждаем, что к «Бессмертию для Окружающих» можно подойти достаточно близко. Для этого следует рассмотреть понятие личности на «инженерном» уровне: необходимо выделить основные черты личности, т. е. составить «конструкционный портрет», и попытаться воспроизвести его в рамках выбранной когнитивной архитектуры. Мы используем оригинальную Естественно-Конструктивистскую Когнитивную Архитектуру (ЕККА), которая изначально содержит интерпретацию логического и интуитивного мышления, подсознания и т. д. Такая система должна обучаться на определенной выборке знаний, присущих конкретному человеку (книги, тексты, изображения, фильмы, д.), при этом упор делается на личностные *вокабуляры* (вербальный, эмоциональный, поведенческий). Показано, что ЕККА включает в себя широкий набор свободных параметров, что позволяет воспроизводить широкий спектр личностных черт, от стиля мышления до темперамента. Мы полагаем, что эта программа могла бы дать возможность создать «виртуальную личность» («цифровой двойник»), способную создать иллюзию общения с конкретным человеком.

**Keywords:** Личность, Конструкционный Портрет, Лексикон\Вокабуляр, Мимика, Парадокс, Подсознание, Снимок Эмоции, «Цифровой Двойник»

## 1 Введение

Можно ли воссоздать чью-либо личность при помощи современных цифровых (нейросетевых) технологий? Этот вопрос бурно обсуждается с начала 2000х и по сей день на самых разных уровнях — религиозном, юридическом, даже морально-этическом (см. например [1]–[4] и ссылки там же). В последнем случае высказывается мнение, что такая «опция» может быть доступна только для очень богатых людей, что есть аморально. Но если бы удалось воссоздать даже самую

«завалиющую» (в моральном отношении) душонку, это был бы истинный переворот не только для науки, но и для морали, религии и т.д. Однако возможно ли это в действительности?

Прежде всего, следует различать «бессмертие для себя» и «бессмертие для окружающих людей». Легко показать, что первая задача не корректна, не интересна и неразрешима.

Действительно, личное бессмертие, т.е. ощущение «я есть вечно» невозможно *делегировать*. Здесь речь может идти только о физическом бессмертии (возможно, путем протезирования всех органов включая нейроны), но это — вопрос технологий очень далекого будущего (if any), для когнитивного моделирования здесь простора нет. Что же касается построения модели, идентичной своему Я, то это затея ненужная и даже вредная. Действительно, представьте, что такая модель создана, т.е. сделан «робот», *ощущения* которого идентичны Вашим, в том смысле, что он чувствует *то же*, что Вы. Но Вы-то не перестаете себя ощущать! Таким образом, этот робот становится Вашим двойником и, следовательно, соперником, вплоть до Вашей физической смерти, когда человек остается наедине с Богом. А далее — «каждому воздастся по вере его»: если человек верит в Бога, то бессмертие ему не нужно, оно у него уже есть. Если же нет, то дальнейшая судьба его двойника, который вроде бы мыслил и чувствовал также, как он, его по большому счету не волнует.

Вторая задача — бессмертие в сознании людей — выглядит гораздо реалистичнее. Более того, в отношении некоторых определенных личностей (Понтий Пилат, Леонардо да Винчи, Альберт Эйнштейн, ...) она практически решена: их помнят. Бессмертие в смысле людской памяти существует с незапамятных времен в форме сказаний, легенд, затем фотографий, видео и т.д. Может ли развитие цифровых (нейросетевых) технологий дать существенно новый импульс в достижении этой цели для любой личности?

Для того, чтобы ответить на этот вопрос, необходимо определить «что есть личность». Этот вопрос многократно обсуждался в литературе в психологическом, философском, этическом и т.д. планах. Это привело (опять же) к множеству определений, среди которых наиболее общим и одновременно конструктивным кажется то, что представлено в Oxford Learner's Dictionary: «Личность — это различные аспекты характера человека, которые в совокупности отличают его от других людей» [5]. Это определение подразумевает практический («инженерный») подход — *как* это можно сделать. Для этого следует выделить наиболее характерные черты данной личности, т.е., создать *конструкционный портрет* личности (по аналогии со словесным портретом).

При этом необходимо иметь в виду, *какая именно цель* преследуется в процессе «реинкарнации». Например, если речь идет о гениальном изобретателе, можно было бы воссоздать его личность в надежде на какое-либо новое изобретение. Тогда «портрет личности» должен включать малейшие подробности профессиональной деятельности. Другой пример: бот гениального преступника мог бы использоваться в роли «доктора Лектора», т.е. консультанта по поимке реальных преступников. Однако наиболее естественной выглядит цель создания вир-

туальной личности, которая могла бы дать *иллюзию общения с реальной*. Конечно, это имеет смысл главным образом для родных и близких, но в некоторых случаях адресная аудитория может быть существенно шире.

С точки зрения когнитивного моделирования следующий шаг — попытаться воспроизвести (или хотя бы имитировать) выделенные черты личности в какой-либо когнитивной архитектуре. Мы рассматриваем эти проблемы в рамках модели Естественно-Конструктивистская Когнитивная Архитектура (ЕККА) [6]–[9]<sup>1</sup>. Эта модель построена на принципах Динамической Теории Информации [10], согласно которой любая развивающаяся система должна состоять из двух разных (но, разумеется, связанных) подсистем, одна из которых ответственна за генерацию информации, обучение и творчество, а другая — за хранение и обработку полученной информации. Это позволяет естественно воспроизводить и интерпретировать понятия логического и интуитивного мышления, подсознания и т.д. Кроме того, эта модель имеет широкий спектр свободных параметров, что дает возможность *настроить* систему под заданный набор особенностей когнитивного процесса (включая его эмоциональную составляющую).

Эта программа может привести к созданию «цифрового двойника», дающего (по крайней мере) иллюзию общения с данной личностью, если так «настроенную» когнитивную архитектуру снабдить соответственно выбранными (из поведенческой части конструкционного портрета) визуальными эффектами.

Работа организована следующим образом. В разделе 2 рассматривается «инженерный» подход к определению понятия личности и обсуждается «конструкционный портрет» личности. В разделе 3 рассмотрены основные особенности ЕККА с учетом цели моделирования конкретной личности. В разделе 4 «конструкционный портрет» личности обсуждается в терминах ЕККА. Выводы и обсуждение представлены в разделе 5. В Аппендиксе представлен проект «конструкционного портрета личности» *Д.С. Чернавского*.

## 2 Что есть личность

Следует сразу оговориться, что мы постоянно будем сталкиваться с терминологическими трудностями, поскольку все термины употребляются в самых разных областях человеческой деятельности, причем в несколько *разных* смыслах (слова «затерты»). Так, понятие «личность» широко обсуждалось в философии, психологии, этике и т.д. В результате появилось множество разных весьма расплывчатых определений. Наиболее емкое и конструктивное дано в Oxford Learner Dictionary: “the various aspects of a person's character that combine to make them different from other people” («различные аспекты характера человека, которые в совокупности делают его отличным от других людей»). Это — почти «инженерный подход»: для того, чтобы определить «что есть личность», нужно *выделить (сформулировать)* ее наиболее характерные черты и особенности. Иными словами,

---

<sup>1</sup> В англоязычных публикациях вместо аббревиатуры ЕККА используется NCCA

речь идет о том, чтобы составить «конструкционный портрет» личности (по аналогии со «словесным портретом»). Эти «черты личности» можно условно разделить на 3 крупных блока:

- Когнитивные особенности;
- Эмоции и эмоциональные реакции;
- Поведенческие особенности.

Разумеется, эти характеристики нельзя механически изолировать, они влияют друг на друга, и тем не менее. Обсудим подробнее эти блоки вместе с возможностью *восстановления* (источники) и *воспроизведения* этой информации при помощи нейросетевых технологий.

## 2.1 Когнитивные особенности

Когнитивные особенности включают в себя

- Знания;
- Умения, профессионализм, интересы;
- Опыт;
- Лексикон\Вокабуляр;
- Цели;
- Воображение, креативность, талант;
- Характер (тип) и стиль мышления.

Этот перечень, разумеется, не полон, и при необходимости его можно продолжить. Рассмотрим подробнее эти пункты.

**Память.** Первые три пункта на самом деле могут и должны быть объединены понятием «память». Действительно, все перечисленное относится к «жизненным накоплениям», т.е. к архиву.

Знания конкретного человека могут быть велики, но не бесконечны и безусловно меньше не только всех знаний, накопленных человечеством, но даже и энциклопедии. Он знает только то, что *узнал и запомнил*.

*Примечание. В свое время была модной идея закачать в робота как можно больше информации, тем самым сделав его умнее любого конкретного человека. Однако обладание информацией не делает человека (равно как и робота) умным, надо еще знать, что, почему и как с ней делать*

Можно выделить несколько принципиально разных *типов* памяти: семантическая, эпизодическая (эксплицитная), и имплицитная.

**Семантическая память.** Это — знания, полученные извне из книг, фильмов, интернета (в обобщенном смысле). Какие именно книги, фильмы, музыку и т.д. знал и любил данный человек, — хорошо знает он сам и его ближайшее окружение, так что эту информацию восстановить достаточно просто. Именно этот пул информации может и должен быть использован как «*обучающее множество*».

Отметим, что, если речь идет о «реинкарнации» *ученого*, в это множество следует включить и его *собственные работы* (статьи, книги, выступления), а также работы, на которые автор ссылался. Этот прием может быть очень эффективным.

*Эпизодическая память.* Это — впечатления и знания, полученные человеком на собственном опыте в течение всей жизни. Источником этой информации могут быть только его собственные рассказы и рассказы о нем (те же «легенды»). В случае яркой личности такой информации может оказаться неожиданно много, хотя она может оказаться не слишком достоверной («ложная память»). В любом случае легенды — бесценный источник «бессмертной» памяти *о личности*.

*Имплицитная память.* То, что человек знает, но сам об этом не догадывается. На наш взгляд, здесь более уместен термин *подсознание*, хотя и он имеет множество толкований. Согласно словарю Вебстера, подсознание есть «совокупность психических процессов, о которых индивидуум не осознает; не-формулируемая и неконтролируемая умственная деятельность» [11]. Согласно этому определению, человек сам не *знает*, что именно хранится в его подсознании; информация об этом может проявляться в неожиданных неясных (*implicit*) ассоциациях, оговорках, опечатках и т.д. К проявлениям подсознания относятся, на наш взгляд, и внезапные догадки, озарения (*insight*).

**Лексикон\Вокабуляр.** Как известно, во времена Остапа Бендера среднестатистический человек знал примерно 30000 слов. Возможно, современный человек знает и больше, но это не так важно. Важно, что человек *реально использует* гораздо меньше слов, т.е. имеет свой личный *вокабуляр* — набор наиболее часто употребляемых слов и фраз. Разумеется, этот набор тоже хранится в памяти, но мы выделили его в отдельный пункт, поскольку именно он в наибольшей степени «иллюстрирует» личность. Источник информации о вокабуляре, как и об эпизодической памяти — рассказы и письма человека, фрагменты его выступлений, память окружающих.

**Воображение, креативность, талант.** Эти свойства личности могут не иметь конкретных внешних проявлений и оцениваются окружающими *postfactum*, по результатам ее деятельности. Креативность, т.е. создание чего-то *нового*, связана с генерацией новой *субъективной* (см. ниже) информации, но основана она на том, чем личность уже обладает — то, что хранится в памяти и, особенно, в подсознании. *Талант* или *дар* определяют как «выдающиеся *врождённые* способности человека, проявляемые в определённой сфере деятельности» [12] — т.е. что это такое конкретно, никто не знает. На наш взгляд *имитировать* (как и *моделировать*) талант невозможно, но креативность, т.е. способность к творчеству, анализировать и моделировать можно. Мы вернемся к этой проблеме в Разделе 4.

**Глобальная Цель.** Речь идет не о текущих намерениях, а о неких глобальных, «жизне-образующих» (по аналогии с термином «градообразующий») целях. Таковыми могут быть: карьера; престиж; благополучие; любопытство; творчество; список можно продолжать. Любое из этих понятий можно трактовать достаточно широко. Так, под благополучием можно трактовать как финансовый достаток, так и спокойное существование без обязательств и проблем («чтобы все оставили

в покое»). Любопытство может выражаться в желании непрерывно путешествовать, или в желании что-либо *понять* (характерно для научных работников).

Цели могут объединяться и смешиваться («карьера-престиж-достаток»), но как правило человеком руководит какая-то главная из них. Однако, следует отметить, что эта цель безусловно характеризует личность, но может не иметь отношения (и даже противоречить) к конкретным реакциям и поступкам. Нельзя также судить о глобальной цели по известным достижениям и общественному положению человека – иногда (и довольно часто) они не пересекаются. Истинную глобальную цель осознает (if any) только сам человек, а воспроизвести ее можно лишь со слов близких ему людей.

**Характер (тип) мышления.** Различают два типа мышления:

- логическое;
- образное (?!) \ интуитивное \ ассоциативное.

В чистом виде они встречаются крайне редко, и такие случаи в определенном смысле являются патологией. Тот и другой тип имеют преимущества и недостатки, но спорить о том, какой лучше, бессмысленно. Человек использует обычно и тот, и другой тип, но какой-то из них может превалировать.

*Логическое мышление.* Этот тип мышления характеризуется последовательным рассуждением, основанным на имеющейся (*запомненной*) информации и *обще-принятых* причинно-следственных связях, а также на *прямых* (очевидных) ассоциациях. При этом ход мысли прозрачен и понятен для окружающих. Как правило, задача решается (*if any*) достаточно эффективно (в *разумное* время).

*Образное\интуитивное\ассоциативное мышление.* Сам термин «образное мышление» вызывает активный протест: в любом случае мысль порождает образы и наоборот, так что любое мышление в определенной степени образное. Термин «интуиция» определяется в Википедии как «способность человека понимать и проникать в смысл событий и ситуаций посредством едино-моментного бессознательного вывода — озарения» [13]. Однако это определение кажется слишком «грубым», сводя интуицию лишь к сиюминутному озарению.

Определение из словаря Britannica представляется более адекватным [14]: “*something that is known or understood without proof or evidence*” = «то, что известно или понятно без доказательства или свидетельства». По сути, здесь подразумевается, что существуют разные *уровни* интуиции. Часто человек *может* в принципе перевести свой внутренний путь к догадке на общепринятый язык, т.е. найти понятные другим аргументы, но ему лень — такой тип мышление более естественно называть «*внутренней логикой*». Но возможно, что догадка *не* ожидаема и удивительна для самого человека, результат не внутренних рассуждений, а именно *озарения* – это есть проявление «*глубинной*» интуиции с отсылкой к подсознанию.

Мы предпочитаем термин «интуитивно-ассоциативное» мышление, имея в виду, что этот ход мысли основан на собственных *индивидуальных (implicit)* ас-

социациях, непонятных для окружающих, или же на смутных (не сформулированных) впечатлениях («что-то *кажется*»), что и есть чисто интуитивная составляющая. Кроме того, для данного типа мышления характерны перескоки на посторонние, никак не связанные (на первый взгляд) проблемы и образы («мысли скачут»). При таком типе мышления задача решается либо сравнительно долго, либо очень быстро – в случае озарения.

*Стиль мышления.* Независимо от того, что превалирует, логика или интуиция, ход рассуждений может неуклонно вести к заданной цели, не отвлекаясь на параллельные проблемы, или, наоборот, уходить по параллельным ассоциациям к сходным проблемам, лишь затем возвращаясь к основной. Эти стили мышления можно условно назвать «вертикальным» и «горизонтальным», соответственно. Стиль мышления особенно ярко проявляется в разговоре с собеседником.

Особенности характера мышления, как правило, достаточно ярко проявляются, и их легко зафиксировать. Можно ли их воспроизвести в какой-нибудь искусственной когнитивной системе? В разделе 4 будет показано, что именно модель ЕККА такую возможность может предоставить.

## 2.2 Эмоции и Эмоциональные Реакции

Эмоции (от лат. *emoveo* = возбуждать, волновать) — наиболее интригующая и загадочная сфера человеческой психики. Определение из Oxford Language Dictionary [15] “сильное чувство, проистекающее из обстоятельств, настроения или отношений с другими” мало проясняет ситуацию. Интуитивно понятно, о чем идет речь, но как это определить с точки зрения «инженерного» подхода? Как можно *вербализовать* (т.е. объяснить словами) и зафиксировать эмоциональные особенности личности?

Этот термин, по словам А. Ребера [16] «...оказался чрезвычайно *резистентным* к попыткам его определить, возможно, никакой другой термин в психологии не сочетает такую неопределенность с частотой его использования». Наиболее конструктивное (на наш взгляд) определение приведено в словаре Britannica [17]: “сложное переживание сознания, телесное ощущение и поведение, отражающее *субъективное оценочное отношение* к ситуации”. Согласно этому определению, у каждой эмоции имеются три составляющие (*‘ипостаси’*):

- психологическая\ когнитивная (*осознание* эмоции);
- физиологическая (изменения пульса, дыхания, «*мурашки*» по коже);
- поведенческая (мимика, интонации, смех, и т.д.).

Забавно, что физиологическая составляющая «субъективной оценки ситуации» вполне объективно может быть измерена и зафиксирована, однако обычно этого никто не делает. Говоря о «выявлении характерных эмоциональных особенностей» личности, можно надеяться только на третью ипостась — поведенческий «портрет» (*snapshot*) типичных для данной личности эмоций. Именно эта информация может быть запомнена окружающими людьми и зафиксирована на фото и видео. Но как ее можно воспроизвести? Современные *роботы* в широком смысле

слова (искусственные когнитивные системы) не могут *испытывать* эмоции, могут лишь *распознавать* и *имитировать* их.

Для *репродукции* эмоционального аспекта личности с помощью современных технологий следует разделять две цели\задачи: *моделировать* и *имитировать* эмоции. В первом случае надо фактически научить робота их *испытывать*, т.е. понять механизм, который может вызывать возникновение той или иной эмоции, встроить его в когнитивную архитектуру, и только потом дать этой эмоции 'название' и связать с ее визуальными (поведенческая «ипостась») проявлениями. Для второй задачи (имитация эмоций) достаточно просто *распознать\назвать* определенную эмоцию и присвоить ей определенную поведенческую реакцию. Для этого необходим *только* «эмоциональный вокабуляр».

**Вокабуляр Эмоций.** Несмотря на, казалось бы, огромное количество *эмоций*, которые испытывает конкретный человек в течение своей жизни, разнообразие его *эмоциональных реакций* не так велико. Можно выделить несколько (10? 20?) типичных ситуаций, вызывающих типичные эмоции, сопровождающиеся типичной для данного человека эмоциональной реакцией (включая физиологические и поведенческие компоненты). Иными словами, каждая типичная эмоция имеет свой «портрет\слепок» (*snapshot*), т. е. специфическое выражение лица, конкретную интонацию, характерный жест и т. д. Конечно, такой «слепок эмоции» индивидуален, хотя многие характерные эмоциональные проявления (слезы, смех, улыбка, поджатые губы и др.) типичны для *конкретной эмоции* и оказываются сходными у многих людей (на этом основана профессия верификатора).

«Конструкционный портрет личности» должен включать в себя «вокабуляр эмоций», т. е. перечень характерных для данного человека эмоциональных реакций вместе с индивидуальным «поведенческим слепком» (*snapshot*) для каждой эмоции. Такой список не должен быть слишком длинным.

Однако для того, чтобы *понять* как именно устанавливается соответствие между эмоцией и ее слепком, необходимо решать первую задачу, т.е. моделирование эмоций. Она представляется нам более интересной (хотя возможно и менее реальной), и далее мы сосредоточимся именно на ней при составлении эмоционального аспекта «конструкционного портрета личности».

**Моделирование когнитивной составляющей эмоций.** В эволюционном плане эмоции представляют собой *более древний* механизм выживания (как инструмент реагирования на изменения окружающей среды), чем когнитивный анализ (см. например [18][19]). Они определяются составом нейро-медиаторов, который связан с продукцией подкорковых структур (*hippocampus, amygdalae, basal ganglia, etc.*), более древних, чем *неокортекс* (т.е. 'old Brain'). Поэтому физиологические и поведенческие реакции (по крайней мере, первичные: «бей-беги-замри») проявляются *раньше*, чем их *осознание*. Во многом благодаря этим факторам, эмоции остаются наиболее загадочной и трудной для когнитивного моделирования



сферой, несмотря на многочисленные попытки [20]–[23]. Однако ниже мы покажем, что в модели ЕККА эмоции встраиваются естественно и неизбежно.

Несмотря на все многообразие причин и ситуаций, в которых человек испытывает эмоции, можно выделить некую общую формулу, которая подходит почти ко всем распространенным «эмогенным» (“emogenic” ~ возбуждающий) ситуациям: «конфликт плюс его разрешение». Рассмотрим это подробнее.

*Прагматические (Рациональные) Эмоции.* Достаточно легко смоделировать прагматические (или *рациональные*) эмоции, т.е. те, которые связаны с достижением определенной цели. Здесь удаление от цели («конфликт») вызывает отрицательные эмоции, а приближение к цели («разрешение») — положительные. При этом характер («качество») и интенсивность эмоций зависят от конкретной цели: эмоции при завоевании олимпийской медали, конечно, отличаются от таковых после удачного похода в магазин, хотя в обоих случаях они положительные (радость, удовлетворение).

Однако следует подчеркнуть, что при решении регулярных (бытовых) задач человек испытывает лишь слабые привычные эмоции, которые почти не имеют поведенческой составляющей. Действительно яркую и выразительную реакцию вызывает нечто *неожиданное, необъяснимое или парадоксальное*. Именно эти реакции, будучи индивидуальными, создают «*вокабуляр эмоций*» личности.

Отметим, что такой эмоциональный всплеск («*слепок*» данной эмоции, типичный для данного человека) может накладываться на любой другой регулярный процесс, оставаясь при этом узнаваемым.

*Эстетические Эмоции.* Гораздо более интересным представляется вопрос об *эстетических* эмоциях – впечатление от созерцания произведения искусства, музыки, явлений Природы и т.д. Эти эмоции *не имеют рациональных* причин, *сугубо индивидуальны и искренни* («мурашки по коже» или есть, или нет, симулировать их невозможно). Тем не менее, в нашей работе [9] обсуждалась гипотеза, позволяющая эти загадки разрешить; мы вернемся к этой проблеме в разделе 4.

*Моделирование Чувства Юмора.* Отдельного обсуждения заслуживает чувство юмора, т.е. «психологическая особенность человека, заключающаяся не только в примечании *противоречий* в окружающем мире, но и в оценке их с комической точки зрения» [24]. Несмотря на то, что, согласно определению, это «психологическая особенность», причем ключевым словом здесь является «противоречие» (*парадокс*), его традиционно относят к эмоциональной сфере, поскольку его проявление стандартно – улыбки, смех (эмоциональные реакции).

Считается, что чувство юмора нельзя объяснить и развить (это либо *дано*, либо нет), но на наш взгляд, это не совсем так. Чувство юмора<sup>2</sup> срабатывает в хорошо знакомых (и прогнозируемых) ситуациях, когда вдруг возникает некое новое обстоятельство, которое заставляет трактовать ее как совершенно иную, но

<sup>2</sup> Следует оговориться: мы не рассматриваем ситуации, когда «бьют грязной тряпкой по лицу», это кажется смешным отнюдь не всем

тоже хорошо знакомую. Таким образом, возникает *парадокс распознавания* (*'unexpected but still familiar'*), резкий переход от удивления к пониманию, который выражается в виде смеха («встряска»). Мы вернемся к этому это в разделе 4.

*Ирония.* Отдельной строкой стоит выделить это свойство, родственное чувству юмора. Его традиционно связывают с эмоциональной сферой: хотя оно скорее относится к особенностям мышления, его проявление обычно связано с улыбкой, насмешливым выражением лица и т.д., что облегчает его визуализацию.

Согласно определению, ирония есть «...противопоставление того, что на первый взгляд кажется правильным, и того, что на самом деле имеет место или ожидается» [25]. Это определение подразумевает, что, в отличие от чувства юмора, связанного с неожиданной сменой взгляда на ситуацию, ирония означает постоянную (и ожидаемую) двойную трактовку («видеть двойное дно»), т.е. свойство замечать не только основные характеристики какого-либо явления, но и сопутствующие неочевидные и вроде бы не важные.

*Эмпатия.* Способность *сопереживать*, т.е. *со-чувствовать* другому человеку по сути связана с когнитивной функцией *воображения* (себя на месте другого), однако безусловно относится к эмоциональной сфере. Именно она отличает нормального социализированного человека от робота и социопата. Однако здесь есть забавный парадокс: известно, что социопат способен блестяще имитировать эмпатию, просто путем *зеркального отражения* эмоциональных проявлений собеседника – в итоге он производит впечатления тонкого психолога и становится (если это ему нужно) «лучшим другом». Воспроизвести такое поведение при помощи цифровых технологий достаточно просто, но зачем? Конкретный (нормальный) человек проявляет эмпатию (когда ее действительно чувствует) по-своему, в свойственной именно ему манере (определенная мимика, интонации,...), и эти проявления обычно достаточно просто детектировать и воспроизвести.

### 2.3 Поведенческие Особенности

Речь идет о тех гранях личности, которые формируются при рождении или в раннем детстве и далее срабатывают на бессознательном уровне («автопилот»).

*Примечание. Не следует путать «бессознательное» с «подсознательным». В первом случае человек действует автоматически, не задумываясь, но при необходимости может четко сформулировать причины поступков («расшифровать автопилот»). Во втором случае действия человека не понятны ему самому, их причины скрыты.*

К таковым относятся:

- темперамент;
- стиль поведения\общения;
- характерные привычки;
- присказки, шутки, восклицания, междометия, т.д.

Эти грани личности наиболее заметны, и именно они остаются в памяти окружающих. Насколько и как их можно воспроизвести в искусственной когнитивной системе, мы обсудим в разделе 4.

**Темперамент.** Считается, что темперамент – врожденные особенности поведения личности, которые проявляются вне зависимости от желания человека и могут лишь корректироваться за счет умственных усилий (то, что называется «выдержка»). Со времен Гиппократов, выделяют 4 основных типов темперамента, которые соответствуют различным, в терминологии И. П. Павлова, *типам* нервной системы:

- сангвиник (активный, веселый) – сильный уравновешенный тип;
- холерик (импульсивный, горячий) – сильный, неуравновешенный тип;
- меланхолик (грустный, подавленный) – слабый, неуравновешенный;
- флегматик (спокойный, медлительный) – слабый, уравновешенный.

В чистом виде эти темпераменты встречаются редко, обычно присутствует некий смешанный тип. Согласно Гиппократу, темперамент зависит от преобладания в организме определенных *жизненных соков* («желтой и черной желчи, лимфы, и крови»). На современном языке это можно интерпретировать как определенный состав нейромедиаторов (стимуляторов и ингибиторов) в организме. Очевидно, темперамент влияет и на эмоции, и на когнитивные процессы, и должен учитываться при когнитивном моделировании личности.

**Стиль поведения.** Ключевую роль в их формировании играет *воспитание*, т.е. влияние семьи и ближайшего окружения. Поэтому информация о семье может значительно помочь в воссоздании эпизодической памяти личности.

Существует несколько широко известных и распространенных типов или *шаблонов* поведения/общения – агрессивный, грубый, интеллигентный, доброжелательный, высокомерный, ироничный, и т.д. Обычно каждый индивидуум использует один выбранный шаблон поведения, но, в зависимости от ситуации, может перейти на другой, тоже шаблонный, режим.

Каждый из этих шаблонов/режимов поведения имеет характерные внешние признаки (мимика, жесты, интонации), поэтому его достаточно легко распознать и воспроизвести. Важно отметить, что каждому режиму поведения соответствует свой *под-лексикон*, т.е. слова и фразы, используемые *только* в такой моде.

**Привычки.** Имеются в виду необъяснимые предпочтения (цвета, вкуса, запаха), неосознанные жесты, машинальные действия. Они не связаны с эмоциональными проявлениями, а срабатывают машинально, независимо от мыслительной деятельности. Набор таких привычек («вокабуляр привычек») обычно невелик и достаточно легко воспроизводим, но он может иметь большой иллюстративный эффект.

**Приказки, Восклицания, Междометия ....** Характерные слова и фразы (включая ругательства), которые человек использует машинально, часто как междометия. Забавно, но сам человек может не осознавать и не помнить этот «мини-вокабуляр», но окружающие запоминают именно это в первую очередь, так что эта информация легкодоступна и иллюстративна.

## 2.4 Сборный конструкционный портрет личности

Таким образом, собрав вместе *воспоминания многих людей* о данной личности, можно выделить из них следующую информацию:

- обучающее множество (семантическая память);
- эпизоды из биографии (эпизодическая память);
- тип и стиль мышления;
- особенности чувства юмора и иронии;
- лексикон\вокабуляр;
- вокабуляр эмоций (по рассказам и фотографиям);
- темперамент;
- поведенческий репертуар (набор используемых шаблонов);
- репертуар\вокабуляр привычек;
- репертуар «восклицаний» (в широком смысле);

можно говорить о *сборном конструкционном портрете* личности («личностный конструктор») и возможности его воссоздания средствами когнитивного моделирования.

## 3 Основные Особенности ЕККА в свете Моделирования Конкретной Личности

Мы разрабатываем оригинальную модели ЕККА, предложенную и развитую в работах [6]–[9]. Одно из базовых оснований этой модели — Динамическая Теория Информации (ДТИ) [10]. Напомним кратко основные выводы этой теории.

### 3.1 Основные выводы ДТИ

ДТИ возникла и развивалась Пригожиным [26], Хакеном [27] и Чернавским [10] как одно из направлений общей теории *сложных развивающихся* (в частности, живых) *систем* (см., например, [28] и ссылки там же). В отличие от теории информации Шеннона [29], разработанной для эффективной *передачи* информации, ДТИ фокусируется на проблеме *возникновения* информации.

**Роль Шума в сложных развивающихся системах.** Шум, т.е. *случайная активация* каких-либо (*ану*) элементов\функций в сложных развивающихся системах,

является **универсальным инструментом приспособления** к **внезапным** и/или **непредсказуемым** изменениям внешней среды. В процессе эволюции удачные случайные решения сохраняются, а неудачные – отмирают (вместе с системами).

В мозге человека шум обеспечивают подкорковые структуры, относящиеся к «old Brain» (*hippocampus, basal ganglia, amygdala* и др.), которые производят *нейромедиаторы* [19]. Все нейромедиаторы делятся на два типа: **стимуляторы** (ускоряющие *все* процессы) и **ингибиторы** (подавляющие активность). Продукция стимуляторов возрастает при изменении среды, обеспечивая тем самым и возможность *случайной* активации (*шум*).

**КАК рождается информация.** Согласно определению Кастлера [30], информация есть «**запомненный выбор** одного варианта из нескольких (многих) равноправных». В зависимости от того, **как** делается этот выбор, различают два пути возникновения информации:

- **генерация** новой информации — свободный *случайный* выбор, акт *создания* информации (творчества);
- **рецепция** информации — вынужденный выбор, навязанный извне. В этом процессе новая информация не возникает, а уже существующая может быть утрачена. Иначе говоря, в этом случае происходит *не выбор, а отбор*.

**Принцип разделения функций.** Функции генерации и рецепции информации *комплементарны* (дополнительны, несовместимы), поэтому любая развивающаяся система должна *состоять из двух* (разумеется, связанных) подсистем. Одна из них должна быть ответственна за генерацию новой информации (обучение, творчество), другая – за сохранение той информации, которая рецептируется как из первой (генерирующей) подсистемы, так и извне. В дальнейшем эта информация используется для обработки вновь полученной информации (распознавание, классификация, и т.д.).

**Субъективная vs Объективная Информация.** В зависимости от того, **кем** делается этот выбор, различаются:

- **Объективная Информация** – выбор, сделанный Природой. Она возникает независимо от системы человека; ее можно (и должно) только *воспринимать (рецептировать)*. В человеческом мозге она воспринимается через сенсорные сигналы, т.е. на уровне “Brain”<sup>3</sup>.
- **Субъективная («Условная»<sup>4</sup>) Информация** – выбор, сделанный в результате взаимодействия (“договоренности” = “convention”) в коллективе.

<sup>3</sup> В английском языке есть четкое разделение терминов “Brain” (мозг) и “Mind” (Разум? Ум? Сознательное?). В русском языке слово «мозг» обычно относят ко всем церебральным процессам, а устоявшегося аналога для термина Mind не существует (пока?). Здесь и далее мы будем использовать английскую терминологию.

<sup>4</sup> Часто при переводе термина «условная информация» путают термины “conventional” и “conditional” — это недопустимо, термин “conditional information” не имеет смысла.

Подчеркнем, что выбор, сделанный в системе «коллектив нейронов», представляет собой ее *индивидуальную* (собственную) *субъективную* информацию, продукт ее когнитивной деятельности, т.е. относится к сфере “Mind”.

**Гомеостаз.** Каждая развивающаяся система (биологическая, экологическая, социальная) должна иметь механизм *гомеостаза*. Согласно общему определению, гомеостаз есть «процесс само-регуляции, который обеспечивает способность системы сохранять внутреннюю стабильность, приспосабливаясь к изменяющимся внешним условиям» [31]. В живых системах под «внутренней стабильностью» следует понимать состояние внутреннего *динамического равновесия* системы.

Конкретный механизм само-регуляции зависит от вида системы. В мозге человека (см. [19] и ссылки там же) само-регуляция основана на «игре» нейромедиаторов, вырабатываемых в подкорковых структурах (тех же, что обеспечивают и шум). Продукция стимуляторов возрастает при изменении среды, а ингибиторы вырабатываются (однако, не сразу) для компенсаторного эффекта. В результате все регулярные процессы в развивающейся системе носят характер *колебаний* (в основном, *затухающих*) вокруг состояния динамического равновесия.

Из того факта, что одни и те же структуры старого мозга контролируют\производят и эмоции, и шум [18][19], следует парадоксальный (на первый взгляд) вывод: именно эти два (казалось бы, анти-стабильных) фактора\процесса обеспечивают само-регуляцию и стабильность (гомеостаз) в мозге человека.

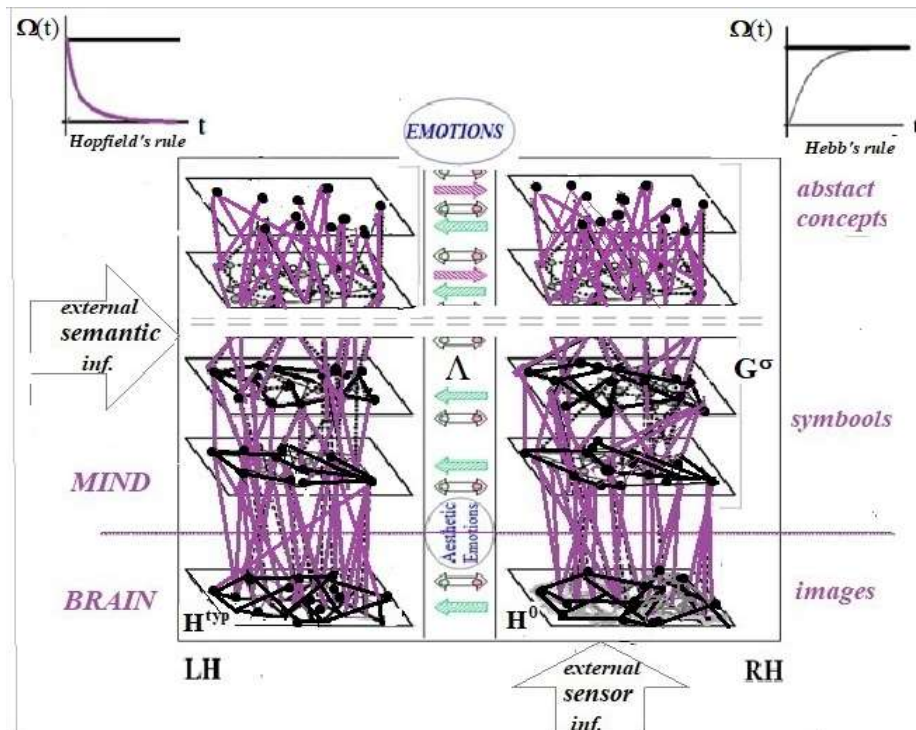
Поскольку внутреннее динамическое равновесие является жизненно важной целью для любой развивающейся системы, движение к нему, т. е. приближение к равновесному состоянию, вызывает положительные эмоции, и наоборот, удаление от него — отрицательные. Заметим, что формула «конфликт плюс его разрешение», предложенная для мотивации *рациональных* («прагматических») эмоций, как раз и представляет собой акт мини-гомеостаза.

**Глобальная Цель.** Согласно ДТИ, основной целью любой развивающейся системы является «*генерация, защита и распространение (expansion) собственной субъективной (условной) информации*» [10].

Отметим, что практически все возможные цели личности, перечисленные в разделе 2, так или иначе соответствуют данному выводу ДТИ.

### 3.2 Основные положения ЕККА

Когнитивная архитектура ЕККА представляет собой сложную иерархическую структуру, составленную из нейропроцессоров типа Хопфилда [32] и Гроссберга [33]. На Рис.1 представлена общая схема ЕККА.



**Fig. 1.** Схема ЕККА с иллюстрацией правил обучения межнейронных связей. Обозначения приведены в английском варианте: LH и RH означают ЛП и ПП, соответственно.

Согласно принципу разделения функций ДТИ, вся система разделена на две подсистемы, ПП (Правая Подсистема) и ЛП (Левая Подсистема). Подчеркнем, что обозначения продиктованы (инспирированы) биологической аналогией: в Природе подобное разделение функций реализовано в полушариях головного мозга: правое полушарие ответственно за генерацию информации (обучение, творчество), а левое полушарие – за хранение и обработку созданной и полученной информации; экспериментальные свидетельства в пользу этой гипотезы приведены в работах клинициста-нейрофизиолога Голдберга [34].

В рамках ЕККА такая специализация обеспечивается двумя факторами:

- наличием шума (случайного самовозбуждения нейронов) в ПП и отсутствие его в ЛП.
- разными («зеркальными») законами обучения межнейронных связей: согласно Хопфилду [32] в ЛП и Хеббу [35] в ПП.

**Принципы построения.** На нижнем ( $\sigma=0$ ) уровне иерархии представлены процессоры типа Хопфилда (*распределенная память*), где каждому объекту соответствует цепочка активных нейронов. Напомним, что распределенная память обеспечивает:

- механизм *восстановления*: активация нескольких нейронов цепочки вызывает активацию и всех остальных;
- *прямые ассоциации*: цепочки, имеющие общие нейроны, ассоциативно связаны, и активация одной из них способствует активации другой.

Принцип работы системы в следующем.

*Объективная* информация от реальных объектов (через различные рецепторы) поступает на пластину  $H^0$  в ПП, где для каждого предъявленного к обучению объекта выбирается (благодаря шуму) некая цепочка возбужденных нейронов, называемая *образом* данного объекта. В подсистеме ПП связи между нейронами цепочки обучаются по Хеббу [35], т.е. усиливаются («чернеют») по мере обучения. Хорошо изученные образы (те, в которых связи «почернели») передаются (по прямым меж-подсистемным связям) в подсистему ЛП для хранения. Пластина ЛП хранит только четкие («черные») образы, которые служат основой для распознавания новых объектов, прогноза и т.д.

В этом процессе нейроны, изначально одинаковые и равноправные, начинают играть *разные роли*. Те нейроны в ПП, связи между которыми почернели до состояния «типичного образа», становятся «*кор*»-нейронами, т.е. несут *типичные признаки* того объекта, образ которого составляют. Информация о них передается в ЛП и на следующую пластину следующего уровня, где становятся *основой символа* этого образа.

Другие нейроны, связи между которыми остаются относительно слабыми («серыми» в разной степени, что пока не важно), становятся «*хало*»-нейронами — они создают «*гало*» вокруг типичного образа, отражая редкие и нетипичные модификации объекта (т.е. *нетипичные признаки*). Информация об этих нейронах и их «серых» связях хранится *только на данном* уровне в ПП и не уходит ни в ЛП, ни куда-либо дальше.

Отметим, что многие нейроны оказываются *не* вовлеченными ни в какой когнитивный процесс (ни в какой образ\символ) — они называются «спящими» и могут быть активированы лишь при специальных условиях (об этом см. [9]).

*Принцип Вертикального Построения*. Достаточно четкие («черные») образы из ПП передаются не только на тот же уровень в ЛП, но и (через меж-пластинные связи) на *следующий уровень* — пластину типа Гроссберга [33], которая обеспечивает *конкурентное* взаимодействие, благодаря чему из всех нейронов, составляющих образ, «выживает» (остается активным) только *один*. Этот нейрон, называемый далее *символом* данного образа, перетягивает на себя все меж-пластинные связи с нейронами образа на хопфилдовской пластине (знаменитый принцип Кохонена [36] «победитель получает все, WTA»), и вся эта информация передается (реплицируется) в ЛП. Эта процедура описана подробно в [9]. После этого активация символа влечет за собой и активацию его образа (символ «разворачивается в картинку»).

Те нейроны, которые уже стали символами, не участвуют более в конкурентном взаимодействии (за право быть символом какого-либо другого образа), но могут, как и на пластине Хопфилда, участвовать в *кооперативном* взаимодей-



ствии. Символы в ПП сами могут объединяться в цепочку, создавая новый ‘образ-из-символов’ («обобщенный образ»), который ведет себя так же, как и на пластине образов  $H^0$ : связи постепенно чернеют, новые ‘черные’ образы передаются в ЛП и на следующий уровень ПП для выбора *их* символов. При этом часть информации снова «теряется», т.е. переходит в разряд *скрытой* (латентной) информации. Те образы-из-символов, которые не стали четкими (черными), не передаются ни в ЛП, ни на следующий уровень, оставаясь лишь на данном уровне ПП.

Подчеркнем, что нейрон-символ, помимо участия в создании обобщенного образа вместе с соседями («горизонтальное» взаимодействие), использует еще и «руки и ноги»: он активизирует свой образ на предыдущем уровне ( $\sigma-1$ ) и, будучи членом нового образа-из-символов, управляется символом этого образа на следующем уровне ( $\sigma+1$ ) — это т.н. «вертикальное взаимодействие».

Эта процедура повторяется на каждом иерархическом уровне  $\sigma$ , благодаря чему система *сама* развивается (*растет*) за счет самоорганизации в “коллективе нейронов”. В физике такой принцип организации называют *скейлинг*.

**Рост и Развитие.** Символы 1-го ( $\sigma=1$ ) уровня являются символами конкретного типичного образа (например, хорошо знакомая конкретная кошка). На более высоком уровне все конкретные кошки объединяются в символ понятия «кошка». Этот *символ понятия* в ПП сопоставляется с *внешней* (семантической) информацией в ЛП и становится *словом*. Так на этих уровнях развития система обретает *язык*.

С ростом номера уровня, новые понятия теряют связь с реальностью (т.е. не имеют какого-то «живого» образа), и превращаются в *абстрактные понятия*. Эти понятия обобщают множество других понятий и могут быть объяснены лишь при помощи многих слов. Например, определение понятия *совесть* требует целого философского трактата. В человеческом мозге за обработку абстрактной информации отвечают лобные доли (*frontal lobe*).

Таким образом, будучи инспирирована разбиением мозга на церебральные полусферы, система ЕККА, уже сама, повторяет его свойства, т.е. развивается от нижних (задних) *визуальных* областей, через *средние языковые* области до уровня абстрактного мышления (аналог *фронтал лобе*). Отметим, что нейроны заранее не знают, какую роль будут играть, роли распределяются в результате процесса развития и самоорганизации. Так возникает субъективная условная информация, которая и составляет Личность.

### 3.3 Математическое описание\обоснование

Математическим обоснованием данной схемы является система дифференциальных уравнений, которые были подробно описаны в работах [8][9]. Здесь мы приводим их в несколько модифицированном виде, наиболее соответствующем цели *репродукции личности* (но, для краткости, избегая деталей).

$$\frac{d}{dt} H_i^0(t) = \frac{1}{\tau_H} \{ \mathfrak{S}_H[H] + \sum_j^n \Omega_{ij}^{R0} H_j^0 + \sum_k \Phi_{ik}^{R1} G_k^{R1} - \lambda^R \cdot \Lambda(t) H_i^{typ} + Z(t) \cdot \xi_i(t) \}, \quad (1)$$

$$\frac{d}{dt} H_i^{typ}(t) = \frac{1}{\tau_i^H} \{ \mathfrak{S}_H[H] + \sum_{i \neq j} \Omega_{ij}^{L0} H_j^{typ} + \sum_k \Phi_{ik}^{L1} G_k^{L1} + \lambda^L \cdot \Lambda(t) \cdot H_i^0 \}, \quad (2)$$

где  $H_i^0(t)$ ,  $H_i^{yp}(t)$  — динамические переменные для нейронов на пластине типа Хопфилда (нулевой уровень  $\sigma=0$ ) в ПП и ЛП, соответственно; функционал  $\mathfrak{S}_H[H_i^0]$  описывает их внутреннюю динамику<sup>5</sup>;  $\tau_i^H$  — характерное время активации. Второй член в обоих уравнениях относится к внутри-пластинным взаимодействиям (через «горизонтальные» связи  $\Omega$ ); третий член относится к меж-пластинным (через «вертикальные» связи  $\Phi$ ) взаимодействиям (влияние символов). Следует подчеркнуть что горизонтальные связи в двух подсистемах обучаются по разным правилам:

$$\frac{d}{dt} \Omega_{ij}^{P0}(t) = \frac{\Omega_0^0}{4\tau^\Omega} [H_i^0 + 1] [H_j^0 + 1] \cdot \zeta(t), \quad (3)$$

т.е., согласно правилу Хебба [35] в ПП, и

$$\frac{d}{dt} \Omega_{ij}^{L0}(t) = -\frac{\Omega_0^0}{2\tau^\Omega} [1 - H_i^{typ} \cdot H_j^{typ}] \cdot \zeta(t), \quad (4)$$

по оригинальному хопфилдовскому правилу [32] в ЛП. Эти правила проиллюстрированы на Рис.1 в верхнем правом и верхнем левом углу, соответственно. Здесь  $\Omega_0$  есть характерная величина горизонтальных связей,  $\tau^\Omega$  есть характерное время обучения;  $\zeta(t)$  фактор насыщения, предотвращающим неестественный рост за пределами возможностей системы.

Переменная  $\Lambda(t)$  обеспечивает взаимодействие между подсистемами ПП и ЛП (по аналогии с *corpus callosum*);  $\lambda^R, \lambda^L$  — свободные параметры (см. ниже); положительная  $\Lambda > 0$  означает передачу активности из ПП в ЛП и наоборот.

Член  $Z(t) \cdot \xi_i(t)$  представляет шум (присутствующий только в ПП), где  $Z(t)$  есть амплитуда шума, а  $0 < \xi_i(t) < 1$  есть стандартная случайная функция.

Напомним, что эти пластины представляют *объективную* (т.е. полученную от «органов чувств») информацию, т.е. относятся к сфере 'Brain'. Следующий блок уравнений описывает взаимодействия на уровне *символов*, так что относится к сфере 'Mind'; точечная линия на Рис.1 иллюстрирует это.

$$\frac{d}{dt} G_k^{R\sigma}(t) = \frac{1}{\tau^\sigma} \left\{ \mathfrak{S}_G[G_k^\sigma] + \Upsilon_{hor}^\sigma \{ \Omega_{kl}^{R\sigma}, G_l^{R\sigma} \} + \Upsilon_{vert}^\sigma \{ \Phi_{lk}^{\sigma-1}, G_l^{R(\sigma-1)}, \Phi_{mk}^\sigma, G_m^{R(\sigma+1)} \} - \lambda^R \cdot \Lambda(t) \cdot G_k^{L\sigma} + Z(t) \cdot \xi_k(t) \right\}, \quad (5)$$

где переменные  $G_k^{R\sigma}(t)$  и  $G_k^{L\sigma}(t)$  относятся к нейронам  $\sigma$ -уровня, которые *уже стали символами* их образов (сама процедура выбора символа подробно описана в [9] и здесь опущена для краткости); функционал  $\mathfrak{S}_G[G_k^\sigma]$  представляет их внутреннюю динамику (см. [6]), причем стабильными состояниями являются  $G=+1$  (активное) и  $G=0$  (пассивное). Функционалы  $\Upsilon_{vert}^\sigma \{ \Phi_{kl}^\sigma, \Phi_{kl}^\sigma, G_l^{R(\sigma-1)}, G_l^{R(\sigma+1)} \}$  и  $\Upsilon_{hor}^\sigma \{ \Omega_{kl}^\sigma, G_l^\sigma \}$  описывают вертикальное и горизонтальное взаимодействие, соответственно. Здесь

<sup>5</sup> Этот функционал устроен так, что стационарными состояниями динамического формального нейрона являются  $H=+1$  (активное) и  $H=-1$  (пассивное); детали см. в [6]

$$\Upsilon_{hor}^{R\sigma} \{\Omega_{kl}^\sigma, G_l^\sigma\} = \sum_i \Omega_{ik}^{R(\sigma-1)} \cdot G_l^\sigma, \quad (6)$$

относится к взаимодействию на данном уровне  $\sigma$ , и

$$\Upsilon_{vert}^{R\sigma} \{\Phi_{lk}^{\sigma-1}, G_l^{R(\sigma-1)}, \Phi_{mk}^\sigma, G_m^{R(\sigma+1)}\} = \sum_l \Phi_{lk}^{R(\sigma-1)} G_k^{R(\sigma-1)} + \sum_m \Phi_{mk}^{R(\sigma+1)} G_m^{R(\sigma+1)}, \quad (7)$$

относится к взаимодействию с символами на предыдущем ( $\sigma-1$ ) и последующем ( $\sigma+1$ ) уровнях. Связи  $\Omega(t)$  и  $\Phi(t)$  в ПП обучаются по аналогии с (3):

$$\frac{d}{dt} \Omega_{kl}^{R\sigma}(t) = \frac{\Omega_0^\sigma}{\tau^\sigma} G_k^{R\sigma} \cdot G_l^{R\sigma} \cdot \zeta(t), \quad (8)$$

и

$$\frac{d}{dt} \Phi_{lk}^{R(\sigma-1)}(t) = \frac{\Phi_0}{\tau^\Phi} G_k^{R\sigma} \cdot G_l^{R(\sigma-1)}, \quad \frac{d}{dt} \Phi_{mk}^{R(\sigma+1)}(t) = \frac{\Phi_0}{\tau^\Phi} G_k^{R\sigma} \cdot G_m^{R(\sigma+1)}, \quad (9)$$

где  $\Phi_0$  – характерная величина меж-пластинных (*семантических*) связей,  $\tau^\Phi$  – характерное время их формирования.

Для ЛП уравнения аналогичны:

$$\frac{d}{dt} G_k^{L\sigma}(t) = \frac{1}{\tau^\sigma} \left\{ \mathfrak{S}_G[G_k^{L\sigma}] + \Upsilon_{hor} \{\Omega_{kl}^{L\sigma}, G_l^{L\sigma}\} + \Upsilon_{vert} \{\Phi_{lk}^{L(\sigma-1)}, G_l^{L(\sigma-1)}, \Phi_{mk}^\sigma, G_m^{L(\sigma+1)}\} - \lambda^L \cdot \Lambda(t) \cdot G_k^{R\sigma} \right\}, \quad (10)$$

где все связи обучаются аналогично (4) с учетом того, что  $G^0 \equiv 1/2 * (N^{yp+1})$ .

Подчеркнем, что в этом блоке уравнений динамика переменных  $\Lambda(t)$  и  $Z(t)$  не определена, они выглядят скорее как параметры. Но это далеко не так.

До этого момента система уравнений (1) – (10), равно как и схема на Рис.1, описывала динамику только нейронов *неокортекса*. Согласно принципам ДТИ эта система *не полна*, поскольку не описывает механизм *гомеостаза*, то есть поддержания динамического равновесия при изменении внешних условий.

Выше говорилось, что в головном мозге человека эти функции выполняют подкорковые структуры ('old Brain'), за счет производства необходимых *нейромедиаторов* (стимуляторов или ингибиторов). Более того, одни и те же структуры отвечают за эмоции и одновременно представляют собой источник шума, который в обобщенном смысле является *инструментом приспособления к новым условиям*.

По аналогии с биологическими механизмами мы ввели следующий блок уравнений, отвечающий как за само-регуляцию, так и за эмоции:

$$\frac{d}{dt} \mu(t) = \frac{1}{\tau^\mu} \cdot \{a_{\mu\mu} \cdot (\mu - \mu_0) + a_{\mu Z} \cdot (Z - Z_0) + F_\mu(\mu, Z)\}, \quad (11)$$

$$\frac{d}{dt} Z(t) = \frac{1}{\tau^Z} \cdot \{a_{ZZ} \cdot (Z - Z_0) + a_{Z\mu} \cdot (\mu - \mu_0) + F_Z(\mu, Z) - X\{\mu, G_k^{R,\sigma}\} + \omega D\}, \quad (12)$$

$$\Lambda(t) = -\Lambda_0 \cdot \tanh\left\{\frac{\tau^Z \cdot \tau^\mu}{\tau^\Lambda} \cdot (Z - Z_0)\right\} \cdot \frac{dZ}{dt}. \quad (13)$$

Здесь  $\mu(t)$  — это новая переменная, относящаяся к подкорковым структурам ('old Brain'); она представляет собой разницу между количеством стимуляторов и ингибиторов;  $Z(t)$  — по-прежнему амплитуда шума, который здесь выступает как посредник между 'old Brain' и неокортексом;  $\Lambda(t)$  представляет инструмент регулирования активности подсистем и имеет непосредственное отношение к обеспечению гомеостаза.

В (11)–(12), параметры  $Z_0$  и  $\mu_0$  — значения этих переменных в состоянии динамического равновесия системы, параметры  $\{a\}$  отвечают за регулярный процесс гомеостаза, обеспечивая, согласно ДТИ, затухающие колебания вокруг положения равновесия;  $\tau^Z$ ,  $\tau^\mu$  — характерные времена вариации переменных  $Z(t)$  and  $\mu(t)$ , соответственно. Функции  $F_Z(Z, \mu)$  и  $F_\mu(Z, \mu)$  описывают нелинейное взаимодействие переменных  $Z(t)$  и  $\mu(t)$ , которое дает возможность описать особые состояния типа стресса (см. [7]). Функционал  $\mathfrak{S}\{Z, H, G\}$  связывает изменения  $Z(t)$  с процессами в *неокортексе* (детали описаны в [8][9]).

Уравнение (13) устроено так, чтобы переменная  $\Lambda(t)$  отслеживала *изменения* амплитуды шума  $Z(t)$ ; здесь  $\Lambda_0$  (характерная величина) и  $\tau^\Lambda$  (характерное время изменения) — *свободные параметры* модели. Отметим, что тангенс гиперболический в (13) представляет собой *ступенчатую* функцию, причем характер (динамика) перехода между предельными значениями  $\Lambda = \Lambda_0$  (активна ЛПП) и  $\Lambda = -\Lambda_0$  (активна ПП) зависит, в частности, от соотношения параметров  $\tau^\Lambda$ ,  $\tau^\mu$  и  $\tau^Z$ .

Теперь система (1) – (13) *полна* в математическом смысле и согласуется с общими принципами ДТИ, поскольку учитывает гомеостаз. Подчеркнем, что для этого необходимы переменные относящиеся и к неокортексу, и к 'old Brain', и *посредник* между этими структурами – амплитуда шума.

### 3.4 Эмоции и “Эмоциональная Температура”

Рассмотрим более подробно роль амплитуды шума, т.е. переменной  $Z(t)$ .

Шум, т. е. случайное самовозбуждение нейронов, присутствует только в ПП и играет решающую роль в поиске *новых* решений. Действительно, интенсивность этого поиска напрямую зависит от величины  $Z(t)$ : чем больше амплитуда шума, тем более вероятно участие «случайного» (т. е. не принадлежащего к образам, уже задействованным в попытке решении задачи) нейрона в процессе поиска, тем больше вероятность найти *новое* решение проблемы. Можно сказать, что увеличение  $Z(t)$  соответствует «напряжению ресурсов» системы.

Имея в виду, что эмоции имеют 3 «воплощения» — поведенческое, физиологическое и когнитивное (*осознание*) [15], — естественно предположить, что физиологическое «воплощение» напрямую связано с продукцией нейромедиаторов (адреналина, дофамина и др.). Тогда когнитивное воплощение должно быть напрямую связано с амплитудой шума  $Z(t)$ , которую можно трактовать как *эмоциональную температуру* системы. Развивая аналогию, можно сказать, что увеличение  $Z(t)$  ( $dZ/dt > 0 \leftrightarrow$  реальная температура человека растет) связано с отрицательными эмоциями (тревога, раздражение, страх), а уменьшение  $Z(t)$  ( $dZ/dt < 0 \leftrightarrow$  температура падает) до нормального значения  $Z_0$  вызывает положительные

эмоции (облегчение, радость, удовлетворение). Эта аналогия оказывается почти буквальной.

Рассмотрим поведение переменной  $\Lambda(t)$ , заданное уравнением (13), учитывая специализацию подсистем. В непредвиденных и непонятных ситуациях, требующих поиска новых решений (т.е. генерации новой информации), ПП должна быть активирована ( $\Lambda < 0$ ). Это сопровождается ростом шума ( $dZ/dt > 0 \leftrightarrow$  организм боится) и, соответственно, удалением от равновесного состояния ( $Z - Z_0 > 0$ ). После того, как решение найдено (что сопровождается положительными эмоциями), ПП может «отдыхать»,  $Z(t)$  уменьшается ( $dZ/dt < 0 \leftrightarrow$  температура в организме нормализуется), система переходит в равновесие, при этом активность переходит в ЛП ( $\Lambda > 0$ , релаксация). Однако чрезмерное падение ниже равновесного значения ( $Z - Z_0 < 0 \leftrightarrow$  чрезмерное понижение температуры, упадок сил, депрессия) снова выводит систему из равновесия, что снова требует участия ПП, т.е. повышения  $Z(t)$ . Однако в этом процессе положительная производная  $dZ/dt > 0$  означает не отрицательные эмоции, а, наоборот, — энтузиазм, воодушевление, кураж, *inspiration*.

Таким образом,  $\Lambda(t)$ , определенное в (13), действительно обеспечивает циклическое взаимодействие подсистем, направленное на восстановление динамического равновесия системы, т.е. *гомеостаз*. Механизм гомеостаза здесь связан с *отслеживанием изменений эмоциональной температуры*.

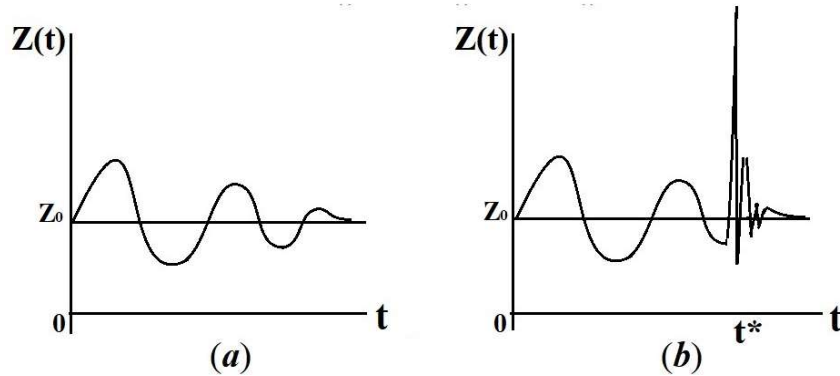
Равновесному состоянию системы соответствует некоторое определенное значение  $Z_0$  («нормальная температура») и определенный состав нейромедиаторов  $\mu \sim 0$  (обычно положительный). Подчеркнем, что эти параметры *индивидуальны* для каждой конкретной системы, что будет важно в дальнейшем.

Что же касается третьей (поведенческой) ‘ипостаси’ эмоций, то она выходит за рамки когнитивного моделирования и должна реализовываться другими средствами.

### **Типичные паттерны поведения амплитуды шума в процессе решении задач.**

Стандартные задачи решаются только в ЛП и не требуют напряжения ресурсов. При этом  $Z(t)$  не меняется, что предполагает отсутствие эмоций (*безразличие*).

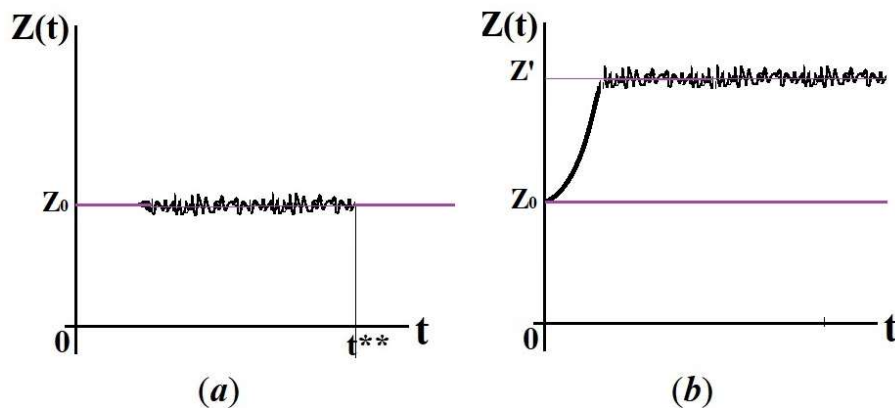
Несколько нестандартные задачи (например, распознавание нестандартного изображения) решаются уже с участием обеих подсистем. Типичная картина регулярного процесса узнавания показана на Рис. 2а. При этом  $Z(t)$  изменяется плавно, что не должно сопровождаться какими-либо особыми поведенческими реакциями.



**Рис. 2.** Типичные паттерны поведения амплитуды шума  $Z(t)$  в случаях: а) процедуры распознавания; б) прогноз в случае шутки (имитация смеха в момент времени  $t^*$ ).

Это не так в случае прогноза по поводу привычной ситуации, прерванного неожиданной (в данном контексте) информацией (например, шутка). В [8][9] было показано, что в случае поступления *не* ожидаемой по ходу прогноза, но тоже знакомой информации, система вынуждена «перескакивать» с одного знакомого образа на совсем другой, но тоже знакомый (рис. 2б). Это сопровождается скачком  $Z(t)$ , которое можно интерпретировать (и иллюстрировать) как *смех*.

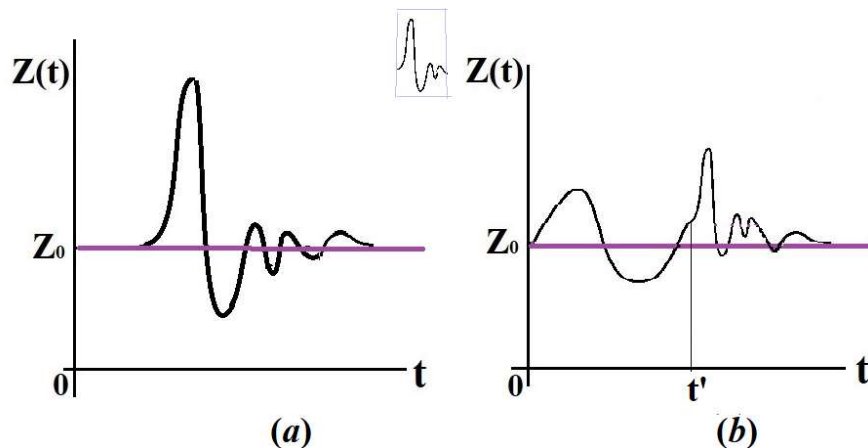
*Паттерн Эстетических Эмоций.* На Рис.3а представлен паттерн  $Z(t)$  в случае эстетических эмоций при ‘созерцании предметов искусства’ (в момент времени  $t^{**}$  созерцание прекращается). В [9] было показано, что они могут быть вызваны рожденными в подсознании смутными (*implicit*) ассоциациями («Brain видит и знает, а Mind не может осознать»), которые сопровождаются малыми колебаниями (вибрация\дрожание)  $Z(t)$  вокруг нормального значения  $Z_0$ . Эта вибрация может интерпретироваться как «мурашки по коже» (*goosebumps*).



**Рис. 3.** Типичные паттерны амплитуды шума  $Z(t)$  в случае: а) переживания эстетических эмоций; б) смутное тревожное раздражение по неясным причинам.

Отметим, что аналогичная картина относится и к случаю *неясного беспокойства* (раздражения), возникающего без видимых причин (Рис. 3б): первоначальный рост  $Z(t)$  до определенного значения  $Z' > Z_0$  (раздражение) сменяется небольшой «вибрацией» вокруг этого значения, которые можно интерпретировать как *внутреннее дрожание*. Здесь снова: «Brain знает, что что-то не так, а Mind не может понять». Отметим, что момента прекращения этого состояния нет, пока Mind не осознает причину.

*Паттерн для общей формулы «Конфликт Плюс Разрешение».* Общая формула эмогенной ситуации, рассмотренная выше, т. е. «конфликт плюс разрешение», должна соответствовать менее радикальным, но очевидным проявлениям (всплеск эмоций). Возможный типичный паттерн («слепок» или «snapshot») представлена на Рис. 4. Такой паттерн должен сопровождаться типичными поведенческими реакциями, разумеется, индивидуальными.



**Рис. 4.** Возможный паттерн  $Z(t)$  («snapshot») для формулы «конфликт плюс разрешение»: (а) — в нормальном состоянии покоя; (б) тот же «слепок», наложенный на регулярный процесс распознавания. Сам «snapshot» представлен в рамке между рисунками.

#### 4 Конструкционный портрет личности в терминах ЕККА

Вернемся к «конструкционному портрету личности», который обсуждался в разделе 2, и обсудим, как его можно воспроизвести в рамках модели ЕККА.

Наиболее просто здесь воспроизводятся два типа мышления: они соответствуют *двум подсистемам*. Логическое мышление обеспечивает подсистема ЛП, а интуитивно-ассоциативное – подсистема ПП.

Действительно, согласно принципам построения этой архитектуры, ЛП хранит полученную информацию (как семантическим, так и эпизодическим способом) и обрабатывает новую информацию в сравнении с общепринятой.

Рассмотрим некую стандартную задачу. Пусть есть начальный образ и образ-цель. Логическое мышление основано на последовательном анализе четких (черных) образов с причинно-следственными связями между ними (основанными на *прямых* ассоциациях через кор-нейроны) — именно то, что мы имеем в ЛП. Образно говоря, имеется карта с множеством дорог и тропинок, и задача состоит в том, чтобы найти кратчайший путь от начального образа к цели.

В ПП, напротив, все образы *размыты*, т.е. имеют некое *гало* благодаря присутствию гало-нейронов со слабыми (серыми) связями, в результате чего хопфилдовский механизм восстановления активирует не прямые (неожиданные, необязательные, побочные) ассоциации. Эти ассоциации могут, казалось бы, уводить от основной цели. Кроме того, присутствие шума приводит к активации совсем посторонних образов, никак не связанных ни с целью, ни друг с другом. Продолжая упомянутую аллегория, — дорог здесь нет, есть перепаханное поле, но и возможность перепрыгнуть произвольно и непредсказуемо между кочками с вероятностью сразу прыгнуть туда, куда нужно (т.е. *insight*).

В Разделе 3.3 было показано, что модель ЕККА содержит достаточно широкий набор *свободных параметров*:

$\Lambda_0, \tau_\Lambda$  — характерная величина и время активации меж-системных связей;

$\lambda^R, \lambda^L$  — параметры «межсистемное асимметрии» (см. ниже);

$\Omega_0, \tau^\Omega$  — характерные величина и время обучения внутри-пластинных («горизонтальных») связей;

$\Phi_0, \tau^\Phi$  — то же для меж-пластинных («вертикальных») связей;

$Z_0, \mu_0$  — амплитуда шума и состав нейромедиаторов в состоянии динамического равновесия (нормальные значения);

$\tau^Z, \tau^\mu$  — характерные времена вариации соответствующих величин.

На самом деле, этот набор еще шире, поскольку характерные параметры связей  $\Omega_0^\sigma$ , и  $\Phi_0^\sigma$  могут отличаться на разных уровнях  $\sigma$ , но это — предмет дальнейшего исследования. Далее мы покажем, что, варьируя эти параметры, можно "настроить" модель ЕККА так, чтобы она максимально соответствовала 'конструкционному портрету' конкретной личности.

#### 4.1 Когнитивные Особенности Личности

**Память.** Схема на Рис.1 представляет «карту памяти» некоторой условной личности. При этом подсистема ПП соответствует *эпизодической памяти*, т.е. тем *знаниям и впечатлениям*, которые человек получает лично, через собственные органы чувств. *Восстановить* их не так просто — только по рассказам человека, его дневникам, письмам и т.д. Отчасти, по воспоминаниям окружающих. Иными словами, источником информации об эпизодической памяти могут быть те же легенды, что и в древнем мире.

Подсистема ЛП содержит как знания (т.е. *хорошо выученную* информацию), транслированные из ПП, так и *семантическую память*, т.е. знания, полученные в «готовом виде» извне. Эти знания приходят уже в символической форме, после



того, как личность освоит какой-либо язык. Восстановить эту информацию достаточно просто, если правильно подобрать обучающее множество – нужные книги, фильмы, музыку и т.д. Личный *лексикон\вокабуляр*, входящий в «конструкционный портрет личности», должен быть специально выделен в обучающем множестве.

**Подсознание.** В модели ЕККА особая роль принадлежит пластине типа Хопфилда  $H^0$  в ПП, точнее, ее специфической части – совокупности гало-нейронов вместе с их серыми связями, которую мы назвали «серое гало». Обсудим это более подробно.

Пластина  $H^0$  в ПП содержит *всю* образную информацию, когда-либо полученную системой на опыте. Затем хорошо выученные отчетливые («черные») образы передаются на пластину  $H^{ур}$  в ЛП и получают свои символы. Та же информация, которая казалась неважной или несущественной в процессе обучения, затрагивает только гало-нейроны, образующие относительно слабые («серые») связи. В результате из смутных образов, смешанных друг с другом, образуется своеобразное серое облако, которое мы называем «серым гало». Именно оно порождает неожиданные личностные *implicit* ассоциации, непонятные в рамках других когнитивных систем. Именно оно может служить источником неординарных и неожиданных (т. е. *не* основанных на запомненной и зафиксированной в ЛП информации) творческих решений, т. е. озарения (*insight*).

Эта информация остается только на уровне Brain, не передаваясь ни в ЛП, ни на следующий символичный уровень. По отношению к системе это некая «вещь в себе»: система ничего о ней не знает.

Поясним. Все символичные пластины относятся, как было показано выше, к понятию Mind, причем символическая информация может быть сформулирована и вербализована. Более того, Mind может управлять и контролировать всю систему: каждый символ может быть активирован либо сверху (через связи с символом более высокого уровня), либо сбоку (через совместные связи с соседями), либо (по крайней мере) снизу (будучи активирован собственным образом). Но это не так в отношении ‘серого гало’ — оно не управляется «сверху» (через Mind): отдельный гало-нейрон может быть активироваться *только случайно!* В то же время само «серое гало» может влиять на процессы «наверху» путем случайной активации каких-то смутных образов, порождающих «невидимые связи» (ассоциации) между отчетливыми (черными), сформулированными и вербализованными образами и процессами. Это естественно интерпретировать как озарение (*insight*).

Сравнивая все сказанное с тем, что известно о человеческом подсознании, естественно сделать вывод, что «серое гало» и есть его реализация.

Эта информация, к сожалению, очень труднодоступна — это так же сложно, как проникнуть в подсознание живого человека (чего он и сам сделать не может). Единственное, что можно сделать, это выявить и зафиксировать необычные (*implicit*) ассоциации человека, оговорки, опечатки и т.п. Заметим, что именно на эти случайные реакции обращал внимание и Зигмунд Фрейд.

**Глобальная цель.** Согласно ДТИ, основной *целью* любой сложной развивающейся системы (в частности, ЕККА) является «генерация, сохранение, защита и распространение собственной субъективной информации» [10]. Поскольку почти все возможные личностные глобальные цели, перечисленные в разделе 2, так или иначе соответствуют этому, ЕККА обеспечивает возможность учета практически любой из них путем варьирования параметров, относящихся к высшим иерархическим уровням  $\sigma \gg 1$ . Однако этот аспект требует специального изучения.

**Воображение, креативность<sup>6</sup>.** Смысл этих терминов интуитивно понятен, но в разных словарях они определяются по-разному. Для того, чтобы интерпретировать эти свойства в терминах ЕККА, мы выбрали наиболее продуктивные (с точки зрения «инженерного» подхода) определения.

Согласно Википедии [37], *воображение* есть «способность человека к спонтанному созданию или преднамеренному построению образов, ... которые в пережитом опыте воображающего в *целостном виде* ранее не воспринимались или же вообще не могут быть восприняты посредством органов чувств». *Креативность* в *Oxford Dictionary* определяется как «использование воображения или оригинальных идей чтобы создать что-то» [38].

На языке ЕККА это можно интерпретировать следующим образом: креативность — это процесс, *обратный* обучению. Воображение означает способность системы самостоятельно (без внешних сенсорных сигналов) создавать *образы* (активное цепочки нейронов) объектов, ранее не виденных («в целостном виде»), а креативность есть стремление привнести эту информацию («плоды воображения») в окружающий мир, т.е. «*обучить мир*».

Отметим, что оговорка «в целостном виде» подразумевает процесс «сборки» новых образов из кусочков, уже имеющихся в памяти (как сборка конструктора «*лего*»); регулярный процесс воображения именно так и происходит. Однако, истинные *шедевры* возникают, если к имеющимся в осознанной памяти (Mind) легикубикам добавляется *нечто*, добытое из подсознания (Brain) и неясное поначалу самому человеку/системе. Именно поэтому в [9] мы определили креативность как «попытку привнести кусочек собственного Brain как в собственный Mind (т.е. понять самому), так и в окружающий Мир».

Очевидно, что за обе функции отвечает ПП, причем параметр  $Z_0$  должен быть по крайней мере не меньше остальных характеристик взаимодействия, т.е.  $\Omega_0$  и  $\Phi_0$ . Возможно, и  $\Lambda_0$  тоже. Однако если  $Z_0 \gg \Omega_0, \Phi_0$ , мышление из креативного переходит в «скачкообразное» и вряд ли продуктивное.

Соотношение параметров  $Z_0$  и  $\Lambda_0$  заслуживает специального анализа. Возможно,  $Z_0 \gg \Lambda_0$  можно интерпретировать как «уход в астрал» — характерная для некоторых творческих людей *отрешенность* от реальных регулярных бытовых проблем (за решение которых отвечает ЛП) и игнорирование необходимости активации ЛП. Однако этот вопрос заслуживает специального исследования.

В [9] было показано, что особое состояние под названием «*муки творчества*» можно интерпретировать как экстремальное увеличение шума  $Z(t)$  до некоторой

<sup>6</sup> Эти проблемы подробно рассматривались в [9]

величины  $Z^{**} \gg Z_0$ , что приводит к *хаосу*, т.е. хаотичным скачкам амплитуды шума  $Z(t)$  — вместе с эмоциями и со всеми другими составляющими когнитивного процесса — от эйфории к отчаянию и обратно. Именно этот хаос, (в согласии с Ницше [39]) и помогает *добывать* ('выкапывать') из Brain решение: "You have to have chaos inside you to give birth to a dancing star".

**Тип мышления.** Оба рассмотренных типа мышления имеют и преимущества, и недостатки, и в «чистом виде» встречаются крайне редко. Как правило, работает смешанный тип мышления, в котором присутствуют оба механизма, т.е. работают поочередно и в зависимости от ситуации обе подсистемы. Тип и стиль мышления конкретной когнитивной системы зависит в общем случае от соотношения параметров  $\Omega_0, \Lambda_0, \sim \Phi_0$  и  $Z_0$ . Обычно предполагается:

$$\Omega_0 \sim \Phi_0 \sim \Lambda_0 \sim Z_0,$$

что означает равное участие обеих подсистем, равно как и всех рассмотренных механизмов взаимодействия нейронов, в процессе мышления. Серьезная асимметрия параметров приводит к ярко выраженным особенностям в характере мышления. Например, если  $Z_0 \gg \Omega_0$ , то ассоциативное мышление меняется на *скачкообразное*: влияние соседних нейронов (направленное на завершение определенного образа) становится незначительным по сравнению со случайным самовозбуждением "посторонних" нейронов. В противоположном случае  $Z_0 \ll \Omega_0$  "скачки мысли" практически отсутствуют, однако вероятность озарения при этом сводится к нулю.

**Логическое vs Интуитивное.** Преобладание логического или ассоциативного типа мышления можно воспроизвести следующим образом.

В уравнениях (1) и (2) присутствуют параметры  $\lambda^R$  и  $\lambda^L$ , которых не было в наших предыдущих публикациях. По умолчанию считалось, что  $\lambda^R = \lambda^L = 1$ , т.е. ПП и ЛП в равной степени влияют друг на друга. При  $\lambda^R = \lambda^L = 0$  (или  $\Lambda_0 = 0$ , что то же) подсистемы вообще «не видят» друг друга и существуют как некие независимые (и ущербные) когнитивные системы. Эта ситуация соответствует прерыванию связей (распиливанию) *corpus callosum*, которое применялось раньше в медицине для лечения эпилепсии и часто приводило к «раздвоению личности». Варьируя  $0 < \lambda^R, \lambda^L < 1$ , можно имитировать разные стили мышления.

При  $\lambda^R \ll 1$  — преобладает ассоциативное (ПП) мышление: ПП *не видит* ЛП, она работает *всегда и сама по себе*. При этом ЛП получает информацию, но не может ни «позвать на помощь» ПП, ни управлять ею. Этот стиль мышления можно интерпретировать как «легкомыслие».

При  $\lambda^L \ll 1$  — преобладает логическое мышление (ЛП): здесь напротив, ЛП не видит ПП и не использует «подсказки» оттуда. Это можно интерпретировать как «ригидность мышления» или *упрямство, самодовольство, самоуверенность*.

Следует подчеркнуть, что скорость реакции на изменение окружающей среды, являясь яркой и заметной характеристикой личности, зависит от соотношения между характерным временем переключения  $\tau^A$  и характерными време-

нами изменения шума  $\tau^Z$  и «гормонального фона»  $\tau^\mu$ . Если  $\tau^\Lambda \ll \tau^\mu \sim \tau^Z$ , переключение ЛП→ПП происходит почти скачком, что соответствует «мгновенной реакции». В противном случае большое время переключения  $\tau^\Lambda$  приводит к «замедленной реакции».

*Стиль Мышления.* Вне зависимости от того, что преобладает, логика или интуиция, отличительной чертой человека является еще и *стиль мышления*. Ранее мы обсуждали два отличительных стиля: «вертикальный» (ведущий прямо к цели) и «горизонтальный» (ведущий посредством параллельных ассоциаций). В рамках ЕККА их очень легко воспроизвести, так как это зависит от соотношения параметров  $\Omega_0$  и  $\Phi_0$  (см. раздел 3.3). При  $\Omega_0 \gg \Phi_0$  доминирует «горизонтальный» стиль, и наоборот. Повторимся: разница в стилях очень заметна в личном общении.

#### 4.2 Эмоциональные Особенности Личности

В разделе 3.4 было показано, что эмоции и эмоциональные реакции связаны с поведением эмоциональной температуры  $Z(t)$ . Поскольку  $Z(t)$  является внутренней переменной системы, ее поведение необходимо как-то визуализировать, т.е. сопоставить с индивидуальными поведенческими реакциями. Эти спецэффекты выходят за рамки когнитивного моделирования.

**Вокабуляр Эмоций.** Было также показано, что, несмотря на все кажущееся разнообразие, можно выделить несколько типичных паттернов поведения  $Z(t)$  (подобных тем, которые представлены на Рис. 2–4), проявляющихся в различных когнитивных процессах. То же верно и для индивидуума: у него есть лишь несколько индивидуальных паттернов, характерных для проявления той или иной эмоции (эмоциональный snapshot) — набор определенных интонаций, мимики, жестов и т. д. Именно эти «слепки» (или snapshot) следует сопоставлять с типичными паттернами поведения  $Z(t)$  — так формируется *эмоциональный вокабуляр*. Именно им следует снабдить желаемую «виртуальную личность» для визуализации общения с ней.

**Темперамент.** Несмотря на то, что темперамент рассматривается как характеристика поведения, он очевидно влияет на эмоциональные реакции. Теперь можно показать, как можно воспроизвести особенности темперамента в рамках ЕККА. Очевидно, темперамент контролируется свободными параметрами модели  $Z_0, \mu_0$ , характеризующими состояние динамического равновесия системы (состояние покоя), а также  $\tau^Z$  и  $\tau^\mu$ , являющимися характерным временем изменения соответствующих переменных.

Значение параметра  $\mu_0$  может быть как положительным (больше стимуляторов), так и отрицательным (больше ингибиторов); он представляет собой те самые «жизненные соки», о которых говорил Гиппократ.

Величину  $Z_0$  следует сравнивать не с  $\mu_0$ , а с ее неокортикальными партнерами  $\Omega_0$  и  $\Phi_0$ , так что (как ни странно) к темпераменту она, похоже, не имеет никакого отношения (скорее только к творческим способностям).

Поскольку  $\mu(t)$ , являясь продуктом old Brain, быстрее (импульсивнее) реагирует на изменение внешней среды, то устойчивость (баланс) системы зависит от того, как быстро начинается этот «гормональный» всплеск (импульс) воздействовать на когнитивные процессы посредством изменения  $Z(t)$ . Если  $\tau^Z < \tau^\mu$ , система начинает «дергаться» еще до того, как осознает это («бежит впереди локомотива»); для нормально функционирующей равновесной системы характерно обратное соотношение  $\tau^Z > \tau^\mu$ .

Теперь мы можем интерпретировать и даже пытаться воспроизвести различные темпераменты в индивидуальных когнитивных системах. В терминах ЕККА четыре типа темперамента по Гиппократу вместе с четырьмя типами нервной системы по Павлову соответствуют:

- сангвиник (активный, веселый) – сильный, устойчивый тип —  $\mu_0 > 0$ ,  $\tau^Z > \tau^\mu$ ;
- холерик (импульсивный, горячий) — сильный, неустойчивый тип —  $\mu_0 > 0$ ,  $\tau^Z < \tau^\mu$ ;
- меланхолик (грустный, подавленный) — слабый, неустойчивый тип —  $\mu_0 < 0$ ,  $\tau^Z < \tau^\mu$ ;
- флегматик (спокойный, медлительный) — слабый, устойчивый тип —  $\mu_0 < 0$ ,  $\tau^Z > \tau^\mu$ .

#### 4.3 Поведенческие Особенности Личности

Как было сказано выше, эти особенности выходят за рамки когнитивного моделирования. Составленный поведенческий вокабуляр должен быть реализован другими средствами и поставлен в соответствие с «вокабуляром эмоций».

## 5 Выводы и Дискуссия

Итак, мы показали, что проблему «цифрового бессмертия для себя» решать не интересно, не возможно, и не нужно. Проблема «бессмертия в сознании людей» представляется гораздо более реалистичной. В каком-то смысле мы приходим к тому, что известно человечеству уже много веков: бессмертие человека заключено в легендах, рассказах, фотографиях, видеосюжетах, его памятных шутках, фразах, оговорках и т. д. Другими словами, бессмертие заключено в творениях человека и памяти, которая остается после него.

Однако современные цифровые (нейросетевые) технологии позволяют продвинуться гораздо дальше и дать новый импульс легендам и сказкам, а именно создать «бота», наделенного чертами данного человека.

Эта цель распадается на две (равнозначные) подцели:

- Сбор информации и составление «конструкционного портрета» личности.
- Воспроизведение этого портрета с использованием современных нейросетевых технологий.

Следует подчеркнуть, что обе эти *под*-цели по-своему важны и интересны. По собственному опыту мы можем сказать, что первая подцель («накопать» информацию и воссоздать портрет) волнует и вдохновляет людей, знавших этого человека (или даже просто слышавших о нем), тем самым уже возрождая память о нем. В итоге оказывается, что информации сохранилось гораздо больше, чем мы полагали заранее.

Вторая подцель может быть достигнута двумя способами:

— путем *моделирования* когнитивного процесса (включая эмоциональную составляющую) с учетом выявленных особенностей данной личности (т. е. с применением составленного конструкционного портрета);

— путем *имитации* персональных реакций.

Последнее означает создание «бота», обученного «репертуару» (правильно подобранное обучающее множество, включая лексику) данного человека. Такой бот может производить впечатление, что он думает так же, как этот человек, по крайней мере, обсуждает проблемы в той же терминологии, что и его прототип. Если снабдить такого бота эмоциональным вокабуляром (в данном контексте, репертуар мимических ‘эмоциональных’ реакций) и характерными реакциями из реконструированного «поведенческого репертуара», впечатление может значительно усилиться. Это не кажется особенно сложным, но может иметь яркий эффект. Такой «бот» может быть дорог (во всех смыслах) в основном родственникам и друзьям данного человека, а насколько он оправдан и доступен, судить им.

Первый способ сложнее. Однако было показано, что ЕККА дает возможность моделировать когнитивный процесс *данного конкретного человека*. Она содержит широкий набор свободных модельных параметров, позволяющих учесть выявленные особенности когнитивного процесса вместе с его эмоциональными аспектами. Затем “вокабуляр эмоций” (коллекцию типичных «снимков» типичных эмоций) надо привести в соответствие с поведенческим (мимическим) репертуаром для визуализации.

Отметим, что мы не рассматривали здесь влияние разнообразия параметров высокого уровня ( $\sigma \gg 1$ ) на особенности мыслительного процесса. Это может стать предметом дальнейших исследований.

Мы полагаем, что предлагаемая программа способна дать возможность создать виртуальную личность («цифровой двойник»), которая могла бы создавать иллюзию общения с данным конкретным человеком.

## References

1. Gordon, B., Gray J.: Digital Immortality. Microsoft Research. Technical report MSR-TR-2000-101, October 2000.
2. Galvao, V.F. et. al.: Discussing human values in digital immortality: towards a value-oriented perspective. *Journal of Brazilian Computer Society* 27, 15–30 (2021).
3. Savin-Baden, M., Burden D., Taylor H. The Ethics and Impact of Digital Immortality. *Knowledge Cultures* 5(2), 10 – 29 (2017).
4. Adali, S., Golbeck, J.: Predicting personality with social behavior: a comparative study. *Social Network Analysis and Mining*, 4(1), 159–179 (2014).

5. Personality: Oxford Learner's Dictionary: <https://www.oxfordlearnersdictionaries.com/definition/english/personality#>
6. Chernavskaya, O. D., Chernavskii, D. S., Karp, V. P., Nikitin, A. P., Shchepetov, D. S.: An architecture of thinking system within the Dynamical Theory of Information. *BICA* 6, 147—158 (2013).
7. Chernavskaya, O. D. et al.: An architecture of the cognitive system with account for emotional component. *BICA* 12, 144—154 (2015).
8. Chernavskaya, O., Rozhylo, Ya.: The Natural-Constructive Approach to Representation of Emotions and a Sense of Humor in an Artificial Cognitive System. *IARIA Journal of Life Sciences*, 8 (3&4), 184—202 (2016).
9. Chernavskaya, O., Rozhylo, Ya.: On Modeling the Creativity and the Concept of Chef-D'oeuvre. *Cognitive System Research* 68, 18—33 (2021).
10. Чернавский Д. С. : Синергетика и Информация. Динамическая теория информации. Изд. 3-е, доп. — М: Книжный дом ЛИБРОКОМ (2009). ISBN 978-5-397-00207-3.
11. Subconscious: <https://www.merriam-webster.com/dictionary/subconscious>, last accessed 2023/04/06.
12. Талант: <https://ru.wikipedia.org>
13. Intuition: <https://ru.wikipedia.org/intuition>
14. Intuition: Britannica Dictionary. <https://www.britannica.com/dictionary/intuition>
15. Emotions: Oxford Language Dictionary <https://www.google.com/search?q=Emotions+definition&oq>
16. Reber, A.S.: Dictionary of psychology. London, Penguin/Viking. Second edition (1995); А Ребер. Большой толковый психологический словарь
17. Emotions: Bratannica <https://www.britannica.com/science/emotion>
18. Panksepp, J., Biven, L.: *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. Norton, N.Y. (2012).
19. Koziol, L. F., Budding, D. E.: *Subcortical Structures and Cognition. Implications for Neurophysiological Assessment*. Springer, Heidelberg (2009).
20. Samsonovich, A.: Emotional biologically inspired cognitive architecture. *BICA* 6, 109 – 125 (2013).
21. Vallverdu, J. et al.: A cognitive architecture for the implementation of emotions in computing systems. *BICA* 15, 34 – 40 (2016).
22. Sellers, M.: Toward a comprehensive theory of emotion for biological and artificial agents. *BICA* 4, 3—26 (2013).
23. Hudlyka, E.: Affective BICA: Challenges and open questions. *BICA* 7, 98 – 125 (2014).
24. [https://ru.wikipedia.org/wiki/Чувство\\_юмора](https://ru.wikipedia.org/wiki/Чувство_юмора)
25. <https://en.wikipedia.org/wiki/Irony>
26. Prigogine, I., Nicolis, G.: *Self-Organization in Non-Equilibrium Systems*. Wiley (1977).
27. Haken, H.: *Information and Self-Organization: A macroscopic approach to complex systems*. Springer, Heidelberg (2000).
28. Graeme, S.: *A General Theory of Complex Living Systems: Exploring the Demand Side of Dynamics*. *Complexity (Journal of the Santa Fe Institute)* 13, 12-20 (2008). 10.1002/cplx.20225.
29. Shannon, C.: *A Mathematical Theory of Communication*. *Bell System Technical Journal* 27, 379–423 & 623–656, July & October (1948).
30. Quastler, H.: *The emergence of biological organization*. Yale University Press, NH (1964).
31. <https://www.britannica.com/science/homeostasis>
32. Hopfield, J. J.: Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79, 2554—2558 (1982).

33. Grossberg, S.: *Studies of Mind and Brain*. Riedel (ed), Boston (1982).
34. Goldberg, E.: *Creativity: The Human Brain in the Age of Innovation*. Carlisle, NY (2018).
35. Hebb, D. O.: *The Organization of Behavior: Neuropsychological Theory*. Wiley & Sons (1949).
36. Kohonen, T.: *Self-Organizing Maps*. Springer, Heidelberg (2001).
37. Воображение: <https://ru.wikipedia.org>
38. <https://www.oxfordlearnersdictionaries.com/definition/english/creativity#:~:text=creativity>  
Nietzsche, F.: *Thus Spoke Zarathustra*. Trans. [Alexander Tille](#). NY: Macmillan (1896).



## 6    **Аппендикс. Предварительный конструкционный портрет Д. С. Чернавского**

Для реализации предложенной программы необходимо собрать следующую информацию оп *Вашему мнению и воспоминаниям.*

Огромная просьба от авторов статьи: *заполнить следующий лист.*

- **Обучающее множество:** *Какие книги\фильмы\музыку Он любил.*
- **эпизоды из биографии** *(что Вы помните)*

• **Тип и стиль мышления:** *по Вашему мнению, более логический или интуитивный?   Ход рассуждений: «горизонтальное» или «вертикальное»?*

- **особенности чувства юмора и иронии** *(примеры)*

- **глобальная цель:** *по Вашему мнению*

- **лексикон\вокабуляр:** *характерные фразы, которые Вы помните*

- **вокабуляр эмоций (ОЧЕНЬ желательны фотографии);**

- **темперамент** – *по Вашему мнению*

- **поведенческий репертуар (набор используемых шаблонов)**  
*Примеры, которые Вы помните;*

- **репертуар\вокабуляр привычек\жестов** *которые Вы помните*

- **репертуар «восклицаний»** *(в широком смысле, что Вы помните)*