

## **Шестнадцать лекций по курсу модели поведения, восприятия и мышления.**

(Лекции читаются студентам факультета ИВТ Московского физико-технического института.  
Курс в значительной степени основан на материале книги А.Л.Шамис, Модели поведения восприятия и мышления. М. Интуит, 2010)

### **Содержание**

#### **Л1. Введение.**

Тайны и загадки мира  
Может ли машина мыслить?  
Сократ. Круги знания.  
Подходы к определению мышления.  
Тест Тьюринга.  
Предварительное определение мышления.

#### **Л2. О работах в области моделирования мышления.**

Кибернетика.  
Бионика.  
Алгоритмический Искусственный интеллект  
Распознавание образов  
Формальные нейронные сети  
Синергетика  
Виталика  
Предварительное определение мышления.

#### **Л3. Существующие предпосылки к инженерному пониманию поведения и мышления.**

Общая оценка результатов работ по ИИ и в смежных областях  
Что такое жизнь?  
Принцип устойчивого неравновесия Э.С. Бауэра  
К определению жизни  
Поведение  
Принцип доминанты А.А. Ухтомского  
Теория функциональной системы П.К. Анохина  
Эксперименты Н.В. Асмояна и Г.А. Голицына  
Роль эмоционального механизма принятия управления поведением  
Общая оценка результатов по изучению поведения

#### **Л4. Формальная модель поведения.**

Характеристика задачи поведения. Внутренняя и внешняя работа.  
Принципы оптимальности.  
Конкретизация задачи поведения  
Принцип maxT  
Формальная модель  
Реальная задача  
Свобода поведения систем с активно поддерживаемым неравновесием.

#### **Л5. Физиологические интерпретации формальной модели поведения.**

- Мотивации и потребности.
- Формулы эмоций
- Выводы

#### **Л6. Распознавание образов и особенности живого восприятия.**

- Распознавание
- Психология машинного зрения.

#### **Л7. Принципы целостного целенаправленного распознавания и их реализация в программах ABBYY FineReader-рукопись и FormReader**

- Задача
- Общие принципы
- Процесс
- Модель
- Предварительная обработка изображений

#### **Л8 Организация.**

- Целостность и системность
- Организация и дезорганизация
- Представление целостного объекта
- Принципы организации
- Принцип maxT в поведении живых и искусственных систем

#### **Л9 Информация.**

- Подходы к определению и измерению информации.
- Ценность информации.
- Познание и информационное отображение мира
- Информация и потребности

#### **Л10 Нейронные модели**

- Свойства и функции нейронных моделей, необходимые для решения базовых задач мышления
- Минимальные сведения об элементах мозга
- Логические нейронные сети Мак-Каллока и Питса
- Перцептрон Ф.Розенблатта

#### **Л11. Признаковые распознающие нейронные сети.**

- Распознавание с помощью разделяющих гиперплоскостей
- Формальная нейронная сеть как модель механизмов мозга
- Нейрокомпьютер

#### **Л12. Активные нейронные модели (А-сети).**

- Активность
- Гипотеза Емельянова-Ярославского
- Вариант свойств нейронов экспериментальной модели (А-сети).
- А-сеть с возвратным возбуждением и возвратным торможением.
- Преактивация и предвидение

#### **Л13. Автомат для распознавания и воспроизведения временных последовательностей**

- Задача
- Схема автомата
- Обучение и воспроизведение опыта
- Вариант с опорными кодами

Эксперимент  
Вероятностная схема  
Возможность реализации в активной нейронной модели

#### **Л14. Основные проблемы, связанные с моделированием нейронных механизмов мозга.**

Еще раз об общей оценке существующих нейронных моделей  
Сознание  
Умозрительное моделирование и свобода воли  
Возникновение и гашение очагов возбуждения  
Получение и запоминание признаков, отношений и метрических характеристик  
Управление процессами в модели с понятийного уровня.  
Построение целостных многоуровневых иерархических описаний  
Проблема моделирования памяти  
Переход от нейрона к нейронному ансамблю  
Проблема моделирования языка и абстрактного мышления  
Проблемы второго плана  
Выводы

#### **Л15 Виды мышления и гипотезы о некоторых механизмах мышления и творчества.**

Общие замечания  
Необходимые свойства формируемой в мозге модели проблемной среды  
Перцептивное мышление  
Когнитивное (познавательное) мышление  
Практическое (поведенческое) мышление  
Структурно-логическая особенность модели проблемной среды человека  
Механизмы простого репродуктивного практического мышления  
Репродуктивное творческое мышление  
Созидательное творческое мышление  
Работа мозга и некоторые функции творчества  
Ненаправленная умственная работа

#### **Л16. Заключение.**

Моделирование мышления миф или реальность?  
Заключение

#### **Приложение**

Актуальность проблемы - риск создания машинного интеллекта.

#### **Литература.**

#### **Лекция 1. Введение**

В этом курсе мы будем говорить о моделях поведения, восприятия и мышления. Естественно, наиболее важной и интересной для нас будет проблема понимания и моделирования мышления. В первой лекции мы затронем вопрос об определении мышления и будем возвращаться к этому вопросу еще много раз на протяжении всего курса.

## **Тайны и загадки мира**

Мышление – это одна из многих интереснейших загадок мира. В начале 20го века среди физиков было распространено мнение, что все физические законы уже известны и в научных исследованиях речь может идти лишь о количественных уточнениях. После этого появились теория относительности, квантовая механика, атомная энергетика и многое другое. Сейчас опять иногда начинают говорить о том, что мы почти все уже знаем и близки к построению “окончательной теории”, объясняющей все в нашем мире. Это заблуждение. Мы практически ничего не знаем о важнейших вещах, касающихся устройства мира. Мы не знаем что такое электричество, внутриатомные взаимодействия, гравитация, магнетизм и распространяющееся в пустом пространстве поле. Мы ничего не знаем об общем устройстве мира и действующих в нем процессах управления и организации, заменяя иногда это знание упрощенными не требующими доказательства или проверки религиозными представлениями. Мы не понимаем, что такое бесконечность и вечность и соответственно не знаем что такое пространство и время. Вполне возможно, что мы ничего не знаем об очень многих вещах, не взаимодействующих с известными нам физическими объектами и потому никак не проявляющих себя в нашем мире.

В ряду этих вопросов стоят загадки жизни и разума. При этом не известно, являются ли существующие на земле жизнь и разум чем-то общим, проявляющемся и значимом во всем мире, или это частные феномены, имеющие значение только для окружающей Землю мельчайшей части бесконечного мира.

В любом случае загадки о жизни и разуме очень интересны. Они не просто интересны, они еще и актуальны. Актуальность проблемы определяется не только перспективой практического использования, но и существующими предпосылками к ее решению. О предпосылках к решению проблемы мышления мы и будем говорить.

## **Может ли машина мыслить?**

Мы будем говорить о том, что такое мышление и можно ли это реализовать вне мозга, т.е. техническими или программными средствами. В связи с этим нас в первую очередь будут интересовать различия между живым и неживым и являются ли эти различия принципиальными. К ответу на этот вопрос нас сможет подтолкнуть рассмотрение некоторых моделей поведения, восприятия и процесса мышления, а также модели некоторых нейронных механизмов мозга. К сожалению, в большинстве случаев мы вынуждены будем ограничиваться качественным (неформальным) уровнем рассмотрения проблем. Это может затруднять понимание, делая его неоднозначным. С другой стороны, неоднозначность может быть полезна, оставляя простор для воображения и фантазии. Многое из того, о чем мы будем говорить, является не устоявшимися признанными фактами, а гипотезами. Будет хорошо, если эти гипотезы подвергнутся вашей критике или вызовут у вас вопросы.

Определенные сведения по рассматриваемым проблемам вы уже получили после того, как прослушали лекции по интеллектуальным системам, распознаванию образов и нейронным сетям. О моделировании интеллекта написаны тысячи книг, с частью из которых вы уже знакомы и, может быть, изучали. В частности, возможно, вы знакомы с вышедшей на русском языке книгой Тьюринга “Может ли машина мыслить?”

В середине прошлого века на фоне бурного начального этапа развития вычислительной техники и программирования в средствах массовой информации, в популярных научных и околонуучных изданиях горячо обсуждался вопрос: может ли машина мыслить? Спорящие разделились на два лагеря - физиков и лириков. Физики это те, кто на вопрос - может ли машина мыслить отвечали да, лирики это те, кто отвечали нет. За ответом да следовали обвинения в грубом механицизме, за ответом нет - в идеализме. Спорящие были уверены в своей правоте, хотя серьезной доказательной аргументации ни в первом, ни во втором случае не было. Так аргументы лириков в основном опирались на утверждение о несводимости психологического уровня рассмотрения к физиологическому, а физиологического к техническому, поскольку мышление и жизнь это особые формы движения материи. (Помните Чехова “не может быть потому, что не может быть никогда”). Аргументы физиков сводились в основном к проповеди материализма и отрицанию идеализма. Все это особого смысла не имело, поскольку не был определен предмет спора. Правильным был бы ответ - не знаю или запрашивающийся встречный вопрос - а что такое мышление? Последнее равносильно ответу - не знаю. Вообще-то неплохо было бы ответить и на вопрос – а что такое машина, или что такое не машина, т.е. жизнь? Ответ также не очевиден.

## **Сократ. Круги знания.**

Плохо ли отвечать "не знаю". Нет не плохо. Плохо это абсолютная уверенность. Существует рассказ о Сократе и его ученике. Говорят, что ученик как-то сказал Сократу: учитель, почему вы так часто говорите "не знаю". Я ваш ученик говорю "не знаю" гораздо реже. В ответ Сократ нарисовал два круга - большой и маленький и сказал. Внутри большого круга то, что знаю я, внутри маленького - то, что знаешь ты. На границе большого круга - то, чего не знаю я, на границе малого круга – то, что не знаешь ты. Как видишь, я не знаю гораздо больше чем ты.

К кругам Сократа можно добавить следующую интерпретацию. Внутри круга знания обладающего мозгом субъекта находится в основном целостная субъективная модель его мира. Модель, во-первых, строится в процессе прямого эксперимента, т.е. она отражает опыт взаимодействия с окружающей средой. Кроме того, если субъект это человек, то значительная часть субъективной модели мира строится на основе уже имеющейся у него модели с добавлением информации, приобретаемой из книг, а также с помощью других людей, в том числе учителей или

проповедников. Информация, лежащая в круге знания, всегда неполна, неточна и субъективна. Поэтому этот круг можно назвать кругом субъективного относительного знания.

На границе круга знания лежит область относительного или актуального незнания. Это то, о чем можно сформулировать какие-то вопросы с использованием имеющегося знания. За пределами области актуального незнания лежит область абсолютного незнания, содержащая неизвестно что.

Пример. Электричество во времена Пушкина у всех людей лежало в области абсолютного незнания. Сейчас у студента физтеха электричество частично лежит в области знания и частично в области актуального незнания. Правда, возможно, что для получения полной разгадки тайны электричества нужно заглянуть в область абсолютного незнания. Можно допустить, что в области абсолютного незнания может существовать что-то вполне материальное, но не просто пока еще неизвестное, но принципиально человеком непознаваемое. Такое допущение может отчасти примерить материалистов и идеалистов.

Пример. Может ли кролик получить хоть какое-то представление о поэмах Гомера? Нет, не может. Модель мира кролика (круг знания) не имеет и не может иметь никаких пространственных и смысловых пересечений с поэмами Гомера, лежащими для кролика в области абсолютного незнания и в подобласти "непознаваемое".

### **Подходы к определению мышления**

На интуитивном бытовом уровне слова мышление и жизнь всем понятны, однако четких и полных определений не существует. Есть надежда, что для человека проблемы "жизнь" и "мышление" лежат в области актуального (относительного) незнания. Поэтому сегодня точное исчерпывающее определение этих понятий и, в первую очередь, мышления невозможно, но станет возможным в будущем. Тем не менее, занимаясь исследованием и моделированием мышления, или только оценкой перспективности этих исследований, нужно хотя бы в каком-то приближении дать определение мышления. Иначе не будет понятно о чем идет речь.

Предлагалось много разных преимущественно индуктивных определений, строящихся на перечислении признаков или характеристик явления. Обычно важными характеристиками мышления считаются: память, обучение, наличие семантической модели проблемной среды, индуктивный вывод, дедуктивный вывод, синтез – анализ, ассоциации по сходству, ассоциации по смежности во времени, интуиция, инсайт (озарение), способность решения неалгоритмических творческих задач, сознание. Этот список, получаемый на основе психологических экспериментов или самоанализа, можно расширять, например, добавляя целенаправленность, волю, воображение, умозрительное конструирование, активность, можно и сокращать, поскольку неизвестно какие характеристики мышления являются необходимыми. В любом случае такое определение может быть лишь очень приблизительным, поскольку входящие в перечень характеристики в свою очередь не являются строгими понятиями и чаще всего сами нуждаются в определении. Перечисленные и другие аналогичные характеристики, выявляемые методами нейрофизиологии или психологии, не давая точного определения мышления, безусловно, полезны, указывая направления возможных теоретических и экспериментальных работ.

### **Тест Тьюринга**

Итак, ответ на вопрос что такое мышление неизвестен. Общего точного определения не существует, поэтому на этот вопрос можно давать разные ответы. Можно и не давать никаких ответов, а просто использовать тест Тьюринга для определения мыслит или нет конкретная программа или машина. В соответствии с этим тестом важен только результат процесса мышления и не важно, как он достигается. Тест состоит в следующем. Есть образец – мышление человека, с которым сравнивается мышление машины. Конкретная программа (машина) мыслит, если человек ведя с ней диалог, не может определить, с кем он общается, с машиной, или с человеком.

Может ли компьютерная программа, созданная в рамках алгоритмического подхода, развиваемого в направлении искусственный интеллект, пройти тест Тьюринга? Такую программу написать можно. Трудности будут состоять в том, что человек пытаясь 'разоблачить' машину будет ориентироваться не только на содержание ответов и вопросов, но и на сложность фраз, экспрессию, эмоциональную окраску текста, чувство юмора и т.п. Но эти трудности преодолимы. Иногда говорят, что с момента когда машины начнут проходить тест Тьюринга, начнется соперничество и борьба за выживание между машиной и человеком. (называются близкие регулярно отодвигаемые сроки, например, 2030 или 2050). Будут ли машины (программы) прошедшие тест Тьюринга соперничать с человеком и стремиться к его подчинению? Нет, не будут, если они останутся чисто алгоритмическими пассивными системами, не имеющими потребностей, целей, желаний и эмоций. Например, не станут системами, которые не только умеют выигрывать в шахматы у человека, но и хотят это делать. В этом аспекте нужно говорить не о перспективах создания искусственного разума, а о перспективах создания искусственной жизни.

Тест Тьюринга не является определением мышления. Делаящаяся оценка, основывающаяся на внешнем сравнении результатов "мышления" машины и человека, может быть лишь очень приблизительной, поскольку никак не очерчивается круг тестовых задач (вопросов). Кроме того, возможно, что мышление это не столько умение решать задачи, сколько способ решения. Нужно подчеркнуть также еще и то, что тест Тьюринга направлен на то, чтобы установить обладает ли конкретная машина "человеческим" мышлением, то есть точно таким мышлением, каким обладает человек. При этом остается вопрос, является ли мышление человека единственным

вариантом. И, наконец, тест Тьюринга ровным счетом ничего не дает для ориентации исследований в области моделирования мышления.

### **Предварительное определение мышления**

Не являются удовлетворительными и упоминавшиеся ранее индуктивные определения, хотя в отличие от теста Тьюринга такие определения, безусловно, полезны для общего понимания проблемы и могут использоваться при планировании необходимых направлений исследований.

Может быть, к ответу на вопрос - что такое мышление, полезно идти от анализа задач, для решения которых возник и эволюционно развивался мозг. Живым организмам, в том числе и человеку, для организации эффективного целенаправленного поведения и, в конечном счете, для выживания необходимо приобретение и использование знаний о своей среде обитания. Приобретаемые знания формируют в мозге активную модель окружающего мира. Эта модель используется для восприятия среды, управления поведением в среде и для наращивания самой модели. Строя в мозге модель мира обладающий мозгом субъект отражает в модели выявляемые причинно-следственные связи между объектами и явлениями мира, включая и самого себя как объект мира, строит иерархии обобщений по линии частное-общее и укрупнений по линии часть-целое. Подчеркнем, что в основе построения модели мира лежит эксперимент, определяющий связи между причинами и следствиями. Средство для построения модели мира это мыслящий мозг, конструирующий эксперимент по выявлению причинно-следственных связей в мире и отражающий в себе результаты эксперимента. Интересно понять, как работает это средство и что с его помощью можно узнать о мире.

Таким образом, будем считать, что мышление это функция мыслящего мозга, строящего на основе эксперимента активную модель окружающего мира и использующего эту модель для активного восприятия мира и активного управления поведением. Это очень общее предварительное определение. В дальнейшем мы постараемся его конкретизировать и приблизиться к ответу на вопрос "Что такое мышление?" Я думаю, что ответ на этот вопрос для человека лежит частично в области знания и частично в области актуального (относительного) незнания. Правда есть вероятность и того, что полный ответ на этот вопрос лежит в области абсолютного незнания или даже в области непознаваемости.

Мышление это сложный многоплановый процесс, имеющий свою специфику в зависимости от решаемой мозгом задачи. Мы будем говорить о различных видах мышления, таких как:

перцептивное мышление, когнитивное мышление, практическое (поведенческое) мышление, абстрактное мышление, образное мышление, словесно-логическое или понятийное мышление, репродуктивное простое мышление, репродуктивное творческое мышление, созидательное творческое мышление.

## **Лекция 2. О работах в области моделирования мышления**

Ключевые слова: *кибернетика, бионика, искусственный интеллект, представление знаний, принятие решений, модель среды, эвристическое программирование, экспертные системы, роботы, распознавание образов, нейронные сети*

В этой лекции я попробую сделать краткий обзор направлений работ, имеющих отношение к моделированию мышления, дать общие оценки тому, что сделано и перспективам дальнейшего продвижения в направлении понимания и моделирования мышления.

### **Кибернетика**

Работы, направлявшиеся на понимание и автоматизацию (имитацию) мышления, велись под разными флагами. Первым и самым большим из этих флагов был кибернетика – наука об управлении в технике, природе и обществе. Кибернетика не стала наукой. Слово кибернетика сейчас чаще всего используется как родовое понятие для разных научных направлений, объединяемых тем, что они занимаются получением, обработкой, передачей, хранением и использованием информации. Содержательная часть кибернетики, предложенной Норбертом Винером в 1948г., состоит в попытке обобщении понятия управления и в провозглашении единства (схожести) принципов и процессов управления в технике, живой природе и обществе. Утверждалось, что в живой природе реализуются те же принципы управления, что и в технике, а именно, управление по отклонению на основе отрицательных обратных связей и управление по возмущению - схема стимул-реакция (т.е. рефлекс).

Эти представления прекрасно сочетаются со сформулированным русским физиологом И.П.Павловым общим принципом поведения живых организмов, звучащем как "уравновешивание со средой". Эти же представления легли в основу концепции гомеостаза и гомеостата (Эшби), а также в основу многочисленных как теоретических, так и технических моделей поведения и мышления, строящихся по схеме стимул - реакция. Об этом еще будем подробно говорить в дальнейшем.

Провозглашение общих принципов кибернетики имело как положительное, так и отрицательное значение. Положительное значение состояло в том, что кибернетика подтолкнула ученых, и в первую очередь представителей точных наук, к исследованию и моделированию информационных процессов, связанных с поведением, восприятием и мышлением. Математики и инженеры стали интересоваться физиологией и психологией. Начались попытки проведения комплексных междисциплинарных исследований.

Отрицательное значение общих принципов кибернетики состояло в том, что на этой основе часто упрощались (и даже примитивизировались) представления о живом и в особенности представления о поведении животных. Подчеркивалось общность и отодвигалось на задний план принципиальное различие между живым и неживым. Излишнее внимание уделялось (а часто уделяется и сейчас) гомеостазу, обратным связям, схеме стимул-реакция, проблеме "уравновешивания со средой". В то же время недостаточное внимание уделялось аспектам активности и целенаправленности. (Об этом мы подробно будем говорить в дальнейшем).

Кроме того, на первых порах с энтузиазмом обсуждался вопрос о сходстве между устройством мозга и устройством вычислительной машины. Появлялась масса статей и книг, таких как, например, "Мозг как вычислительная машина" (Джордж), "Конструкция мозга" (Эшби), "Алгоритмы разума" (Амосов).

В начале работ по кибернетике часто в разных аспектах проводились неправомерные прямые параллели между мозгом и вычислительной машиной. Так очень часто, сравнивая мозг и вычислительную машину, в качестве важного различия говорили, а иногда и сейчас говорят, о том, что ЭВМ - это последовательный вычислитель, а мозг - огромный, включающий 140 млрд. нейронов, параллельный. Однако, уже давно совершенно очевидна неправильность такого поверхностного сравнения. Говорить о том, что мозг это мощный вычислитель и сопоставлять операции, которые выполняет человеческий мозг, с операциями вычислительной машины нельзя. Это совершенно разные операции.

Человек, несмотря на якобы параллельную организацию своего "вычислителя", не может делать 100 тысяч сложений в секунду. Он, как правило, и одного сложения за секунду сделать не может. При этом, человеку удастся решать некоторые задачи, которые машина при всей своей вычислительной мощности решать не может, либо решает долго, на основе перебора. И дело здесь не в том, что машина - это последовательный вычислитель, а мозг - параллельное "устройство". Просто мозг и вычислительная машина решают свои задачи совсем по-разному. Тем не менее, до сих пор иногда возникают работы, в которых говорится (как об открытии) о том, что мозг и вычислительная машина работают совершенно по-разному, хотя в этом уже давно можно не сомневаться.

В дальнейшем мы постараемся более или менее подробно обосновать, что говоря о работе мозга, не следует использовать термины "алгоритм" и "программа", поскольку то, что управляет работой мозга в процессе мышления, ни алгоритмом в строгом смысле, ни программой, по-видимому, не является. Не является то, что происходит в мозге и вычислениями.

## **Бионика**

Так называемая интеллектуализация машин велась под разными флагами. Один из них назывался бионика. Бионика инициировалась кибернетикой, и общая цель бионики формулировалась как перенос в технику "изобретений" природы. Под этим флагом открывались многочисленные лаборатории, часто объединявшие инженеров и физиологов. Например, многочисленные лаборатории в различных организациях Министерства радиопромышленности, НИИ Нейрокибернетики РГУ, студенческие группы в РГУ, которым в равном объеме читались физиология и математика, группы с бионическим уклоном на физфаке МГУ и т.п.

Одной из наиболее важных задач бионики считалось использование в вычислительной технике знаний из области нейрофизиологии мозга. Однако достаточно быстро выяснилось, что переносить из нейрофизиологии в вычислительную технику нечего, при этом не только потому, что физиология не располагает достаточной конструктивной и целостной информацией о работе мозга, но и попросту потому, что эта информация вычислительной технике не нужна. Современная вычислительная машина на мозг не похожа. Устройство и принципы работы современной вычислительной машины не имеют ничего общего с устройством и принципами работы мозга.

Взаимодействие между физиологами и инженерами не получилось. Совместные междисциплинарные исследования в основном остались лозунгом.

Естественно, что бионический бум 60х годов, когда основной и единственной целью провозглашалось создание технических устройств или практически используемых программ, окончился полным крахом. Правда были и исключения не связанные с мышлением. Например, некоторые результаты исследований эхолокации летучих мышей использовались в радиолокации.

В целом, междисциплинарные исследования принципов работы мозга, которые велись под общим флагом – кибернетика или под частным флагом – бионика, принесли свою пользу хотя бы потому, что было привлечено внимание инженеров и математиков к проблемам, которые до этого считались исключительно делом психологов и физиологов. Это привело к многочисленным попыткам применения к описанию мозга и мышления как формальных аналитических средств, так и средств программного или технического моделирования, что, безусловно, принесло пользу для общего понимания проблем.

## Алгоритмический искусственный интеллект

Еще одним флагом, под которым велись и ведутся работы в рассматриваемой области, был и до сих пор остается "Искусственный интеллект".

Термин "Искусственный интеллект" (ИИ) используется в двух разных смыслах. Во-первых, ИИ может отождествляться с искусственным (машинным) мышлением. В этом случае вопросы: возможен ли искусственный интеллект? и может ли машина мыслить? – это один и тот же вопрос. Во-вторых, ИИ это название научного направления.

В Советском Союзе направление "Искусственный интеллект" пришло на смену кибернетическому буму первой половины 60-х годов.

Сначала оптимистам казалось, что вот-вот произойдет революция и вычислительная машина начнет думать по-человечески. Однако ничего подобного не произошло. Стало ясно, что полноценного искусственного мышления быстро построить нельзя. Поэтому стали говорить, что нужно заниматься не научными спекуляциями и не исследованиями с непонятной перспективой, а искусственным интеллектом - т.е. просто машинным решением трудно формализуемых задач - таких задач, которые человек решает, а машина нет. Таким образом, первоначально ИИ не претендовал на прямое моделирование мышления человека.

И все-таки, с самого начала явно или неявно предполагалось, что эти работы позволят сформулировать обобщения и выработать специфические методы ИИ, ведущие, в конечном счете, к машинному мышлению. Представители возникшего направления полагали (и полагают), что к пониманию и моделированию мышления полезно идти от специфики задач к методам их решения, вводя "интеллект" как механизм, необходимый для решения.

Какие задачи традиционно относят к области ИИ? Оказалось, что таких задач много. Это понимание машиной естественного языка, т.е. вопрос-ответные системы и доступ к базам данных на естественном языке, перевод с одного языка на другой, распознавание образов, анализ изображений 3-х мерных сцен, логистические системы представления знаний и логический вывод, эвристическое программирование, доказательство теорем, принятие решений, игры, базы данных, базы знаний, роботы, экспертные системы и др.

Перечисленные задачи, по-видимому, можно действительно считать специфическими трудно формализуемыми "человеческими" задачами, и возникает проблема выявления в этих задачах общности, определяющей необходимость мышления.

На определенную общность в первую очередь ориентировались направления: эвристическое программирование, построение систем представления знаний и разработка методов поиска решений.

Эвристическое программирование основано на психологическом анализе и попытках обобщения и формализации методов, используемых человеком при решении конкретных задач. Серьезных обобщений получить не удалось. Универсальный решатель проблем (GPS Ньюела, Шоу и Саймона) не получился. Часто не удавалось формализовать методы решения и одной единственной задачи. Например, не удалось формализовать мышление шахматиста.

Системы представления знаний. Разработанные главным образом для решения задач машинной лингвистики, иерархические системы представления и использования знаний, обычно строятся на основе математической логики, фреймов, семантических сетей или графов. Все такие системы можно считать шагом в нужном направлении, т.е. в переходе от данных к знаниям, хотя они и являются не активными моделями, а пассивными описаниями. На разнице между моделью и описанием мы еще будем останавливаться в дальнейшем.

Говорят, что любая "интеллектуальная" система, например, система управления, перевода или восприятия должна уметь строить и использовать семантическую модель мира. Это, безусловно, верно, как верно и то, что достаточно общего решения эта проблема пока еще не имеет.

Коротко о проблеме поиска решений.

Часто поиск решения сводится к общей схеме: (So) -> {A} -> (Sk)

Начальная ситуация - "So" может иметь интерпретацию от простого сигнального стимула до сложной ситуации. Соответственно идентификация "So" в среде решаемой задачи может быть разной - от очень простой до очень сложной, требующей специальных методов распознавания.

"Sk" может иметь интерпретацию от конкретной целевой ситуации до ситуации, определяемой как обобщение. В последнем случае необходимо формирование конкретных целей (дедукция) и выбор между ними.

"A" может иметь интерпретацию от простой реакции до сложной последовательности действий, ведущих к переходу в промежуточные ситуации. От однозначной реакции типа рефлекса (ситуационное управление) до выбора из множества возможностей.

В общем случае большинство задач может представляться как поиск пути на графе путем перебора. Перебор сокращается, если возможны точные или эвристические оценки промежуточных ситуаций в отношении приближения к цели. В этом случае в одноэкстремальных задачах возможно построение градиентных алгоритмов управления. В многоэкстремальных задачах полный градиентный алгоритм невозможен, но можно на основе какой-то информации о задаче и эвристик пытаться упорядочивать перебор или сокращать пространство поиска.

Если нет никакой обобщающей информации о многоэкстремальной задаче, то ее решение может гарантировать только полный перебор. Таким образом, анализ многих работ по ИИ приводит к выводу, что одной



из наиболее важных задач мышления можно считать задачу построения иерархической модели среды и ее использования для сведения многоэкстремальных задач к одноэкстремальным. (Более подробно об этом будет говориться в следующих лекциях).

Среди конкретных задач наиболее актуальными были машинная лингвистика и машинное зрительное восприятие. Эти задачи с определенными ограничениями доводятся до практических результатов, но их решение еще очень далеко до интеллектуального, т.е. человеческого. В лингвистических системах проблемой является полноценный семантический анализ. В системах машинного зрительного восприятия принципиальная проблема это управление с понятийного уровня, т.е. “восприятие с пониманием”.

Определенные комплексные исследования велись по направлениям ‘экспертные системы’ и ‘интегральные роботы’. В экспертных системах центральными моментами были базы данных, информационный поиск и общение человека с базами данных, распознавание ситуаций по признакам и связывание ситуаций с рекомендациями и, в некоторых случаях, с управляющими действиями. Это связывание может происходить с использованием знаний экспертов как по схеме стимул-реакция, так и на основе формальных методов поиска решений – главным образом с использованием исчисления предикатов и его расширений. Серьезных обобщений, полезных для общего понимания проблемы мышления на этом направлении получено не было.

При создании роботов главные вопросы это зрительное восприятие трехмерных сцен и управление движением тележки или манипулятора. Роботы используются в программируемых технологических операциях, например таких, как покраска автомобиля, либо как игрушки. Восприятие среды в этих системах упрощено, а принятие решений не идет дальше априорно задаваемых формальных схем.

Можно считать, что конструктивных обобщений методов ИИ, применимых к разным задачам и принципиально отличающихся от традиционных способов решения создано не было. В частности, сейчас машины прекрасно играют в шахматы. Практический уровень игры машины сопоставим с уровнем игры чемпиона мира. Машина решает эту задачу за счет мощных вычислительных возможностей в основном перебором огромного числа вариантов и использованием как точных, так и эвристических правил оценки сравниваемых позиций.

Если задача машинной игры в шахматы поддалась “силовому” решению, то успех в решении многих других задач практически отсутствует. Главным инструментом в системах ИИ была и остается логика предикатов и ее расширения (вероятностная логика, логика с часами и т.п.). Принципиальных прорывов не было и нет. Поэтому нет и серьезных практических успехов.

Так не был реализован японский проект ЭВМ 5-го поколения, направлявшийся, главным образом, на создание естественных зрительных и речевых форм общения между человеком и компьютером и дружелюбного “интеллектуального” интерфейса. Ставилась задача радикального упрощения программирования. Приводились кривые роста населения земного шара и роста числа программистов. Эти кривые при их экстраполяции пересекаются. Выяснилось, что для реализации этого проекта недостаточно построить “глаз” и “ухо”. Оказалось, что нужно еще и понять, как работает мозг. (Впрочем, это было ясно многим с самого начала).

По тем же причинам отсутствует качественное продвижение в решении таких задач, как анализ 3-х мерных сцен и перевод с одного языка на другой. Точно так же и роботы, как и 30 лет назад, остаются либо устройствами для выполнения сложных стандартных технологических операций, либо все более и более сложными игрушками, не идущими, тем не менее, в своем обучении и поведении дальше условного рефлекса, динамического стереотипа, т.е. фиксированной последовательности действий, или принятия решений по простому набору правил.

## **Распознавание образов**

К исследованиям в области ИИ примыкают работы по распознаванию образов. Первоначально ставившаяся общая задача восприятия сложной среды постепенно выродилась в упрощенную задачу классификации. Традиционные программы или устройства распознавания образов это чаще всего пассивные признаковые системы классификации объектов, рассматриваемых по отдельности. Серьезных теоретических успехов, имеющих значение для понимания механизмов мышления, на этом направлении не получено. Эти системы не обладают такими свойствами живого восприятия как целостность, целенаправленность и “распознавание с пониманием”, основывающееся на использовании модели среды.

В дальнейшем мы еще будем говорить о перцептивном мышлении (мышлении при восприятии), которое включает не только узнавание отдельных объектов, но и целостное восприятие окружающей среды. При этом должно происходить управление процессом восприятия с понятийного уровня. Восприятие должно основываться на предвидении, формирующем “акцептор восприятия” и гипотезы восприятия. Должен использоваться контекст и максимально полная целостная семантическая модель проблемной среды.

Классическое распознавание образов прямого отношения к перцептивному мышлению не имеет. “Восприятие с пониманием”, о котором мы еще будем говорить в дальнейшем, можно считать упрощенной, но в то же время важной частью перцептивного мышления.

Для алгоритмического ИИ характерно отсутствие заметных продвижений в решении проблемы понимания и моделирования мышления. Это, связано с тем, что отсутствует общая теория мышления. В то же время, целостное мышление не может возникнуть из суммы решенных отдельных ‘интеллектуальных’ задач. Кроме того,

недостаточно изучать методы решения таких задач. Необходимо также понять работу активных нейронных механизмов решения.

## **Формальные нейронные сети**

Моделирование нейронных сетей проводилось, но эти работы велись и ведутся независимо от алгоритмического ИИ. Не ориентировались эти работы и на решение обычных ‘интеллектуальных’ задач. Единственной практической задачей, успешно решаемой нейронными сетями, было и остается распознавание образов. Можно выделить три разные вида формальных нейронных сетей – логические сети, растровые распознаватели, признаковые распознаватели.

Серьезных теоретических успехов, имеющих значение для понимания механизмов мышления у направления, называемого нейронаукой, нет.

Это в первую очередь относится как к логическим сетям, так и к окутанном чуть ли не мистическим туманом искусственным или формальным растровым и признаковым распознающим нейронным сетям. Часто их рассматривают даже как какой-то аналог мозга. В последнее время в текстах просто пишут “нейронные сети”, опуская прилагательные искусственный или формальный. На самом деле, признаковая распознающая нейронная сеть (ПРНС) – это просто пассивный признаковый классификатор, строящий разделяющие гиперплоскости в пространстве признаков.

Используемый в этих сетях формальный нейрон – это сумматор с пороговым элементом, подсчитывающий сумму произведений значений признаков на некоторые коэффициенты, являющиеся ни чем иным, как коэффициентами уравнения разделяющей гиперплоскости в пространстве признаков. Если сумма меньше порога, то вектор признаков находится по одну сторону от разделяющей плоскости, если больше – по другую. Вот и все. Кроме классификации по признакам никаких чудес. Сеть из формальных нейронов может также аппроксимировать плоскостями нелинейные разделяющие поверхности и объединять по результату несвязанные области пространства признаков. Это и делается в многослойных сетях.

Во всех случаях ПРНС – это пассивный признаковый классификатор, выделяющий области в фиксированном пространстве признаков. Никаких других задач ПРНС решать не может, причем задачу распознавания ПРНС решает не лучше обычных признаковых распознавателей, использующих аналитические методы.

Еще меньшими возможностями в плане распознавания образов обладают идущие от модели Хебба и перцептрона Розенблатта растровые нейронные распознаватели. Хотя такие распознаватели, в которых образуются ансамбли одновременно активируемых входной информацией нейронов, могут в некоторых случаях обладать ассоциациями по сходству и, в принципе, ассоциациями по смежности во времени. Эти аспекты работы растровых (ансамблевых) нейронных сетей перекликаются с некоторыми представлениями о работе мозга.

К понятию формальная нейронная сеть часто добавляется понятие нейрокомпьютер. Этот термин не очень удачен. Системы, к которым он обычно применяется, компьютерами не являются. Нейрокомпьютеры это обычно сложные обучающиеся формальные нейронные сети, которые могут решать задачи, отличающиеся от простого распознавания отдельных объектов. Вход у этих систем может быть сложным многопараметрическим. Также сложным и многопараметрическим может быть выход. Однако эти системы остаются по сути пассивными распознавателями, связывающими вход с выходом и работающими по схеме стимул-реакция.

Оценивая успехи так называемой нейронауки, можно, в отличие от распространенного мнения, сказать, что создание пассивных признаковых и растровых распознавателей, а также строящихся на их основе устройств или программ, называемых нейрокомпьютерами, не является заметным шагом в понимании и моделировании нейронных механизмов мозга.

Более подробно формальные пассивные нейронные сети, а также гипотезы об активных нейронных сетях будут рассмотрены в дальнейшем.

## **Синергетика**

Работам в области ИИ в последнее время часто противопоставляется синергетика. Развиваемый в этом новом направлении подход иногда называют теорией всего и говорят о возможности применения аппарата синергетики к описанию любых систем, в том числе живых и даже мозга. Синергетика использует аппарат нелинейной динамики для описания изменений в мультистабильных системах, теряющих устойчивость при внешних воздействиях и переходящих в результате происходящих изменений в новое стабильное состояние. Переход в новое состояние определяется градиентом действующих сил, а также сочетанием внешних и внутренних случайных факторов. Такой процесс характерен для теряющих устойчивость пассивных неживых систем. Новое стабильное состояние может быть как равновесным, так и неравновесным. В последнем случае это состояние должно поддерживаться постоянным внешним воздействием, например притоком энергии. Такое образование стабильных неравновесных систем можно назвать пассивной самоорганизацией. В дальнейшем мы будем подробно говорить о том, что активные живые системы постоянно находятся в состоянии неравновесия. Теряя устойчивость и начиная движение к состоянию равновесия, они за счет внутренней и внешней работы возвращаются в состояние неравновесия. Переход от неравновесия к равновесию в каких-то фазах существования и развития живого также существует.

Однако представляется, что главное, как в поведении живых организмов, так и в работе мозга это активное целенаправленное поддержание неравновесия. Этот процесс можно назвать активной самоорганизацией.

В дальнейшем мы будем говорить о том, что две главные решаемые мозгом задачи это построение иерархической активной модели среды и использование этой модели для целостного целенаправленного восприятия, а также для быстрого решения многопереборных многоэкстремальных задач поведения путем сведения этих задач к малопереборным одноэкстремальным. Все это в целом аппаратом нелинейной динамики не описывается. В связи с этим претензии синергетики на роль теории всего не выглядят обоснованными.

Таким образом, ни в направлении ИИ ни в смежных областях значительных обобщающих результатов, ведущих к пониманию мышления нет. Более того, становится все более ясно, что пассивные алгоритмические системы к пониманию и моделированию мышления прямо не ведут. Сумма решенных “интеллектуальных” задач мышления не образует. Часто считается плохим тоном говорить о моделировании человеческого мышления, априори относя эти разговоры к спекуляциям.

Не случайно в названиях некоторых популярных книг и статей появляются такие слова, как: “голый король”, “король умер”, “новый ум короля”.

### **Виталика**

Все сказанное естественно приводит к следующему выводу. К пониманию мышления, по-видимому, нужно идти не от алгоритмического ИИ, т.е. не от решения трудных задач, связанных с необходимостью обработки сложной информации, а от понимания того, чем живое отличается от неживого. Можно надеяться на то, что эти отличия не только лежат в основе особой формы существования материи, но и в основе мышления. Можно надеяться на то, что эти отличия материальны, моделируемы и находятся в области относительного незнания. Поиск, точное определение и моделирование этих отличий можно считать отдельным научным направлением. Это направление можно назвать, например, “виталика” (от латинского *vitalis*). При этом не нужно смешивать это направление с идеалистическим витализмом, понимая цель виталики как стремление найти материальную и точно определяемую основу для таких сущностей как ‘жизненная сила’, ‘одушевленность’, ‘энтелехия’. Пытаться продвинуться в этом направлении мы будем, рассматривая поведение, восприятие и работу нейронных механизмов мышления. Возможно, что проявляющаяся в этих задачах важная суть отличий живого от неживого определяется принципом устойчивого неравновесия Э.С.Бауера. Об этом мы еще будем говорить в следующих лекциях.

### **Предварительное определение мышления**

Мозг возник и эволюционно развивался для обеспечения существования животных, т.е. для выживания. Возможно простое функциональное определение мышления, основывающееся на представлении о том, для чего нужно мышление (человеку или животному). Такое предварительное определение было введено в предыдущей лекции. Повторим его с небольшими добавлениями.

Мышление это **активный процесс в живом мозге** направленный на:

1. построение в мозге активной иерархической модели среды, необходимой и достаточной для восприятия среды и управления **активным целенаправленным поведением в многоэкстремальной среде**;
2. реализацию процесса восприятия среды;
3. реализацию процесса управления поведением в многоэкстремальной среде;
4. реализацию процесса обучения;
5. решение неалгоритмических (творческих) задач.

Обратите внимание на слова **активный процесс** в живом мозге. Этими словами подчеркивается, что для понимания мышления важно понять не только то, что делает мозг, но и то, как он это делает. Принципиально также подчеркивание многоэкстремальности задач поведения.

Это очень общее предварительное определение в дальнейшем будет использоваться как исходная рамка, которая будет заполняться конкретным и уточняемым в значительной степени гипотетическим содержанием.

И как уже было сказано, мы постараемся двигаться к пониманию мышления от задач восприятия и управления поведением, также как и от понимания принципов работы нейронных механизмов, рассматривая эти задачи с единых позиций отличий живого от неживого.

## **Лекция 3. Существующие предпосылки к инженерному пониманию поведения и мышления**

Итак, мы хотим понять, что такое мышление, оценить возможность его моделирования и найти подходы к моделированию. От чего можно отталкиваться?

В предыдущей лекции я в основном критиковал кибернетику, бионику и направление исследований, называемое искусственный интеллект (ИИ). В науке принято считать плохим тоном отрицание всего, что было сделано ранее. Хорошим тоном считается говорить, что предшествующие результаты были полезны и послужили отправной точкой для дальнейшего развития. В нашем случае так и есть.

## **Общая оценка результатов работ по ИИ и в смежных областях**

Кибернетика и бионика стимулировали проведение междисциплинарных исследований и привлекли внимание представителей точных наук к проблемам нейрофизиологии и психологии. Были сделаны первые попытки моделирования работы мозга. Эти попытки были полезны, хотя чаще всего они ограничивались моделированием гомеостаза, условнорефлекторного поведения (схема стимул-реакция) и очень упрощенных формальных нейронных сетей.

Много полезного для понимания мышления было получено и в направлении алгоритмического ИИ. Работы по системам представления знаний привели к выводу, что мышление должно основываться на создании семантической модели проблемной среды и что эта модель должна быть иерархически построена как по линии частное-общее, так и по линии часть-целое. Разработки алгоритмов принятия решения для разных “интеллектуальных” задач показали, что эти задачи чаще всего представляются графом состояний или ситуаций и поиск решений основывается на переборе. Предложены различные в основном эвристические методы сокращения перебора. Эти методы хорошо работают в одноэкстремальных задачах и значительно хуже в многоэкстремальных. Попытки разработки алгоритмов для восприятия сложной среды, так же как и алгоритмов для понимания естественного языка, определили необходимость использования семантической модели проблемной среды. Эта модель должна быть не пассивным описанием, а активной системой, управляющей процессом восприятия и делающей его целенаправленным. Правда, последний вывод делается на основе анализа результатов и напрямую из работ по ИИ не вытекает.

Примыкающие к направлению ИИ формальные нейронные сети можно считать моделями нейронных механизмов мозга, но лишь в очень отдаленном приближении.

Перечисленные и другие полученные в ИИ полезные результаты не объединены в целостную систему и к моделированию мышления прямо не ведут. Более того, сейчас считается плохим тоном говорить о проведении в рамках ИИ работ по моделированию мышления. Априори считается, что такие утверждения с очень большой вероятностью могут быть спекуляцией.

Итак, исследования, проводящиеся в направлении ИИ подталкивают к выводу, что для понимания и моделирования мышления недостаточно ограничиваться разработкой алгоритмов и программ решения трудноформализуемых задач. Нужно понять в чем состоит специфика живого как на уровне восприятия и внешнего поведения, так и на уровне работы нейронных механизмов мозга. Таким образом, необходимо исследование принципов поведения, восприятия и мышления. Это направление работ можно назвать, например, “Искусственный разум” (ИР), чтобы не смешивать его с формальным алгоритмическим направлением “Искусственный интеллект” (ИИ), или еще лучше, как отмечалось в предыдущей лекции, назвать это направление - “Виталика”.

Теперь поговорим о предпосылках к конструктивному пониманию и определению перспектив моделирования поведения и мышления, которые дает биология или физиология.

## **Что такое жизнь?**

Мы уже договорились о том, что будем считать мышление активным процессом в живом мозге. А что такое жизнь? Чем живое отличается от неживого? И что такое активность?

Существуют многочисленные определения понятия жизнь. Приведем классическое определение Ф.Энгельса – “Жизнь есть способ существования белковых тел, и этот способ существования заключается по своему существу в постоянном обновлении их химических составных частей путем питания и выделения.” Типичное современное определение: “Жизнь – высшая по сравнению с физической и химической форма существования материи, закономерно возникающая при определенных условиях в процессе ее развития. Живые объекты отличаются от неживых объектов обменом веществ – непрерывным условием жизни, способностью к размножению, росту, активной регуляции своего состава и функций, к различным формам движения, раздражимостью, приспособляемостью к среде и т.д.” (А.А.Малиновский, энциклопедический словарь)].

Определение жизни, данное Ф.Энгельсом, значительно лучше чем определение Малиновского, поскольку в нем делается попытка выделить общий принцип (способ существования, заключающийся в том ...). Подобные определения сливаются с объяснением. Определение Энгельса, конечно, не является исчерпывающим, а дается на некотором уровне понимания проблемы. Остается много вопросов и главный – Зачем живому обновлять свои составные части?

Определение А.Малиновского, так же как и многие другие современные определения, хуже не только по своей сути, хотя это конечно очень важно, но и потому, что оно по своему характеру является индуктивным, то есть строится в основном на перечислении свойств или признаков определяемого объекта и не дает читателю информации, которая была бы полезна для какого-то обобщения. Индуктивные определения неизбежны на эмпирическом этапе развития любой науки.

Многие определения ставят на первое место функциональную или поведенческую сторону существования живого. Я уже отмечал, что характерными и типичными являются представления о подвижном равновесии между организмом и средой, схема стимул – реакция, “уравновешивание со средой” по И.П.Павлову, принципы гомеостаза Кеннона и гомеостатического регулирования Эшби и др. Все эти подходы можно назвать “кибернетическими”. Они основываются на принципе поддержания устойчивости при внешних воздействиях. Например, типичным является утверждение А.А.Ляпунова, который в своей статье “Кибернетический подход к теоретической биологии” одной из наиболее важных характеристик живого предложил считать наличие повышенной относительной устойчивости за счет сохраняющих реакций на внешние воздействия со стороны среды.

### **Принцип устойчивого неравновесия Э.С.Бауэра**

Теперь остановимся на результатах, полученных Э.С.Бауэром, книга которого “Теоретическая биология” вышла еще в 1935г. По Бауэру в основе жизни лежит принцип устойчивого неравновесия: “все и только живые системы никогда не бывают в равновесии и исполняют за счет своей свободной энергии постоянно работу против неравновесия, требуемого законами физики и химии при существующих внешних условиях”.

Этот принцип, по мнению Э.С.Бауэра, является всеобщим законом биологии, определяющим отличие живой материи от мертвой. В соответствии с этим принципом выделение и использование свободной энергии в живой системе сопровождается постоянным разрушением и восстановлением структуры живой ткани, обладающей свободной энергией и находящейся в неравновесном состоянии. Для этого процесса необходимо поступление в систему извне определенных, в первую очередь, органических веществ. Таким образом, направленное специфическое взаимодействие со средой в целевых ситуациях поведения необходимо для существования живой материи. Поэтому живая система должна быть открыта и постоянно активна. Поведение – непрерывная задача.

### **К определению жизни**

Я не буду пытаться дать законченное и точное общее определение жизни. Отмечу лишь, что это определение, повидимому, должно опираться на принцип устойчивого неравновесия и активный динамический способ существования в необходимой для существования среде. Устойчивое неравновесие и как следствие активность и целенаправленность являются важнейшими необходимыми, но, может быть, недостаточными признаками живого. Эти принципы, повидимому, могут быть реализованы и в искусственных системах. То, что действительно необходимо для жизни и не реализуется ни в каких искусственных системах – это фотосинтез и биосинтез живых органических структур. Для получения полного определения жизни необходимо также привязать устойчивое неравновесие к конкретному биологическому субстрату. Ф.Энгельс считал таким субстратом живой белок. Э.С.Бауэр говорил не о живом белке, а об особом структурном состоянии живой материи. В современной биологии по этому поводу имеются и другие мнения. Кроме того, в определении жизни наряду с устойчивым неравновесием, фотосинтезом и биосинтезом должны входить такие основные необходимые свойства, как создание модифицированной копии (размножение).

На некотором уровне понимания и объяснения живое можно охарактеризовать еще и следующим образом. В отличие от неживого живое имеет цель – сохранение во времени своей качественной определенности и средства для активного достижения этой цели...

Итак, можно сделать вывод, что непрерывное поддержание неравновесия нужно всему живому функционально и на клеточном и на поведенческом уровне. Следствие этого – целенаправленность и активность. На этом основаны принципы работы любых живых механизмов. Слова жизнь и активность хотя и не синонимы, но они относятся к тесно и неразрывно связанным между собой понятиям. Кроме того, представляется, что в определении жизни, во всяком случае, в определении ее эволюции, полезно включить также с необходимостью возникающее в живых организмах свойство синергии (взаимосодействия). На этом я еще буду останавливаться в дальнейшем. Кроме того, я постараюсь показать, что активность и синергия являются не только необходимыми свойствами живого организма в целом, но и необходимыми свойствами нейронных механизмов живого мозга.

Большой интерес в плане основной интересующей нас общей темы, т.е. мышления, представляют процессы возникновения и эволюции живых организмов. Мы не будем здесь подробно рассматривать очень сложный в своих деталях и не до конца еще определенный процесс возникновения жизни. Приблизительно описывая самый общий уровень процесса, можно сказать, что зарождение жизни произошло при случайном объединении элементов в органическую неравновесную систему, обладающую свободной энергией, свойством самоподдержания неравновесия за счет специфического взаимодействия со средой и свойством запоминания, размножения и передачи своей копии потомству. Взаимодействие со средой ограничивалось на первом этапе приемом из среды углерода, приемом солнечного света и фотосинтезом неравновесных органических структур. Реализуемый в

растениях фотосинтез остается и сейчас необходимым и единственным созидательным фактором существования жизни на земле. Без фотосинтеза неравновесных, обладающих свободной энергией органических соединений жизнь на земле существовать не может. Помимо фотосинтеза существует и биосинтез белка. Для биосинтеза нужно поступающее из среды питание, состоящее из мертвых тканей других организмов. Биосинтез белка является необходимой функцией существования животных. Как для фотосинтеза и биосинтеза, так и для воспроизведения при размножении нужна информация о строении организма, т.е. память. Эта память реализуется соответственно в молекулах РНК и ДНК.

Таким образом, понятие живой организм шире, чем понятие животное. Живыми организмами являются и растения. Живое это то, что построено из органических соединений и по определению Э.Бауера “никогда не бывает в равновесии и исполняет за счет своей свободной энергии постоянно работу против равновесия”, поддерживая тем самым неравновесие и уровень свободной энергии. В этом плане, растение это такой же живой организм, как и животное. Более того, жизнь на земле создана и поддерживается растениями. Мы уже говорили о том, что жизнь возникла тогда, когда возникла первая клетка, реализующая фотосинтез обладающих свободной энергией неравновесных органических структур. Фотосинтезирующими были некоторые бактерии и растения. Необходимая для существования среда растений – это солнечный свет, находящийся в воздухе углерод, находящиеся в почве вода и неорганические элементы, в то время как необходимая для существования, то есть удовлетворения первичных потребностей и поддержания неравновесия, среда животных – это другие живые организмы, в том числе (обязательно) и растения. Таким образом, растения несут в себе созидательное начало – создают жизнь, а животные как созидательное, так и разрушительное – поддерживают жизнь за счет разрушения жизни. Растения без животных на земле существовать могут, животные без растений нет.

## Поведение

Вернемся к задаче поведения. По сравнению с поведением животных поведение растений обладает принципиальными особенностями, определяемыми отличием потребностей и среды. Как мы уже отметили, потребности растущих на земле растений это свет, воздух, тепло, вода, элементы почвы. Для удовлетворения таких потребностей у растений реализуется простое градиентное одноэкстремальное поведение. Это поведение сводится к захвату из среды необходимых элементов, направленному росту корней, стеблей и листьев, а также к элементарным движениям листьев и цветов (тропизмы растений). Для управления таким поведением мозг не нужен.

Животные могут перемещаться в пространстве. Это значительно расширяет их возможности по достижению необходимых целевых ситуаций взаимодействия со средой. Для перемещающихся организмов фотосинтез становится необязательным. Питание может осуществляться за счет других организмов, в том числе и за счет организмов, реализующих фотосинтез (растений). В то же время задача поведения значительно усложняется. Тропизмы (градиентные механизмы) перестают работать. Нужно управление для оптимизации поведения. Для целенаправленного перемещения в среде требуется дистанционное восприятие, требуется умозрительное моделирование процессов в среде (воображение). Для всего этого нужна модель среды. Кроме того модель нужна для такого представления среды, на котором стал бы вновь возможен градиентный одноэкстремальный способ управления поведением. Для всего этого и нужен мозг.

Каковы физиологические предпосылки к инженерному, т.е. конструктивному пониманию поведения и мышления? Законченной целостной теории работы мозга при поведении, восприятии или мышлении в целом в науке, называемой физиологией высшей нервной деятельности, нет. Нет законченной целостной теории и в психологии.

Как на физиологическом уровне изучать мышление непонятно. Изучать можно то, что наблюдается в эксперименте. В физиологическом эксперименте, во-первых, можно наблюдать электрическую активность отдельных нервных клеток и интегральную электрическую активность отделов головного мозга. О таких результатах мы еще будем говорить в дальнейшем. Эти результаты очень интересны но они, не образуют целостную теорию работы головного мозга в процессе мышления, и могут только помогать при построении такой теории.

Во-вторых, можно экспериментально изучать поведение и более или менее просто наблюдать рефлексы (преимущественно млекопитающих) и динамические стереотипы (преимущественно насекомых, птиц и рыб). Поэтому с самого начала очень многие работы посвящались изучению безусловных и условных рефлексов. Фундаментальное значение в отечественной физиологии имеют классические труды И.М.Сеченова и И.П.Павлова. Основная работа Сеченова – “Рефлексы головного мозга”. Изучение рефлексов много дало для понимания работы мозга, но теорий поведения и мышления на этой основе построить нельзя.

Мы уже говорили о том, что в кибернетике на первых порах значительное внимание уделялось моделированию поведения, ограничивающегося безусловными и условными рефлексами, а также динамическими стереотипами (фиксированными планами). Представления о поведении долгое время базировались на схеме стимул-реакция плюс гомеостаз (поддержание постоянства внутренней среды организма). До сих пор как в технических, так и в физиологических работах иногда (часто) вспоминают тезис Павлова “уравновешивание со средой” и рассматривают поведение в рамках схемы стимул-реакция.

Дальше мы остановимся на результатах исследования более сложного поведения. Эти исследования не привели к построению законченной теории работы мозга, однако есть много достаточно интересных результатов, которые можно считать важными предпосылками к построению целостной теории. Как часто говорят физиологи, в мозге строится модель среды для реализации целесообразного или “эффективного приспособительного” поведения. Что такое целесообразное и эффективное приспособительное? Целесообразный и эффективный вариант поведения это лучший по каким-то критериям вариант среди множества возможных. Каковы эти критерии? Как в физиологии высшей нервной деятельности трактуются целесообразность и эффективность поведения?

В связи с этими вопросами нужно остановиться на результатах некоторых классических отечественных работ в области нейрофизиологии, посвященных исследованию и описанию закономерностей, характерных для целенаправленного поведения живого организма в сложной среде, то есть поведения более сложного, чем просто реакции на внешние раздражители.

В общем виде представление о поведении как об активном процессе, опирающемся на опережающее моделирование, содержится еще в классических работах Сеченова и Павлова. Так Сеченов подчеркивал активную основу процесса восприятия, отмечая, что мы не слышим и видим, а слушаем и смотрим, а разработанная Павловым теория условнорефлекторной деятельности подразумевает наличие в нервной системе опережающего возбуждения модели того, что должно произойти. На это прямо указывают в опытах с выработанными условными рефлексами у собак эксперименты по исследованию химического состава слюны, который соответствует ожидаемому подкреплению.

### **Принцип доминанты А.А.Ухтомского**

В плане рассматриваемых ниже проблем особый интерес представляют работы А.А.Ухтомского, в которых сформулирован принцип доминанты. В соответствии с этим принципом доминанта поведения и соответствующий ей очаг возбуждения в нервной системе возникают как результат “конкурентной борьбы” многих доминант, отражающих существующие в данный момент первичные и вторичные потребности. “Победившая” главная доминанта фиксируется как стабильный очаг возбуждения. Остальные доминанты затормаживаются. Одновременно получают возбуждение все элементы нервной системы, которые как генетически, так и в соответствии с приобретенной и ситуативной информацией должны (или могут) участвовать в достижении поведенческой цели, соответствующей главной доминанте. При достижении цели доминанта ликвидируется (гасится очаг возбуждения) и возникает новая. В теории Ухтомского не конкретизируется вопрос о том, каковы критерии выбора фиксирующейся доминанты.

Основные аспекты адаптивного поведения в плане построения и использования модели среды показаны в физиологических исследованиях Анохина, Беритова и Бернштейна. Для всех этих работ характерно продвижение от Павловского “уравновешивания со средой” к направленному целесообразному и необходимому взаимодействию со средой, основанному в первую очередь на предвидении результатов взаимодействия.

### **Теория функциональной системы**

П.К.Анохин разработал понятие динамически складывающейся “функциональной системы” как “замкнутого физиологического образования с обратной афферентацией”, направленного на достижение конечного приспособительного эффекта. Составляющие функциональной системы это: “афферентный синтез, акцептор действия, формирование действия и обратная афферентация о его результатах”. (Афферентация это сигналы, поступающие в нервную систему от органов восприятия среды и рецепторов тела. Афферентация - это сигналы, идущие из нервной системы к периферии). Направленность целостного поведенческого акта определяется актуальной физиологической потребностью, которая с учетом текущей ситуации формирует в нервной системе модель состояния взаимодействия со средой, при котором потребность удовлетворяется. Достижение конечного приспособительного эффекта фиксируется получением “санкционирующей афферентации”, совпадающей с моделью (акцептор действия). Акцептор действия, т.е. предвидение, возникает при любом действии. При нарушении акцептора действия (нарушении предвидения) поведенческий акт прерывается и возникает ориентировочно-исследовательская реакция, направленная на уточнение восприятия, корректировку предвидения или корректировку модели среды.

Важный момент теории П.К.Анохина состоит в том, что функциональная система и акцептор действия работают всегда и непрерывно. Во –вторых, важно, также, что центр внимания смещается от рефлекторного поведения в сторону активного поведения, направленного на достижение целевого состояния взаимодействия со средой. И, наконец, важным аспектом теории функциональной системы является сформулированное в ней свойство динамической синергии, т.е. оперативно возникающего взаимодействия между функциональными подсистемами организма в любом поведенческом акте. Все это было сформулировано еще в 60-е годы прошлого века.

Близкие по смыслу идеи содержатся в работах Н.А.Бернштейна, который выделяет две линии изучения поведенческого акта – физиологию регуляций и физиологию активности. Регуляция основывается на “вероятностном прогнозировании по воспринимаемой текущей ситуации”. Кроме того, организм “не просто

взаимодействует с окружающим миром, но и активно воздействует на этот мир, стремясь изменить его в потребном себе отношении”. Активность поведения направлена на достижение “закодированного в нервной системе отображения или своего рода модели потребного будущего”. Выработка программы действия состоит в установлении связи “между наличной ситуацией и тем, во что она должна быть превращена”.

Вопросы, связанные с построением в нервной системе модели среды и использованием этой модели в процессе поведения нашли отражение в работах И.С.Беритова и многих других авторов. К сожалению, в большинстве этих работ, в том числе в теории А.А.Ухтомского о принципе доминанты, в теории функциональной системы П.К.Анохина и в теории поведения Н.А.Бернштейна нет четких и строгих представлений о том, какая потребность становится актуальной, как выбирается целевое состояние взаимодействия со средой и вариант достижения цели, что такое полезность и “приспособительный эффект”. Поэтому обо всех таких теориях нельзя говорить как о теориях поведения. В частности, функциональная система Анохина это не теория поведения, а лишь теория отдельного поведенческого акта, поскольку в этой теории на центральном месте находится нераскрываемый “черный ящик” – “блок афферентного синтеза”.

### **Эксперименты Н.В.Асмояна и Г.А.Голицына,**

Определенное исключение представляет работа Н.В.Асмояна и Г.А.Голицына, посвященная экспериментальному исследованию принципов принятия поведенческого решения животным в ситуации выбора. На основании экспериментальных результатов авторы смогли установить количественное локальное правило принятия решения и пришли к естественному выводу о том, что поведение животных подчиняется некоторому принципу оптимальности. Однако, убедительного объяснения “приспособительного” смысла этого принципа в работе Н.В.Асмояна и Г.А.Голицына получено не было. Мы еще поговорим об этих очень интересных результатах в следующей лекции.

Роль эмоций.

Многие авторы, например П.К.Анохин, Г.А.Голицын, П.В.Симонов и др. выделяют важную роль эмоций в поведенческом акте. Так Анохин пишет, что эмоции играют роль “пеленга”, определяя направленность поведения, Голицын говорит об эмоциональном механизме сравнения и выбора вариантов поведения. (Это очень важно и не вполне очевидно, ведь еще великий Дарвин писал, что эмоции это просто атавистические свойства, имея, по-видимому, ввиду лишь мощные эмоциональные состояния, такие как горе, радость, страх и т.п.)

#### **Общая оценка результатов по изучению поведения**

Результатов, относящихся к исследованию поведения, много и они весьма разнообразны. Мне представляется, что обобщая эти результаты можно выделить следующие наиболее важные моменты:

- построение в мозге модели среды;
- непрерывное предвидение на основе модели;
- активность при восприятии среды;
- активность поведения, направленная на достижение целевых ситуаций;
- доминанта, возникающая в результате конкуренции целей;
- оптимальный (вариационный) принцип принятия решения в ситуации выбора;
- эмоциональный механизм оценки и сравнения вариантов.

Дает ли это ответ на поставленные выше вопросы об алгоритмах (правилах) выбора целей и механизмах поведения? Нет, не дает. Перечисленные моменты целенаправленного поведения проливают свет преимущественно на общую структуру поведенческого акта, причем не на все его важные аспекты. Как правило, несколько в стороне остаются вопросы о том, как выбираются целевые ситуации, что такое в количественном смысле “наилучший приспособительный эффект”, от чего зависит и как влияет на выбор и закрепление плана поведения “величина положительного подкрепления”, как измеряется эта величина и как она связана, с одной стороны, с целевой ситуацией и, с другой стороны, с пройденным путем к цели.

В сложных случаях, к которым можно отнести большинство реальных ситуаций, однозначная фиксация конкретного наилучшего плана поведения невозможна ввиду разнообразия и неповторяемости в точности среды. Удачная программа поведения может быть зафиксирована и воспроизведена только в обобщенном виде. При этом остается вопрос: по какому количественному критерию и как происходит в случае сложного поведения не только умозрительный выбор цели, но и выбор пути к ней. Вернее, оцениваемые и сравниваемые варианты целенаправленного поведения должны включать совокупную оценку пути и цели. И главный вопрос – что при этом оценивается? Выбор варианта поведения основывается на эмоциональных оценках и их сравнении. А что такое эмоция и как она связана с “полезностью” или “приспособительным эффектом”?

Забегая вперед можно сказать, что более или менее точный (хотя и гипотетический) ответ на большую часть таких вопросов возможен, по-видимому, с учетом рассматриваемого ниже принципа тахТ и предположений о сути базовой задачи мозга, то есть задачи построения иерархической модели среды, сводящей реальные многоэкстремальные многопереборные задачи поведения к умозрительным одноэкстремальным, малопереборным. Об этом я буду говорить в следующей лекции.



В предыдущих лекциях мы рассматривали очень общее определение мышления, требующее конкретизации таких понятий, как живое, активность, модель среды, управление поведением в многоэкстремальной среде, восприятие, обучение, активные нейронные механизмы. В этой лекции мы сформулировали определение и в некотором приближении конкретизировали понятие “жизнь”. Конкретизацию других входящих в определение мышления понятий мы попробуем продолжить рассматривая задачи управления поведением и восприятия среды.

#### Лекция 4. **Формальная модель поведения**

##### **Характеристика задачи поведения. Внутренняя и внешняя работа.**

Рассмотренные в предыдущей лекции принцип устойчивого неравновесия Э.С.Бауэра, а также вопросы, касающиеся поведения, позволяют сделать ряд выводов. В живом организме нужно различать внутреннюю и внешнюю работу. Внутренняя работа это ассимиляция необходимых для биосинтеза веществ, биосинтез и диссимиляция. Внешняя работа бывает трех разных видов.

Первый вид внешней по отношению к клетке работы это работа внутренних органов (сердце, легкие, печень, почки и т.п.). Управляет этой работой вегетативная нервная система. Главная цель этой работы доставка к клеткам элементов питания и поддержание оптимального состояния внутренней среды организма. К этой работе относится понятие гомеостаз и идущие от кибернетики представления о роли отрицательных обратных связей и регулировании по отклонению. Мышление в этой работе не участвует.

Второй вид внешней работы (поведения) это реакции на внешние воздействия. К этой работе относятся понятия безусловный и условный рефлекс, схема стимул-реакция, сохраняющие реакции (по А.Ляпунову), динамический стереотип и идущие от кибернетики представления о разомкнутом контуре управления, т.е. об управлении по возмущению. К этому виду работы относится и сформулированное И.Павловым представление об уравнивании со средой. Центральная нервная система (кора головного мозга) в управлении этим поведением уже участвует как при обучении, так и при выборе реакций. Правда, эта работа мозга относительно проста и в плане рассматриваемой нами темы значительного интереса не представляет. Развитие схемы стимул-реакция это сознательный поведенческий акт и его теория в виде функциональной системы П.К.Анохина. В схеме функциональной системы наибольший интерес могли бы представить блок афферентного синтеза и принципы выработки решения о действии. Однако, эти вопросы в теории функциональной системы не раскрываются.

Третий вид внешней работы это целенаправленное поведение. Рассмотрение этого вида работы с учетом принципа устойчивого неравновесия позволяет сконцентрировать внимания на следующих моментах:

Первый: живой организм, как на клеточном уровне, так и в целом – это неравновесная система с активно поддерживаемым неравновесием и высоким уровнем свободной энергии.

Второй: поддержание устойчивого неравновесия на уровне организма происходит в результате целенаправленного поведения в среде, направленного на обеспечение материала и условий для биосинтеза. Это осуществляется путем удовлетворения первичных физиологических потребностей. Схема поведения должна строиться с учетом также и вторичных потребностей. Вторичные потребности в первую очередь характерны для человека. Эти потребности определяются групповым социальным способом существования и особенностями эмоционального механизма управления поведением.

Третий: живой организм должен быть постоянно активен – поведение непрерывная задача.

Таким образом, поведение не ограничивается ответными реакциями на внешние стимулы и не преследует цели “уравнивания со средой”. Необходимость всего этого существует, но главная и постоянная задача поведения – это удовлетворение потребностей, для которого необходимо достижение целевых ситуаций в среде. В основном для решения этой задачи эволюционно формировался мозг и развивалось мышление.

Главные проявляемые в поведении свойства живых организмов это активность и целенаправленность.

Все это нужно учитывать при построении модели поведения. Кроме того, нужно учитывать полученный Асмаяном и Голицыным предварительный экспериментальный результат, состоящий в том, что принятие решения животным в ситуации выбора подчиняется экстремальному принципу  $\max m_i V_i$  ( $m$ -мотивация,  $V$ -скорость ее удовлетворения).

Итак, по рассматриваемой теме можно сделать следующие выводы. Интересующее нас целенаправленное поведение, естественно, не осталось без внимания физиологов. Однако, целостной модели целенаправленного поведения в физиологии нет. Теория функциональной системы П.К.Анохина ограничивается рассмотрением отдельного поведенческого акта. За скобками остаются вопросы о возникновении, конкуренции и выборе доминант. Теория доминант А.А. Ухтомского говорит об этом только в самом общем виде без конкретизации принципов и механизмов выбора. При этом, ограничиваться, как это часто делают в физиологии, словами о конечном приспособительном эффекте явно недостаточно. Хотелось бы сформулировать задачу более точно. Целевых ситуаций может быть много. Какие ситуации, в каком порядке и почему становятся целями направленного поведения? Какая задача решается в поведении? Каковы правила принятия решения в поведении? Я

попробую сформулировать ответы на эти вопросы в виде принципа оптимальности (экстремального принципа) и вытекающих из этого принципа следствий.

## **Принципы оптимальности**

Сначала остановимся на принципах оптимальности в физике и биологии. Развитие науки обычно идет от эксперимента к теории, от наблюдения фактов к их обобщению. Эти обобщения могут фиксироваться в виде законов, которые в совокупности объясняют наблюдаемые факты и предсказывают новые. Такими законами, например, являются законы механики Ньютона, или законы геометрической оптики. Но возможен и следующий шаг обобщения в виде принципа оптимальности, который формулируется как требование минимума или максимума какой либо величины. Так все законы механики обобщает принцип наименьшего действия, а законы геометрической оптики обобщает принцип скорейшего пути Ферма. (Действие это произведение массы, пути и скорости. Авторство принципа наименьшего действия связывается с разными именами – Мопертюи, Эйлер, Лагранж.)

Многие ученые давно пришли к мысли о том, что в природе все делается оптимально и все обобщения, а значит и факты в любой области естествознания, могут быть выведены из единого принципа оптимальности. Нужно только понять, что экономит природа в объектах и явлениях, относящихся к этой области естествознания. В пределе может существовать и общий принцип оптимальности, определяющий все в нашем мире.

Мысль о том, что в природе все подчиняется какому-то принципу оптимальности и всю науку можно строить не снизу вверх - от эксперимента к теории, а сверху-вниз, от принципа оптимальности к частным законам была высказана Эйлером. Однако, ни Эйлеру, ни кому-либо другому найти такой общий принцип не удалось.

Что же экономит природа? В качестве вариантов напрашиваются и чаще всего рассматриваются: энергия, вещество, действие ( $mvs=\min$ ), энтропия (негэнтропия), информация.

А, может быть, экономится время?

Кстати принцип Ферма действительно говорит о том, что при распространении светового луча экономится время. В этом плане безусловный интерес представляют и термодинамические принципы наименьшей диссипации энергии и наискорейшего спуска. Диссипацией называется процесс перехода энергии из упорядоченной формы в неупорядоченную тепловую. С этим процессом связано возрастание энтропии. Согласно принципу наименьшей диссипации энергии в устойчивом состоянии любой термодинамической системы скорость диссипации энергии в ней минимальна. Указанный принцип является конечной характеристикой равновесного состояния. Для описания переходного процесса формулируется принцип наискорейшего спуска: в процессе приближения термодинамической системы к равновесному или стационарному состоянию функция внешней диссипации убывает наискорейшим возможным способом. Т.е. и в первом и во втором случае процессы оптимизируются по времени. (Принципы предложены А.И.Зотиним и А.А.Зотиним. Авторы считают, что эти термодинамические принципы оптимальности включают в себя второе начало термодинамики и охватывает практически все явления природы)

В плане рассматриваемой нами темы очень интересно также и общее утверждение Спинозы о том, что основной закон вещей – стремиться к максимально долгому существованию. На качественном уровне это понятно. Существует то, что “умеет” более или менее длительно существовать. Остальное разрушается и исчезает. Это переключается с рассматриваемым ниже принципом  $\max T$  в поведении живых и не только живых систем.

В биологии в качестве принципов оптимальности главным образом предлагались принципы экономии энергии, экономии ресурсов, максимума энтропии и максимума информации. Эти идеи оказались в определенной степени продуктивными. Предложенные принципы многое объясняют, но далеко не все, поскольку они работают при определенных как правило нестрогих формулируемых условиях. Например, экономия энергии при условии удовлетворения нужд организма или при условии достижения целевой ситуации, экономия ресурсов при достижении заданного результата, или максимум взаимной информации между стимулами и реакциями при условии достижения определенного результата и т.п.

С помощью подобных нестрогих условий (оговорок) можно практически всегда объяснить несовпадения результатов эксперимента и теории. Кроме того, почти всегда то, что задается только как находящееся на втором плане нестрогое внешнее условие, необходимое для выполнения этих принципов оптимальности (заданный результат, приспособительный эффект, полезность и т.п.), на самом деле должно находиться в центре внимания и определять цели и принципы поведения.

Более общим и естественным представляется рассматриваемый ниже принцип  $\max T$ .

## **Конкретизация задачи поведения**

Итак, первой задачей, для решения которой возник и эволюционно развивался мозг, возникло и развивалось мышление, явилась задача поведения живого организма в сложной среде.

Я уже говорил, что часто задачи поведения, так же как и задачи автоматического управления, рассматриваются по следующей упрощенной схеме: система путем управления (поведения) стремится удержать оптимальное по каким либо собственным критериям состояние, отклонение от которого происходит за счет

воздействий, поступающих на систему из среды. При этом причина поведения - воздействие, изменяющее состояние (стимул, возмущение) находится вне системы, а цель - восстановление состояния – внутри.

Рассматривая такую схему часто говорят, что единственная цель поведения живого организма в среде это гомеостаз, т.е. обеспечение постоянства внутренней среды организма при поступающих из среды возмущениях. Из этого вытекают представления о необходимости поддержания равновесия в системе организм- среда, об основной схеме работы по типу стимул-реакция, условный и безусловный рефлекс, об обратной связи, как об основном регулирующем факторе. Все это представляется серьезным упрощением, порожденным тем, что не учитываются или отодвигаются на второй план принципы организации живой материи.

В свете рассмотренных в предыдущей лекции представлений о неравновесности живого и принципе устойчивого неравновесия представляется, что гораздо естественней рассматривать поведение в рамках прямо противоположной схемы. По этой схеме причина поведения – собственная накапливающаяся неустойчивость, приводящая к ухудшению состояния, в том числе и к отклонению от нормы гомеостатических переменных, находится в самой управляемой системе, а цели – внешние воздействия (факторы), необходимые для улучшения состояния (в том числе и для поддержания гомеостаза) – в среде.

Повторим то, о чем уже говорилось в предыдущей лекции. Выделение и использование свободной энергии в живой системе сопровождается постоянным разрушением и восстановлением структуры живой ткани, находящейся в неравновесном состоянии и обладающей свободной энергией. Для этого процесса необходимо поступление в систему извне определенных, в первую очередь, органических веществ. Таким образом, направленное специфическое взаимодействие со средой в целевых ситуациях поведения необходимо для существования живой материи. Живая система должна быть открыта и постоянно активна. Поведение – непрерывная задача.

## **Принцип maxT**

В связи с этим естественным представляется принцип maxT – принцип максимизации времени пребывания системы внутри условной области существования, определяемой как область допустимых значений регулируемых переменных.

Введение принципа maxT и упрощенной формальной модели поведения базируется на следующих конкретизирующих предпосылках. Для живых организмов характерно наличие потребностей. Удовлетворение физиологических потребностей, являющееся необходимым условием

существования, может происходить только в определенных ситуациях взаимодействия организма со средой. Эти ситуации являются альтернативными целями направленного поведения. Поведение животных в каждый момент направлено, в общем случае, на достижение одной цели, соответствующей одной потребности. Ситуации необходимого целевого взаимодействия со средой в общем случае альтернативны.

Для живого организма, рассматриваемого как целое, характерно накопление неустойчивости в том смысле, что физиологические потребности имеют общее свойство, заключающееся в их нарастании с течением времени. Поддержание устойчивого неравновесия - непрерывная задача живых организмов, решаемая на клеточном уровне за счет внутренней работы, то есть ассимиляции органических веществ и синтеза живой неравновесной структуры. Обеспечивается эта работа на уровне целого организма за счет удовлетворения потребностей путем активного целенаправленного поведения в среде.

Физиологические параметры (переменные), определяющие наличие и величину первичных физиологических потребностей, должны иметь согласованные значения. Можно упрощенно предположить, что в многомерном пространстве физиологических переменных имеется область, соответствующая нормальному состоянию организма. Можно также предположить, что имеется еще одна более широкая область – область допустимых значений, выход за пределы которой губелен для организма, и объективная задача поведения – максимально долго поддерживать величины физиологических переменных в пределах этой области.

Задача поведения не исчерпывается прямой задачей выживания отдельной особи, т.е. необходимостью поддержания значений первичных физиологических переменных организма и соответствующих им потребностей внутри области допустимых значений. К первичным физиологическим потребностям самого организма добавляются потребности, определяемые необходимостью продолжения рода, а также вторичные потребности, косвенно влияющие на первичные. Последнее особенно характерно для человека вследствие его сложного общественного способа существования. Добавление в рассмотрение вторичных потребностей не меняет общей схемы: система (живой организм) обладает внутренней неустойчивостью – неудовлетворяемые потребности увеличиваются. Потребности в общем случае альтернативны, то есть удовлетворяются раздельно и поочередно.

Еще раз определим принцип оптимальности в поведении. Цель поведения - это максимизация времени пребывания системы внутри области допустимых значений регулируемых переменных (первичных и вторичных потребностей) – принцип maxT.

Из принципа оптимальности должны вытекать все частные закономерности. Мы постараемся как следствие принципа maxT вывести формальное правило принятия решения животным (и человеком) в ситуации выбора. Кроме того, как следствие принципа maxT формулируются гипотезы о формировании животными составляющих эмоциональных оценок как своего состояния, так и реальных и умозрительных ситуаций взаимодействия со

средой. В дальнейшем мы рассмотрим принцип maxT более широко, то есть попробуем применить его не только к поведению живых организмов, но и к оптимальному управлению (“поведению”) в некоторых искусственных системах.

## Формальная модель

Рассмотренные представления о сути поведенческого акта позволяют представить себе упрощенную качественную модель поведения, основанную на принципе оптимальности (maxT) и вытекающей из этого направленности в выборе целей. От этих качественных представлений можно попробовать перейти к формальной модели. Упрощенная формальная модель поведения основывается на решении следующей задачи.

Пусть имеется n-параметрическая система  $X(x_1, x_2, \dots, x_n)$ , параметры которой увеличиваются с постоянными скоростями  $U(U_1, U_2, \dots, U_n)$ . В каждый момент путем управляющего воздействия возможно уменьшение любого, но только одного из параметров  $x_i$  с фиксированной скоростью  $V_i$ . Пусть область допустимых значений параметров системы ограничена выпуклой поверхностью  $F(x)=0$  в первом ортанте т.е.  $x_i \geq 0$ .

Система представляется точкой X в n-мерном пространстве. Если никаких управляющих действий не делать, то точка X движется к границе области допустимых значений. Требуется найти условия, определяющие возможность организации управления, удерживающего систему внутри области допустимых значений параметров, а также и саму тактику управления, оптимального в том смысле, что при этой тактике максимизируется время пребывания системы внутри области допустимых значений

Можно получить различные алгоритмы решения этой задачи, основанные на определении точки выхода системы на границу области допустимых значений. Любой из этих алгоритмов естественно назвать глобальным, поскольку для его реализации необходимо предварительное рассмотрение всей задачи в целом и определение конечной точки системы. Однако такой подход к решению задачи затрудняет физиологические интерпретации, поскольку трудно предположить, что в живой системе в какой бы то ни было форме заложены сведения о ее конечной точке. Кроме того, реализация глобального алгоритма даже в случае когда задача полностью и точно определена часто требует слишком большого времени. Поэтому на алгоритм управления естественно наложить требование, чтобы решение, принимаемое системой в каждый момент времени, было функцией только состояния системы и среды в этот момент времени. Такой алгоритм управления будем называть локальным.

Решение сформулированной задачи с условием получения локального правила выбора управляющего действия позволяет, не вдаваясь в подробности, сформулировать следующие результаты.

Введем функцию, оценивающую состояние системы:

$$Z(X) = F(X) - F(0) \quad j$$

К этой функции мы будем обращаться в дальнейшем при рассмотрении разных тем.

Свойства функции  $Z(X)$  таковы. Значение функции в начале координат  $Z(0)=0$ . В любой точке границы области допустимых значений переменных X при  $F(X)=0$  значение функции  $Z(X)$  равно постоянной величине, равной  $F(0)$ . Всегда можно сделать так, чтобы значение  $F(0)$  было положительно. Тогда в любой точке границы области допустимых значений параметров X значение функции  $Z(X)$  одинаково, положительно и максимально. Поскольку поверхность  $F(X)=0$  выпукла, функция  $Z(X)$  изменяется монотонно при монотонном изменении любых аргументов X.

Используя формулу полного дифференциала можно определить приращение функции  $Z(X)$  для j-го варианта управления:

$$\Delta Z_j = \sum_i \frac{\partial F(X)}{\partial x_i} U_i - \frac{\partial F(X)}{\partial x_j} V_j$$

К решению задачи ведет требование минимума приращения функции Z на выбираемом варианте (шаге) управления. Поскольку первое слагаемое приращения одинаково для всех j (вариантов управления), то оптимальная тактика, определяющая необходимое в каждый момент действие задается вторым слагаемым. Таким образом, получаем локальное правило выбора варианта управления

$$\max \frac{\partial F(x)}{\partial x_j} V_j$$

$$j=(1, \dots, n)$$

где j это вариант управления на каждом оцениваемом шаге поведения, другими словами, в рассмотренной упрощенной задаче каждый вариант управления (j) - это достигаемый в один шаг вариант цели поведения.

Для оперирования полученным правилом выбора управляющего действия нужно знать функцию  $F(X)$  и уметь находить ее частные производные. Проще всего случай, когда переменные  $X_i$  независимы и область их допустимых значений n-мерный параллелепипед.

Тогда  $0 \leq X_i \leq X_{i\max}$  или  $F(X) = \Pi (X_{i\max} - X_i) = 0$

и локальное правило выбора управляющего действия имеет вид:

$$\max(V_j / (X_{j\max} - X_j))$$

j

Если переменные  $X_i$  зависимы, то с учетом ограничений, накладываемых на вид функции  $F(X)$ , проще всего и естественнее считать, что граница области допустимых значений переменных аппроксимируется функцией второго порядка, т.е. функция  $F(X) = 0$  определяет  $n$ -мерный эллипсоид.

В этом случае локальное правило выбора управляющего действия имеет вид:

$\max_j K_j X_j V_j$

j

Приведенные примеры показывают, что процедура определения частных производных функции  $F(X)$  для оперирования полученным соотношением при выборе управляющего действия может быть достаточно простой. В первом случае необходимо измерение близости переменных к предельному значению, во втором случае необходимо измерение отклонения величин переменных от нулевого значения.

Состояние системы (организма) характеризуется текущим значением вектора  $X$ . В зависимости от конкретных условий то есть от соотношения скоростей увеличения и уменьшения параметров  $X$ , а также от реализуемого алгоритма управления, точка, отображающая состояние системы в пространстве  $\{X\}$ , может перемещаться к границе области допустимых значений, удаляясь от начала координат, или наоборот двигаться к началу координат, удаляясь от границы области регулирования.

В первом случае система движется к разрушению, и полученное правило управления максимально замедляет это движение ( $\max T$ ). Во втором случае система удаляется от границы области допустимых значений и то же правило управления максимально ускоряет это движение. При этом можно говорить, что управление (поведение) подчиняется принципу оптимальности  $\min T$  – минимизируется время прихода системы в начало координат. В физиологической интерпретации можно говорить, что минимизируется время удовлетворения всех потребностей. Обе интерпретации принципа оптимальности  $\min T$  и  $\max T$  полностью эквивалентны. В недетерминированной среде, то есть при возможности изменения условий задачи управления, максимально быстрое удаление от границы области допустимых значений создает запас времени и максимизирует время выхода системы на границу области регулирования в случае, когда система начнет вынужденно двигаться к этой границе при оптимальном управлении, т.е. мы опять приходим к принципу  $\max T$ .

Рассмотренная формальная задача является серьезным упрощением реальной задачи поведения. Главное упрощение - это одноэкстремальность задачи, позволившая получить локальное правило управления. Тем не менее, эта задача, по-видимому, отражает принципиальную суть реальной задачи, которая сводится к рассмотренной упрощенной схеме. Общие принципы этого сведения мы рассмотрим теперь и уточним в дальнейшем.

## Реальная задача

Приведенная формальная задача полностью соответствует интереснейшим экспериментам по изучению принятия решения животным в ситуации выбора, описанным в работе Н.В.Асмаяна и Г.А.Голицына. [2]. Эти эксперименты мы будем подробно рассматривать в следующей лекции. Собственно, рассмотренная задача управления и упрощенная формальная модель поведения и возникли с учетом этих экспериментов. Так же как в физиологических экспериментах в рассмотренной упрощенной задаче все целевые ситуации достигаются за один шаг. Это, так же как и другие упрощения (постоянство векторов  $U$  и  $V$ , выпуклость области допустимых значений регулируемых переменных) позволили вывести локальное правило выбора действия, то есть не решать задачу до конца и не просчитывать время достижения границы области допустимых значений переменных на всех возможных траекториях управления. Рассмотренная упрощенная задача поведения одноэкстремальна. Многоэкстремальности в этой задаче при любом варианте поведения возникнуть не может, поскольку все цели достигаются за один шаг.

Однако эта одноэкстремальная задача принципиально отличается от многоэкстремальной задачи поведения животных и человека в реальной среде. Тем не менее, поведение живых организмов все же, по-видимому, подчиняется принципу оптимальности  $\max T$  и локальному правилу выбора действия, суть которого определяется полученной формулой. Противоречие снимается если предположить, что мозг в процессе обучения и мышления успешно выполняет сведение реальной многоэкстремальной задачи к рассмотренной одноэкстремальной.

Все изменения в неживом мире направлены по градиенту – равнодействующей всех действующих на систему физических сил. Изменения происходят в направлении быстрого уменьшения свободной энергии системы и заканчиваются при достижении ее ближайшего локального минимума. Сходным образом ведут себя и растения, обладающие элементарным восприятием среды и реализующие простое градиентное поведение. Растения, так же как и животные имеют потребности, но они всегда “знают” что им нужно делать: они тянутся к свету, теплу, питательным элементам - туда, где в данный момент лучше (так называемые тропизмы растений), то есть растения в своем поведении могут пользоваться локальным, градиентным правилом выбора. Для управления таким поведением мозг не нужен.

В отличие от растений животные могут перемещаться в пространстве. Это значительно расширяет их возможности по удовлетворению потребностей и, в то же время, значительно усложняет задачу поведения. Животные вынуждены жить в условиях необходимости решения многоэкстремальных задач. Для достижения какой-то цели они могут преодолевать боль, страх, препятствия, подвергаться нападениям, вступать в борьбу, тратить энергию, ухудшать свое функциональное состояние с целью в конечном счете его улучшить. Другими словами, в процессе поведения нужно уметь находить глобальный экстремум, преодолевая локальные.

Таким образом, в отличие от неживой природы и растений для животных характерна необходимость движения против градиента непосредственно действующих сил. В отношении причин, обуславливающих поведение высокоорганизованных живых организмов, обладающих развитой нервной системой допустимо применение термина “психические силы”. Однако, поведение живых организмов может не определяться и равнодействующей непосредственно действующих “психических сил”, если считать непосредственно действующими те силы (мотивации и эмоции), которые вызываются внутренним физиологическим состоянием организма и его взаимодействием с непосредственно воспринимаемым окружением (средой).

Однако, движение против градиента “психических сил” высокоорганизованных живых организмов в конечном счете является кажущимся. Дело в том, что в мозге (животного, человека) формируется при обучении общее целостное отображение среды и модель взаимодействий со средой. Реальная ситуация на входе системы рассматривается только как актуализированный частный фрагмент некоторой обобщенной и укрупненной ситуации, развернутой во времени и пространстве за пределы непосредственного восприятия. Эта укрупненная и обобщенная, зависящая от опыта субъективная ситуация и формирует “психические” силы, по градиенту которых направляется поведение.

Глобальный экстремум можно находить путем полного перебора. Человек действует по-другому. Он пытается свести многоэкстремальную задачу, которую можно решать только методом перебора, к градиентной, одноэкстремальной, которую вообще не надо решать, так как в каждой точке нужно оценить и сравнить очень ограниченное число вариантов. То есть человек создает новое информационное отображение среды: он решает задачу с помощью обобщения и укрупнения. За счет этого в знакомой среде можно действовать почти автоматически – по прогнозируемому градиенту эмоциональной оценки ситуации. В принципе, так же действуют и животные. Несколько подробнее эти вопросы будут рассмотрены в последующих разделах.

Таким образом, мозг нужен человеку и животному для создания модели среды, позволяющей решать многоэкстремальные задачи поведения, полагаясь на локальное правило принятия решения и выбора среди альтернативных вариантов поведения. Эта модель среды должна строиться на иерархии обобщаемых и укрупняемых представлений, дающих возможность определять и фиксировать доминантные целевые ситуации, и находить одношаговые переходы к целевым ситуациям при умозрительном планировании поведения. На уровне одношаговых переходов задача становится одноэкстремальной.

## **Свобода поведения систем с активно поддерживаемым неравновесием.**

Итак, живые организмы, т.е. системы с активно поддерживаемым неравновесием должны путем поведения (управления) удерживать величины регулируемых переменных (потребностей) внутри области допустимых значений -  $D_x$ . Состав переменных  $\{X\}$  может определяться по-разному. У животных регулируемые переменные это первичные физиологические потребности, удовлетворение которых необходимо для физического существования как отдельной особи, так и вида. У человека в состав потребностей, определяющих область существования системы, помимо первичных физиологических потребностей входят и вторичные, чаще всего субъективные потребности. Например, такими потребностями может быть: играть в карты, ходить в театр, писать стихи, сажать капусту и тому подобное.

Выход значений переменных за пределы области существования системы, построенной с учетом вторичных потребностей, не ведет, в общем случае, к физической гибели. Человек, как правило, может построить взамен разрушенной иную систему с иной областью существования, включающей другие вторичные потребности, или, в пределе, даже не включающей никаких вторичных потребностей. Хотя иногда творческий человек, например, художник или композитор может с большой долей истинности утверждать, что для него творить так же жизненно необходимо, как пить, есть или дышать.

Активные неравновесные системы можно разделить на две группы. Критерий деления – характер необходимого внешнего поведения и взаимоотношения со средой. Первая группа – это автономные, относительно свободные системы, поведение которых (управление) определяется следующими факторами.

1. Взаимодействие со средой осуществляется системой активно, по всем параметрам, поддерживаемым в состоянии устойчивого неравновесия. Поведение полностью и непосредственно обеспечивает все необходимые для существования внешние компоненты..
2. Существуют ограничения возможностей взаимодействия со средой.
3. Существенным фактором, определяющим поведение, может являться конкуренция с другими активными системами и борьба между ними.

4. Помимо необходимых полезных взаимодействий со средой возможны и вредные воздействия, непосредственно ухудшающие состояние системы. Эти воздействия могут быть как пассивными ненаправленными, так и активными направленными. Один из аспектов поведения – это избегание, компенсация или преодоление этих воздействий. Решается эта задача автономно – силами самой системы.

Примеры автономных относительно свободных систем это животные в свободном поведении, человек вне общественных связей, например отшельник или Робинзон, натуральное хозяйство, независимая фирма.

Единственный результат функционирования относительно свободной автономной системы – это ее существование как некоторой качественной определенности. Промежуточные приспособительные результаты направлены на достижение этого главного результата.

Вторая группа – это специализированные системы с ограниченной свободой поведения. Примерами таких систем могут быть клетка в составе организма или человек в обществе. Системы с ограниченной свободой поведения входят в качестве элементов в состав некоторой макросистемы и, в общем случае, независимо существовать не могут. Поведение таких систем обладает следующими особенностями:

Работа микросистемы направлена не прямо на обеспечение необходимого взаимодействия со средой, а на создание некоторого фактора (продукта), используемого в макросистеме;

Деятельность (поведение) микросистемы узкоспециализирована;

Необходимое взаимодействие микросистемы со средой и ее защита от агрессивных внешних воздействий обеспечивается макросистемой.

Таким образом, деятельность системы с ограниченной свободой поведения в основном направлена на обеспечение существования макросистемы, которая, в свою очередь, обеспечивает удовлетворение потребностей входящих в нее микросистем и их защиту от воздействий среды. То есть микросистемы не прямо, а косвенным образом, работая на макросистему и тем самым на другие микросистемы, обеспечивают себе простую, беспроблемную искусственную среду, достаточную для существования.

Мы выделили крайние точки спектра от полной автономности и полной свободы до полной зависимости. Как всегда, возможны промежуточные, смешанные варианты.

Биологическая эволюция прошла путь от полностью автономной и полностью свободной клетки к многоклеточному синергическому организму, имеющему в составе полностью зависимые клетки с регламентированным поведением. Социальная эволюция идет от полностью автономных и свободных животных и человека к человеку в обществе, иными словами к синергическим (взаимозависимым) системам с ограниченной свободой поведения. На этом пути жизнь человека, то есть борьба за существование значительно упрощается, а поведение все более регламентируется. В то же время, неизбежно связанные с этим ограничения свободы поведения могут приводить к уменьшению возможностей возникновения и удовлетворения вторичных потребностей.

К чему это придет в пределе неизвестно. Много зависит от того, как будут формироваться вторичные потребности микросистем и будет ли социальная макросистема обеспечивать удовлетворение этих потребностей. Некоторые возможные, но, безусловно, не обязательные варианты, описываются в фантастических романах о тоталитарных обществах будущего. Кстати заметим, что без проблем удовлетворяющей первичные и не имеющей вторичных потребностей системе с ограниченной свободой поведения, например клетке в составе многоклеточного организма, мозг не нужен.

В связи со всем этим возникает вопрос: свобода, за которую часто борются с оружием в руках, это хорошо или плохо? Полностью свободной автономной активной неравновесной системе нужно самостоятельно решать задачу выживания путем организации всего *необходимого* взаимодействия со средой. Полностью зависимой специализированной системе с максимально ограниченной свободой поведения задачу выживания, а в пределе и никаких других задач кроме выполнения своей специализированной функции, решать не нужно.

Таким образом, полностью свободной активной динамической системе в общем случае жить трудно, ее жизнь целиком зависит от среды и может быть сложна, многообразна, опасна и тяжела. Один из аспектов этого выразил Бернард Шоу, сказав: “Свобода это ответственность. Вот почему большинство дюдей боится ее”. Полностью зависимой системе с ограниченной свободой поведения жить просто, но не всегда легко, и обычно не интересно, поскольку интерес к жизни чаще всего связывается с наличием вторичных потребностей или со злоупотреблением первичными. Общего надежного рецепта для определения, выбора и достижения оптимального и одинакового для всех уровня свободы не существует, хотя и имеются претендующие на это различные социологические теории. Повидимому, выбрать и зафиксировать оптимальный уровень свободы нельзя. Можно только понять закономерности развития синергического человеческого социума и степень неизбежности этих закономерностей.

## Лекция 5. Физиологические интерпретации формальной модели поведения

## Мотивации и потребности

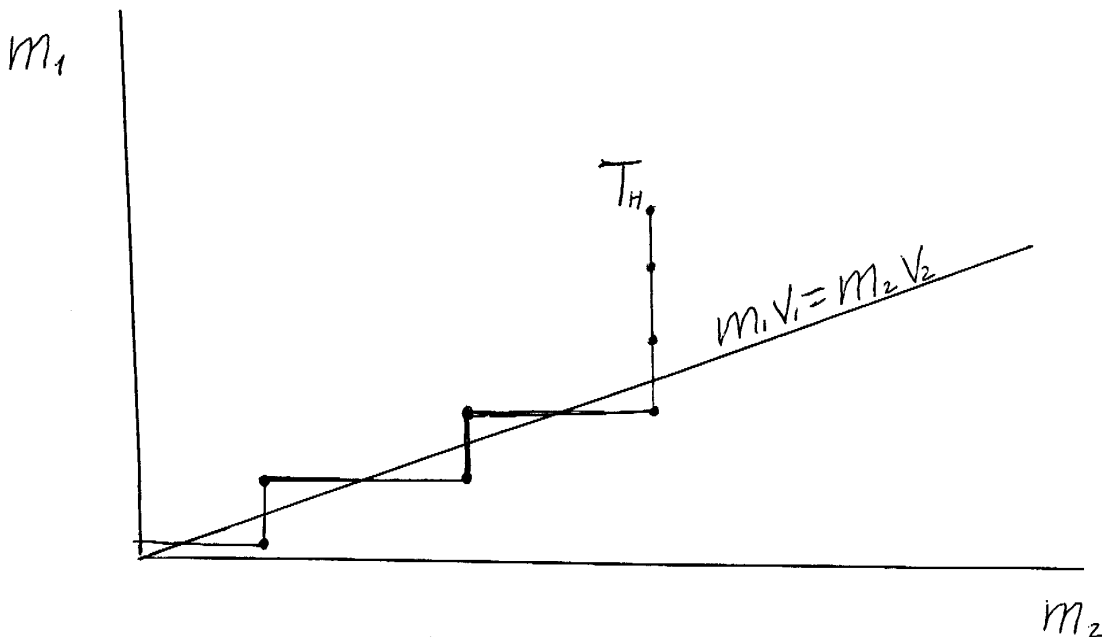
Введенная упрощенная формальная модель поведения допускает следующие физиологические интерпретации. Уже отмечалось, что если функция  $F(X) = 0$  определяет границу области допустимых значений регулируемых переменных как поверхность второго порядка, то приращение целевой функции для  $j$ -го варианта поведения равно

$$\Delta Z_j = \sum_i K_i X_i U_i - K_j X_j V_j$$

и общее локальное правило выбора управляющего действия преобразуется в правило  $\max_j K_j X_j V_j$

Физиологические интерпретации этого правила выбора управляющего действия основаны на очень интересных экспериментальных результатах, полученных Н.В.Асмоляном и Г.А.Голицыным. Общая схема и результаты экспериментов были следующими. У собаки вырабатывалось знание о том, что в ответ на световой сигнал она может нажать на рычаг А или рычаг Б и получить соответственно порцию воды или порцию корма. Вырабатывались также знания о размерах (небольших) этих фиксированных порций. Размеры порций определяют скорости удовлетворения потребностей голода и жажды, обозначаемые ниже как  $V_1$  и  $V_2$ . Экспериментаторы научились измерять потребности и мотивации у собаки и определили, что мотивации линейно связаны с потребностями, т.е.  $m = kx$ , где  $m$  – мотивация,  $x$  – потребность.

Поведение собаки в эксперименте иллюстрирует рисунок.



Начальное состояние – точка  $T_n$ , определяется начальными значениями мотивации голода– $m_1$  и жажды– $m_2$ . Собака в зависимости от начального состояния несколько раз подряд в ответ на световой сигнал нажимает на один и тот же рычаг и только ест или только пьет (движение точки, отражающей на рисунке состояние, параллельно оси  $m_1$  или оси  $m_2$ ) до тех пор, пока точка  $(m_1, m_2)$  не попадет на прямую  $m_1 v_1 = m_2 v_2$ . После этого поведение собаки изменяется – она ест или пьет в такой последовательности, что движение точки, отображающей состояние, происходит в скользящем режиме вдоль прямой  $m_1 v_1 = m_2 v_2$ .

Полученный экспериментальный результат требует ответа на вопрос: почему животное в ситуации выбора принимает решение и действует в соответствии с приведенным правилом, или в чем состоит приспособительный смысл такого поведения? Ответ не очевиден. Г.А.Голицын пытался привлечь для объяснения результатов эксперимента универсальную теорию управления В.Ф.Кротова, но убедительного результата не получил.

Повидимому, отсутствие содержательной интерпретации результатов экспериментов стало одной из причин того, что Н.В.Асмолян и Г.А.Голицын свою интереснейшую экспериментальную работу по определению принципов принятия решения животным в ситуации выбора не продолжили. Голицын, занимаясь поиском принципа оптимальности в поведении животных, пришел к принципу максимума информации. Кстати, этот принцип описанных экспериментальных результатов не объясняет.



Эти результаты можно объяснить на основе принципа  $\max T$  и рассмотренной нами упрощенной формальной модели поведения. Эксперимент показал, что в случае двоичного выбора между возможностями удовлетворения мотиваций голода и жажды поведение собаки хорошо согласуется с формулой выбора:

$$\max_j m_j v_j = \max_j K_j X_j v_j \quad j = 1, 2$$

$j$

Правило выбора, полученное при решении формальной задачи для случая, когда граница области регулируемых переменных является поверхностью второго порядка, и правило выбора, экспериментально полученное в физиологическом эксперименте совпадают. С учетом этого, приращение целевой функции  $Z(X)$  можно выразить через мотивации:

$$\Delta Z_j = \sum_i m_i U_i - m_j V_j$$

Как мы уже отмечали, полученное для упрощенной задачи локальное правило выбора варианта поведения должно работать и в полной многоэкстремальной задаче. Это обеспечивается наличием уровней укрупнения и обобщения в строящейся в мозге модели проблемной среды. Реализация локального правила сравнения вариантов поведения, направленного на выполнение принципа  $\max T$  без определения и анализа всех точных далеко прослеживаемых возможных траекторий поведения, основана на получении оценок по шкале хорошо-плохо, т.е. на эмоциональном выборе.

### Формулы эмоций

Эмоции складываются из нескольких (многих) компонент. Мы рассмотрим качественный смысл этих компонент и формальное выражение некоторых из них применительно к приведенной выше упрощенной модели поведения. Поведение в общем случае должно сводиться к этой упрощенной одноэкстремальной модели. Поэтому представляется, что эмоции и в общем случае должны иметь тот же смысл и ту же структуру.

Во-первых, на общий эмоциональный фон влияет совокупное значение текущих потребностей. В плане решенной выше формальной задачи эта оценка определяется положением вектора первичных и вторичных потребностей внутри области допустимых значений, т.е. формулой, определяющей целевую функцию

$$E_1 = k_1 Z = k_1 (F(X) - F(0))$$

Во-вторых, на эмоциональный фон влияет соотношение необходимого и возможного. При решении приведенной формальной задачи было получено, что возможности управления (поведения) характеризуются величиной

$$R = \sum_i \frac{u_i}{v_i} - 1$$

При  $R > 0$  и любом управлении система с течением времени выходит за пределы области допустимых значений параметров. Аналогия - ситуация, имеющая отрицательную эмоциональную окраску, определяемую тем, что живой организм не справляется с задачей управления, т.е. не может минимизировать потребности путем доступного ему поведения.

При  $R \leq 0$  возможно управление, сколь угодно долго удерживающее систему в пределах области допустимых значений. Аналогия - ситуация имеющая положительную эмоциональную окраску, определяемую тем, что живой организм справляется с задачами поведения, определяемыми потребностями и средой. Таким образом, знак и величина  $R$ , определяемые формулой задают еще одну составляющую эмоции, а именно вторую составляющую общего эмоционального фона.

Таким образом,

$$E_2 = k_2 R = k_2 \left( \sum_i \frac{u_i}{v_i} - 1 \right)$$

Количественным критерием оценки возможных шагов поведения (управления) с точки зрения улучшения или ухудшения состояния системы является величина ожидаемого приращения значения целевой функции  $\Delta Z_j$ . Эта оценка сходна с эмоциональной оценкой живыми организмами возможных вариантов поведения в ситуации альтернативного выбора. Таким образом, третья составляющая эмоции можно условно определить как

$$E_3 = E_j = -\Delta Z_j = m_j V_j - \sum_i m_i U_i$$

Третья составляющая эмоции, также как и вторая, может быть как положительной, так и отрицательной. В положительную сторону действует мотивация и соответствующая ей потребность, удовлетворяемая в рассматриваемом альтернативном варианте поведения, в отрицательную все остальные.

Переменные  $\{X\}$ , входящие в правила выбора действия, и в формулы, определяющие первую, вторую и третью составляющие эмоций, естественно назвать "эмоциональными переменными".

Говоря об эмоциях в контексте принятия поведенческого решения в ситуации выбора нужно подчеркнуть, что в данном случае речь идет лишь об эмоциональной оценке альтернативных вариантов поведения, а также и об оценке ситуации в целом, определяемой возможностью удовлетворения первичных и вторичных потребностей.

Таким образом, для упрощенной формальной модели поведения можно определить поведенческую составляющую эмоций как

$$E_p = E_1 + E_2 + E_3$$

Роль эмоций не исчерпывается принятием решения в ситуации выбора и планированием поведения. Эмоции участвуют также и в непосредственном управлении восприятием и поведением. Как в восприятии, так и в поведении большую роль играет прогноз или предвидение ситуаций в среде и ситуаций взаимодействия со средой. Это предвидение выражается открытым П.К.Анохиным и Е.Стреж принципом акцептора действия и рассматриваемым ниже акцептором восприятия. Совпадение реальной ситуации с акцептором (предвидением) сопровождается слабой положительной эмоциональной оценкой. Несовпадение предвидения с реальной ситуацией сопровождается отрицательной эмоциональной оценкой, приостановкой поведения и ориентировочно-исследовательской реакцией. Нарушение предвидения (акцептора) может сопровождаться эмоциональной переоценкой ситуации по всем приведенным выше критериям. Переоценка возможна как в положительную, так и в отрицательную сторону. Результатом эмоциональной переоценки может быть изменение поведения.

Кроме того, эмоциональные оценки возникают при формировании в нейронной сети, поддержании и гашении стабильных и кратковременных очагов возбуждения в процессе постановки и решения поведенческих и формальных задач мышления. Эмоциональные оценки динамически возникающих и гасящихся очагов возбуждения сопровождают восприятие произведений искусства. Эти вопросы еще будут обсуждаться в дальнейшем.

Таким образом, эмоциональная оценка живым организмом своего состояния, своих возможностей, вариантов поведения и ситуаций взаимодействия со средой очень многопланова. Выразить все это условной единой невычислимой формулой эмоций, как это делается в некоторых работах, конечно можно, но не представляется полезным. Общая формула, объединяющая отмеченные компоненты будет слишком громоздкой. Кроме того, выделенные выше составляющие эмоций проблему не исчерпывают.

Рассматриваемые аспекты схемы формирования и использования эмоций в процессе поведения и восприятия коррелируют с некоторыми существующими “поведенческими” определениями эмоций, введенными, например, П.К.Анохиным, Г.А.Голицыным или П.В.Симоновым и др.. Однако, существуют и иные аспекты. Один из них – мощные эмоциональные состояния, например, страх, горе, восторг, отчаяние. Мы уже отмечали, что Чарльз Дарвин считал такие экстремальные эмоциональные реакции атавизмом. Однако можно предположить, что подобные эмоциональные реакции также имеют приспособительный смысл и играют, главным образом, роль прерываний или блокировок механизма эмоционального выбора среди альтернативных линий поведения, в ситуации, когда все действия бесполезны, не нужны или вредны, либо полезны только рефлекторные действия, например, замереть, бежать, кричать и т.п.

Есть еще один аспект эмоций, который иногда противопоставляют поведенческому. Этот аспект можно выразить, немного перефразируя Гумилева: “Все это так, но как же быть с зарею, которую ни съесть, ни выпить, ни поцеловать?”. Можно также вспомнить, что почему-то приятно смотреть на пламя костра, морской прибой, приятна или неприятна музыка, абстрактная живопись.

Эти, как и любые другие эмоциональные реакции, также должны иметь приспособительный смысл. Предположения о механизме и приспособительном смысле подобных не имеющих очевидного поведенческого аспекта эмоциональных реакций, по-видимому, можно сформулировать только с учетом представлений о некоторых нейронных механизмах. Более того, представляется, что ‘продвинутые’ модели нейронных механизмов мозга с необходимостью должны объяснять подобные эмоциональные реакции.

Этот вопрос еще будет затрагиваться в следующих лекциях, однако существование не связанных с удовлетворением потребностей эмоциональных реакций дает возможность заранее, т.е. еще до рассмотрения моделей нейронных механизмов мозга, предположить следующее. Во-первых, работа нейронных механизмов мозга в процессе восприятия может получать эмоциональную оценку вне зависимости от смысловой (поведенческой) нагрузки восприятия и эта оценка дополняет общую эмоциональную оценку состояния организма.

Во-вторых, нейронные механизмы мозга могут работать в специфических эмоционально оцениваемых режимах, определяемых характеристиками поступающих из среды входных воздействий. В-третьих, для таких воздействий среды чаще всего характерны ритмическая составляющая - музыка, прибой и/или случайная (шумовая) составляющая – прибой, свет костра.

На основании сказанного можно утверждать, что одной из отправных точек в исследовании нейронных механизмов мозга, включая исследование структуры и режимов работы нейронных сетей, и, в конечном счете, в построении теории работы мозга, может быть исследование особенностей характеристик внешних воздействий, получающих не связанные с внешним поведением положительные и отрицательные эмоциональные оценки, и отражение этих особенностей на нейрофизиологическом уровне.

В частности, было бы интересно установить формальную связь между параметрами нейронов и структуры нейронной сети и гармоническими характеристиками музыки. Экспериментально определить, чем отличается работа нейронов и нейронной сети при восприятии гармонии, мелодии, какофонии, диссонансов и т.п.

Интеграция оценок, строящихся в зависимости от значений эмоциональных переменных, и выбор варианта поведения происходит в эмоциональном центре нервной системы. Поведение, в конечном счете, направлено на максимизацию положительных и минимизацию отрицательных эмоциональных оценок. В животном мире управление поведением, основанное на локальных эмоциональных оценках, как правило, работает правильно, реализуя направленность поведения на максимизацию времени существования, то есть выживание, особи и вида (maxT).

У человека в современном цивилизованном обществе с его широкими возможностями формирования и удовлетворения многочисленных, не вполне естественных, а часто и вполне неестественных вторичных потребностей, это, к сожалению, не всегда так. Формирование расширенной области управления с добавлением неестественных эмоциональных переменных может приводить к получению неправильного результата управления поведением, то есть принцип maxT, реализуемый в искусственно расширенной области управления, может приводить к результату, в значительной степени не отвечающему задаче максимизации времени существования особи и вида.

Еще значительно хуже и опасней возможность непосредственного химического или электрического воздействия на эмоциональный центр. Такое воздействие может сопровождаться прямым получением положительных эмоций и иногда почти полностью нарушать работу эмоционального центра по управлению поведением. Иллюстрацией этого являются наркомания, а также известные классические эксперименты Экклза и других авторов, посвященные исследованиям самораздражения. В этих экспериментах животное может постоянно нажимать на рычаг, посылая через вживленный в мозг электрод идущие в эмоциональный центр электрические импульсы, и делает это до полного истощения, забывая обо всех других потребностях.

## **Выводы**

**Рассмотрение задачи поведения позволяет сделать ряд выводов.**

**Первый: для живых организмов характерна накапливающаяся неустойчивость. Поэтому поведение - непрерывная задача. Живой организм должен быть постоянно активен.**

**Второй: поведение подчиняется принципу оптимальности – maxT**

**Третий: поведение управляется локальными правилами принятия решения на основе эмоциональных механизмов сравнения и выбора вариантов.**

**Четвертый: для решения задачи поведения необходимы умозрительное моделирование и предвидение.**

**Пятый: мозг решает многоэкстремальные задачи поведения за счет построения многоуровневой иерархической модели проблемной среды. С использованием этой модели в процессе мышления ищутся оцениваются и сравниваются обобщенные и укрупненные одношаговые переходы между текущей и целевой ситуациями. (Этот вопрос еще будет подробно рассматриваться).**

## **Лекция 6. Особенности живого восприятия**

Мы хотим придти к пониманию феномена мышления, идя от задач поведения и восприятия, т.е. от задач, для решения которых возник и эволюционно развивался мозг.

В предыдущих лекциях мы говорили о поведении. Теперь посмотрим, что дает для понимания феномена мышления задача восприятия. Мы рассмотрим некоторые принципы “интеллектуального” восприятия, конкретизирующиеся на примере решения задачи автоматического чтения рукописных символов. Практическая ориентация не привела, как это часто бывает, к упрощению и выхолащиванию проблемы восприятия. Наоборот, для получения работоспособного решения потребовалось введение “интеллектуальных” составляющих, ориентированных на распознавание “с пониманием”.

## **Распознавание образов**

С самого начала развития кибернетики машинное восприятие изображений чаще всего выбиралось для исследования и моделирования интеллекта и, в частности, таких очевидных составляющих мышления, как построение системы обобщенных знаний о среде и использования этих знаний в процессе принятия решений. Восприятие зрительной информации представлялось наиболее удобным для моделирования и, в то же время, наиболее практически значимым.

Сразу было ясно, что для полного решения задачи машинного зрительного восприятия необходимо “интеллектуальное” распознавание или распознавание “с пониманием”. Часто даже пытались сводить мышление к восприятию попросту ставя между ними знак тождества. В дальнейшем мы увидим, что мышление и восприятие

неразрывно связаны, но это далеко не одно и то же. Поэтому исследования живого восприятия ( в первую очередь зрительного) безусловно полезны для понимания процесса мышления, но проблему в целом далеко не решают. В то же время, практическая ориентация работ в области автоматического анализа зрительной информации, стремление к технической реализуемости привели к серьезной трансформации проблемы. Оказалось практически почти вынужденным упрощение рассмотрения процесса восприятия путем сведения его к классификации по признакам простых объектов, рассматриваемых по отдельности. Это направление стало называться “Распознавание образов”.

Распознавание образов к направлению Искусственный интеллект (ИИ) чаще всего не стали относить, поскольку в отличие от задач ИИ в распознавании образов появился хорошо разработанный математический аппарат и для не очень сложных объектов, оказалось возможным строить практически работающие системы распознавания (классификации). В результате традиционное распознавание образов, с одной стороны, не решает задачу машинного анализа сложных изображений и, с другой стороны, не является серьезным инструментом для моделирования интеллекта. Рассмотрим связанные с этим вопросы более подробно..

Для любого распознавания нужны эталоны или модели классов распознаваемых объектов. Классификация методов распознавания возможна по типам используемых эталонов или, что почти то же самое, по способу представления объектов на входе распознающей системы. В большинстве систем распознавания изображений обычно применяются растровый, признаковый или структурный методы.

Растровому подходу соответствуют эталоны, являющиеся изображениями, либо какими-то препаратами изображений. При распознавании представленное в виде точечного раstra входное изображение сопоставляется точка-в-точку со всеми эталонными и определяется с каким из эталонов изображение совпадает лучше, например, имеет больше общих точек. Входное изображение и эталонное должны быть одного размера и одной ориентации. Например, в так называемых multifold-распознавателях печатного текста это достигается построением разных эталонов не только для разных шрифтов, но и для разных размеров символов (кеглей) в пределах одного шрифта. Распознавание таким способом рукописных символов невозможно ввиду их слишком большой вариабельности по форме, размеру и ориентации.

Возможен также вариант использования растрового распознавания с приведением входного изображения к стандартным размерам и ориентации. В этом случае распознавание рукописных символов растровым методом становится возможным после кластеризации каждого распознаваемого класса и создании отдельного растрового эталона для каждого кластера.

В общем случае, получение инвариантности по отношению к размерам форме и ориентации распознаваемых по растру объектов является сложной, а часто и неразрешимой проблемой. Другую проблему порождает необходимость выделения из изображения его фрагмента, относящегося к отдельному объекту. Эта проблема является общей для всех классических методов распознавания образов.

В подавляющем большинстве систем распознавания и, в частности, в существующих omnifold системах оптического чтения (OCR), основным является признаковый метод. При признаковом подходе эталоны строятся с использованием выделяемых на изображении признаков. Изображение на входе распознающей системы представляется вектором признаков. В качестве признаков может рассматриваться все что угодно – любые характеристики распознаваемых объектов. Признаки должны быть инвариантны к ориентации, размеру и вариациям формы объектов. Желательно также, чтобы вектора признаков, относящиеся к разным объектам одного класса, принадлежали выпуклой компактной области пространства признаков. Пространство признаков должно быть фиксировано и одинаково для всех распознаваемых объектов. Алфавит признаков придумывается разработчиком системы. Качество распознавания во многом зависит от того, насколько удачно придуман алфавит признаков. Какого либо общего способа автоматического построения оптимального алфавита признаков не существует.

Распознавание состоит в априорном получении полного вектора признаков для любого выделенного на изображении отдельного распознаваемого объекта, и лишь затем в определении того, какому из эталонов этот вектор соответствует. Эталоны чаще всего строятся как статистические, либо как геометрические объекты. В первом случае обучение может состоять, например, в получении матрицы частот появления каждого признака в каждом классе объектов, а распознавание в определении вероятностей принадлежности вектора признаков каждому из эталонов.

При геометрическом подходе результатом обучения чаще всего является разбиение пространства признаков на области, соответствующие разным классам распознаваемых объектов, а распознавание состоит в определении того в какую из этих областей попадает соответствующий распознаваемому объекту входной вектор признаков. Затруднения при отнесении входного вектора признаков к какой либо области могут возникать в случае пересечения областей, а также, если области, соответствующие отдельным распознаваемым классам не выпуклы и так расположены в пространстве признаков, что распознаваемый класс от других классов просто, например одной гиперплоскостью, не отделяется. Эти проблемы решаются чаще всего эвристически, например, за счет вычисления и сравнения расстояний (не обязательно евклидовых) в пространстве признаков от экзаменуемого объекта до центров тяжести подмножеств обучающей выборки, соответствующих разным классам. Возможны и более

радикальные меры, например, изменение алфавита признаков или кластеризация обучающей выборки, или то и другое одновременно.

Структурному подходу соответствуют эталонные описания, строящиеся в терминах структурных частей объектов и пространственных отношений между ними. Структурные элементы выделяются, как правило, на контуре или “скелете” объекта. Чаще всего, структурное описание может быть представлено графом, включающем структурные элементы и отношения между ними. При распознавании строится структурное описание входного объекта. Это описание сопоставляется со всеми структурными эталонами, например, ищется изоморфизм графов.

Растровый и структурный методы иногда сводят к признаковому подходу, рассматривая в первом случае в качестве признаков точки изображения, а во втором, структурные элементы и отношения между ними. Сразу заметим, что между этими методами есть очень важное принципиальное различие. Растровый метод обладает свойством целостности. Структурный метод может обладать свойством целостности. Признаковый метод свойством целостности не обладает.

Что такое целостность и какую роль она играет при восприятии.

Классическое распознавание образов обычно организуется как последовательный процесс, разворачивающийся снизу-вверх (от изображения к пониманию) при отсутствии управления восприятием с верхних понятийных уровней. Этапу распознавания предшествует этап получения априорного описания изображения. Операции выделения элементов этого описания, например, признаков, или структурных элементов выполняются на изображении локально, части изображения получают независимую интерпретацию, то есть отсутствует целостное восприятие, что, в общем случае, может приводить к ошибкам – рассматриваемый изолированно фрагмент изображения часто можно интерпретировать совершенно по-разному в зависимости от гипотезы восприятия, т.е. от того, какой целостный объект предполагается увидеть.

Во-вторых, традиционные подходы ориентированы на распознавание (классификацию) объектов, рассматриваемых по отдельности. Этапу собственно распознавания должен предшествовать этап сегментации (разбиения) изображения на части, соответствующие изображениям отдельных распознаваемых объектов. Методы априорной сегментации обычно используют специфические свойства входного изображения. Общего решения задачи предварительной сегментации не существует. За исключением самых простых случаев, критерий разделения не может быть сформулирован в терминах локальных свойств самого изображения, т.е. до его распознавания.

Строчный, даже рукописный текст не является самым сложным случаем, но и для таких изображений, выделение строк, слов и отдельных символов в словах может оказаться серьезной проблемой. Практическое решение этой проблемы часто основывается на переборе вариантов сегментации, что совершенно не похоже на то, что делает мозг человека или животного в процессе целостного целенаправленного зрительного восприятия. Вспомним Сеченовское – мы не слышим и видим, а слушаем и смотрим. Для такого активного восприятия необходимы целостные представления объектов всех уровней – от отдельных частей до полных сцен и интерпретация частей только в составе целого.

. Таким образом, недостатки большинства традиционных подходов и, в первую очередь, признакового подхода, это отсутствие целостности восприятия, отсутствие целенаправленности и последовательная однонаправленная организация процесса снизу-вверх, или от изображения к “пониманию”.

Распознавание возможно также с использованием окутанного чуть ли не мистическим туманом искусственных или формальных распознающих нейронных сетей (РНС). Иногда их рассматривают даже как какой-то аналог мозга. В последнее время в текстах просто пишут “нейронные сети”, опуская прилагательные искусственный или формальный. На самом деле, РНС – это чаще всего просто признаковый классификатор, строящий разделяющие гиперплоскости в пространстве признаков.

Используемый в этих сетях формальный нейрон – это сумматор с пороговым элементом, подсчитывающий сумму произведений значений признаков на некоторые коэффициенты, являющиеся ни чем иным, как коэффициентами уравнения разделяющей гиперплоскости в пространстве признаков. Если сумма меньше порога, то вектор признаков находится по одну сторону от разделяющей плоскости, если больше – по другую. Вот и все. Кроме построения разделяющих гиперплоскостей и классификации по признакам никаких чудес.

Введение в формальном нейроне вместо порогового скачка от -1 к 1 плавного (дифференцируемого), чаще всего сигма-образного перехода ничего принципиально не меняет, а лишь позволяет использовать градиентные алгоритмы обучения сети, то есть нахождения коэффициентов в уравнениях разделяющих плоскостей, и делать “размазывание” разделяющей границы, присваивая результату распознавания, то есть работе формального нейрона вблизи границы, оценку, например, в диапазоне от 0 до 1. Эта оценка, в определенной степени может отражать “уверенность” системы в отнесении входного вектора к той или иной из разделяемых областей пространства признаков. В то же время, эта оценка, строго говоря, не является ни вероятностью, ни расстоянием до разделяющей плоскости.

Сеть из формальных нейронов может также аппроксимировать плоскостями нелинейные разделяющие поверхности и объединять по результату несвязанные области пространства признаков. Это и делается в многослойных сетях.

Во всех случаях ПРНС – это признаковый классификатор, строящий разделяющие гиперплоскости и выделяющий области в фиксированном пространстве признаков (характеристик). Никаких других задач ПРНС решать не может, причем задачу распознавания ПРНС решает не лучше обычных признаковых распознавателей, использующих аналитические методы.

Кроме того, помимо признаковых распознавателей на формальных нейронах могут строиться растровые или ансамблевые распознаватели. В этом случае сохраняются все отмеченные недостатки растровых распознавателей. Правда, могут быть и некоторые преимущества, о которых мы еще будем говорить в дальнейшем.

Во избежание недоразумений, следует заметить, что на формальных нейронах, в принципе, можно построить универсальный компьютер, с использованием как разделяющих плоскостей в пространстве переменных, так и легко реализуемых на формальных нейронах логических функций И, ИЛИ и НЕ, однако, таких компьютеров никто не строит и обсуждение связанных с этим вопросов выходит за рамки рассматриваемых проблем. Нейрокомпьютерами обычно называют либо просто нейронный распознаватель, либо специальные системы, решающие задачи, близкие распознаванию образов и фактически использующие распознавание на основе построения разделяющих гиперплоскостей в пространстве признаков, либо на основе сравнения растра с эталоном.

Выше уже отмечалось, что для моделирования мышления очень важно, а может быть, и необходимо, понять, как работают нейронные механизмы живого мозга. В связи с этим возникает вопрос: а не являются ли формальные распознающие нейронные сети если и не решением проблемы моделирования нейронных механизмов мозга, то хотя бы важным шагом в этом направлении. Представляется, что, к сожалению, ответ должен быть отрицательным. В отличие от активной живой нейронной сети РНС это пассивный признаковый или растровый классификатор со всеми недостатками традиционных классификаторов. Аргументы, на основании которых сделан этот вывод, более подробно мы рассмотрим в дальнейшем.

Итак, традиционные в первую очередь признаковые системы распознавания, основывающиеся на последовательной организации процесса распознавания и классификации объектов, рассматриваемых по отдельности, эффективно решать задачи восприятия сложной зрительной информации не могут, главным образом, по причине отсутствия целостности и целенаправленности восприятия, отсутствия целостности в описаниях (эталонах) распознаваемых объектов и последовательной организации процесса распознавания. По этой же причине такие системы распознавания образов мало что дают для понимания живого зрительного восприятия и процесса мышления.

## **Психология машинного зрения**

В этом разделе ваше внимание будет привлечено к некоторым аспектам живого восприятия. Предполагается, что для построения практически эффективной системы машинного восприятия недостаточно изобретение изолированных алгоритмов классификации отдельных объектов. Необходима *организация* целостного, целенаправленного, управляемого контекстом *процесса* восприятия, то есть, восприятия “с пониманием”.

Название раздела рассчитано на то, чтобы рождать у специалистов ассоциации с популярной в 80-е годы книгой “Психология машинного зрения”. Интерес к этому сборнику был связан, в первую очередь, с именами Патрика Уинстона и Марвина Минского. Их работы, как впрочем и большинство других работ сборника, объединяет убежденность в том, что эффективное машинное восприятие с необходимостью должно и может обладать многими свойствами живого восприятия и, в первую очередь, целенаправленностью.

Работы М.Минского и П.Уинстона стали определенным шагом от традиционного распознавания образов к созданию теории машинного зрительного восприятия сложных изображений. Минский выдвинул теорию фреймов, содержащую предположения о том, как строится и функционирует иерархическая структурная зрительная модель мира. Модель Минского, к сожалению, не работала и работать не могла, поскольку она была в значительной степени качественной. Тем не менее, пользу для общего понимания процесса восприятия среды с использованием идей структурности и целостности эта модель безусловно принесла. К сожалению, в экспериментальном плане ее никто серьезно не развивал, а в теоретическом плане ее не развивал и сам Минский.

П.Уинстон обратил внимание на необходимость реализации целенаправленного процесса машинного восприятия. Цель должна управлять работой всех процедур, в том числе и процедур нижнего уровня, то есть процедур предварительной обработки и выделения признаков. Должна иметься возможность на любой стадии процесса в зависимости от получаемого результата возвращаться к его началу для уточнения результатов работы процедур предшествующих уровней. Такие системы Уинстон предложил называть гетерархическими. У Уинстона также как и у Минского до решения практических задач дело не дошло, хотя в 80-е г.г. прошлого века вычислительные мощности больших вычислительных машин позволяли начать решать подобные задачи.

Безусловный интерес для теории восприятия представляет высказанный М.М.Бонгардом принцип имитации. По этому принципу, относящемуся еще к концу 60-х годов, в основе задачи узнавания (расознавания) должна лежать умозрительная имитация устройства, строящего объекты распознавания. От принципа имитации уже напрашивается переход к рассматриваемым ниже принципам структурности, целостности и отображаемости.

Представляется, что общее решение задачи машинного восприятия изображений должно основываться на организации процесса, с включением таких составляющих, как целостность восприятия, целенаправленность и предвидение (гипотеза), то есть, моментов, характеризующих наше сегодняшнее представление о процессе зрительного восприятия человека. Естественно, что кроме этого должны максимально использоваться контекст восприятия и знания о среде (максимально полная семантическая модель среды).

Важнейшим аспектом восприятия является предвидение на основе иерархичной модели мира и многоуровневого процесса восприятия. В знакомой среде и знакомых ситуациях восприятие идет на уровнях обобщений (общее-частное) и укрупнений (целое-часть) и состоит в подтверждении предвидения на этих уровнях. Обращение к уровню детального восприятия происходит только по мере поведенческой необходимости или при рассогласовании предвидения и реального входа.

Мы уже говорили об известном в физиологии принципе акцептора действия. Напомним, что этот принцип состоит в том, что в нервной системе всегда (непрерывно) при любом действии стоит модель ожидаемой обратной афферентации, поступающей от результата действия. Рассогласование модели и реальной обратной афферентации вызывает ориентировочно-исследовательскую реакцию. (В качестве примера можно вспомнить ощущения, возникающие при входе на неподвижный эскалатор метро). Все это очень похоже на то, что происходит и при "чистом" восприятии, то есть при восприятии информации от среды не связанным с действием. Может быть, полезным было бы введение термина "акцептор восприятия".

В акте зрительного восприятия на всех этапах и всех уровнях восприятия тесно переплетаются и взаимодействуют два процесса, которые можно обозначить как процессы «сверху-вниз» (от понимания к изображению) и «снизу-вверх» (от изображения к пониманию).

Важнейший момент восприятия - это формирование гипотезы о содержании изображения. Гипотеза возникает при взаимодействии процесса «сверху-вниз», разворачивающегося на основе модели среды, модели текущей ситуации и текущего результата восприятия, и процесса «снизу-вверх», основанного на непосредственном грубом в первую очередь признаковом восприятии.

Далее происходит подтверждение гипотезы или уточнение восприятия. На этом этапе также взаимодействуют оба процесса - операции над информацией из модели и операции на изображении. При этом в рамках текущей гипотезы с использованием модели среды и информации о контексте восприятия осуществляется целенаправленный поиск, включающий сегментацию изображения на искомые в соответствии с гипотезой части, и совместную интерпретацию выделяемых частей.

Важнейшим аспектом восприятия является его целостность: результаты локальных операций интерпретируются только совместно в процессе интерпретации целостных фрагментов и всего изображения в целом. Используемая при восприятии целостная модель должна быть структурной, полной и отображаемой. Последнее означает, что должна иметься возможность мысленно представить себе объект по его модели.

Используемая при восприятии модель проблемной среды должна включать иерархию целостных представлений. Применительно к задаче автоматического восприятия изображений можно сказать следующее. Помимо совместной интерпретации элементов изображения в составе распознаваемых объектов, принцип целостности восприятия предполагает также интерпретацию самих распознаваемых объектов в составе более крупных целостных образований - конструкций, отображающих те взаимосвязи из внешней задачи, в которых участвуют распознаваемые объекты. Эти взаимосвязи образуют внешний контекст распознавания. Использование внешнего контекста распознавания позволяет не только правильно интерпретировать те объекты, изображения которых допускают при их отдельном восприятии неоднозначную интерпретацию, но и повысить надежность распознавания всех объектов, задействованных в той или иной семантической конструкции, за счет их целенаправленной и совместной интерпретации.

Таким образом, в соответствии с принципом целостности восприятия, в общем случае, можно говорить об иерархии уровней интерпретации элементов распознаваемого изображения: от интерпретации в составе самих распознаваемых объектов, до интерпретации в составе наиболее крупных семантических конструкций, представленных на данном изображении.

Очевидно, что в общем случае, чем выше уровень интерпретации, то есть чем крупнее те целостные образования и конструкции, задающие внешний контекст, в составе которых осуществляется интерпретация тех или иных элементов входного изображения, тем выше надежность распознавания этого изображения.

Такая организация процесса распознавания в системе машинного зрения необходима, если мы хотим получить действительно эффективное решение сложных практических задач. Естественно, машинное зрительное восприятие не может пока еще соревноваться со зрительным восприятием человека. Главная причина этого состоит в том, что мы не умеем строить и использовать полную машинную семантическую модель среды восприятия. Однако, для повышения эффективности систем машинного зрения и, в частности, систем машинного чтения, отмеченные выше принципы двунаправленности (от изображения к модели и от модели к изображению) предвидения (гипотеза), целостности, целенаправленности и максимального использования информации о проблемной среде в определенной степени реализовать не только можно, но и необходимо.

Эти принципы настолько, насколько это оказалось возможным, реализованы в пакете программ Графит, в программах FineReader-рукопись и FormReader - для распознавания рукописных символов и, частично, в

программе FineReader для распознавания печатных текстов. Пакет Графит был разработан в НИЦЭВТе в 80е годы, программы FineReader и FormReader в ABBYY.

## **Лекция 7. Принципы целостного целенаправленного распознавания и их реализация в программах ABBYY FineReader-рукопись и FormReader**

### **Задача**

Частным случаем задачи машинного зрительного восприятия сложных изображений является автоматическое чтение рукописных и печатных текстов.

Практическое значение этой задачи определяется необходимостью представления, хранения и использования в электронном виде чрезвычайно большого количества накопленной и вновь создаваемой текстовой информации. Кроме того, большое значение имеет оперативный ввод в информационные и управляющие системы информации с машиночитаемых бланков, содержащих как напечатанные, так и рукописные тексты, например, ввод банковских платежных документов, бланков заказов, налоговых деклараций, анкет и т.п. Практическое значение имеют и многочисленные другие задачи автоматического анализа изображений.

В системах ABBYY FineReader-рукопись и FormReader при распознавании рукописных текстов используются структурный, растровый, признаковый, дифференциальный и лингвистический уровни. Основным является структурный уровень.

### **Общие принципы**

Главный принцип, положенный в основу построения структурного уровня, это принцип целостности. В соответствии с этим принципом, распознаваемый объект рассматривается как целое, состоящее из структурных частей, связанных между собой пространственными отношениями. Фрагмент изображения интерпретируется как конкретный целостный объект, в нашем случае – рукописный символ, если на изображении присутствуют все структурные части этого объекта, и эти части находятся в определяемых структурно-метрическим описанием объекта отношениях. С другой стороны, части распознаваемого объекта получают интерпретацию только в составе объединяющего их целого. Распознаваемый объект (целое) сначала присутствует в виде гипотезы, и это позволяет целенаправленно выделять на изображении и интерпретировать образующие его структурные части.

Гипотезы формируются следующим образом. Признаковый и нормируемый растровый уровень порождают список гипотез, упорядоченный по вероятности их подтверждения. В формировании списка проверяемых гипотез участвуют не только растровые эталоны и выделяемые на изображении локальные признаки, но и текущий результат распознавания, а также информация из моделей, определяемых проблемной средой, то есть информацией из внешней задачи.

Требования к признаковому и растровому уровням на этом этапе состоят лишь в том, что истинная гипотеза должна присутствовать в формируемом списке и находиться в нем максимально близко к началу. Кроме того, при сильно нарушенной структуре символа результату, полученному на структурном уровне может соответствовать низкая уверенность или отказ от распознавания. В этом случае предварительным результатом распознавания является упорядоченный список гипотез, полученных на признаковом и растровом уровнях.

### **Процесс**

Проверка на структурном уровне выдвинутых гипотез сводится к целенаправленному поиску непосредственно на входном изображении совокупности элементов, удовлетворяющих структурному описанию (модели) данного класса распознаваемых объектов. При этом процесс проверки гипотез характеризуется следующими особенностями.

Сегментация изображения на отдельные элементы, а также интерпретация этих элементов должны осуществляться непосредственно в ходе распознавания (этапы предварительной сегментации и описания отсутствуют) и проводятся исключительно в рамках целостного образования – проверяемой гипотезы.

Сегментация изображения и интерпретация получаемых элементов осуществляется целенаправленно и управляется проверяемой гипотезой и текущим результатом, причем все результаты интерпретируются совместно.

После окончания проверки очередной гипотезы, все полученные в ходе ее проверки результаты, связанные с выделением структурных элементов, аннулируются, поскольку они имеют смысл только по отношению к этой гипотезе. Одновременно с проверкой гипотезы вычисляется интегральная оценка качества ее подтверждения.

В полной схеме автоматического зрительного восприятия, ориентированного на анализ сложных изображений, процессы “снизу-вверх” и “сверху вниз” должны разворачиваться в диапазоне от исходного непрепарированного изображения до наиболее высокого уровня, отражающего полную информацию о проблемной среде и семантике решаемой задачи. В описываемых программах этот диапазон несколько сужен снизу и значительно сверху.

Нижним уровнем структурного анализа в программах чтения рукописных символов являются линии изображения, аппроксимированные отрезками прямых. Верхний уровень – это, чаще всего, достаточно



ограниченная информация о простых семантических конструкциях внешней задачи, а также лингвистическая информация (словарь). Информация из внешней задачи в наибольшей степени используется при распознавании и вводе форм, например таких, как анкеты, налоговые декларации или банковские платежные документы.

## Модель

Структурные описания задают обобщенные метрические характеристики структурных частей объекта и пространственные отношения между ними. Очень приближенно, без какой либо математической нагрузки, структурное описание можно интерпретировать как ориентированный, раскрашенный, мультисвязный граф с петлями.

Целостное, структурное описание объекта (знака), состоит из обобщенных описаний структурных частей объекта и обобщенных пространственных отношений между ними. Обычно разделяют обобщения и укрупнения как что то независимое, при этом в проблематике ИИ чаще рассматриваются обобщения (индукция – дедукция). Укрупнения, т.е. взаимодействия между частями и целым (анализ - синтез) рассматриваются реже.

В рассматриваемой системе описание знака, является укрупнением, поскольку связывает в единое целое образующие его части. В то же время, описание знака является обобщением, соответствующем не отдельному объекту, а всем объектам данного класса. Это достигается тем, что описание определяет не конкретные значения, а некоторую многомерную область допустимых значений характеристик частей и отношений между ними.

Назовем метрические характеристики структурных элементов и отношения между ними структурными параметрами. Описания разных классов изображений независимы и, в общем случае, имеют разное число и состав структурных параметров. Тем не менее, интегральные оценки качества подтверждения разных гипотез должны быть сопоставимы. Сопоставимость достигается тем, что оценки подтверждения гипотез имеют общий физический смысл.

Пусть некоторому классу изображений соответствуют структурные параметры  $(X_1, X_2, \dots, X_n)$ . Конкретное изображение класса отображается точкой в n-мерной системе координат –  $X(X_1, X_2, \dots, X_n)$ .

Для всех параметров известны 'идеальные' и предельно допустимые значения. Если определить параметры оценки изображения  $x_i$  как разность между текущим и идеальным значением, то все  $x_i \geq 0$ .

Входящие в описания класса параметры независимы. Это не является, как может показаться, недостатком описаний, поскольку в их состав могут входить любые отношения между параметрами, в том числе и отношения между отношениями.

С учетом введенных ограничений, также как мы уже делали ранее, граница области допустимых значений параметров класса - D задается выпуклой поверхностью  $F(X) = 0$ . Назовем  $x_{imax}$  разность между максимальным и идеальным значением i-го параметра. Поскольку все предельно допустимые и идеальные значения параметров  $i$  независимы и известны, область D – это n-мерный параллелепипед и уравнение границы области допустимых значений параметров:

$$F(X) = \prod_i (x_{imax} - x_i) = 0 \text{ ИЛИ } \prod_i (1 - x_i / x_{imax}) = 0$$

Конкретное изображение отображается в пространстве  $\{X\}$  точкой

$X(x_1, x_2, \dots, x_n)$ . Качество изображения тем выше, чем ближе эта точка к началу координат и соответственно, дальше от границы области допустимых значений - D.

Так же как в модели поведения введем оценочную функцию

$$Z(X) = F(0) - F(X)$$

И с учетом уравнения границы области допустимых значений параметров

$$Z(X) = 1 - \prod_i (1 - x_i / x_{imax})$$

Поскольку функция  $F(X)$  выпукла, функция  $Z(X)$  монотонна.

При  $X = 0$ , т.е. когда все параметры изображения идеальны, функция  $Z(X) = 1$ . В любой точке границы области допустимых значений параметров функция принимает максимальное значение  $Z(X) = 0$ .

Величина, оценивающая качество изображения как представителя конкретного класса

$$W = 1 - Z(X) = \prod_i (1 - x_i / x_{imax})$$

Эта же величина является интегральной оценкой качества подтверждения гипотезы.

Интегральная оценка используется для принятия решения на структурном уровне в том случае, когда изображение допускает неоднозначную структурную интерпретацию. Кроме того, неоднозначность может сниматься парными дифференциальными классификаторами, которые строятся для путающихся пар, а также лингвистическим уровнем (словарь) или контекстом, идущим от внешней задачи.

Используемые при подтверждении гипотез структурные описания должны быть отображаемыми в объекты своего класса и только в объекты своего класса. Принципиальная отображаемость является необходимым критерием целостности. С помощью отображения описаний, то есть получения изображений для точек  $X(X_1,$

X<sub>2</sub>...X<sub>n</sub>), лежащих вблизи границы области допустимых значений параметров, разработчик мог бы визуально контролировать качество созданных им описаний.

Построение и отладка структурных описаний – это трудоемкий процесс, не поддающийся пока автоматизации. Также не удается для общего случая создать алгоритмы построения изображений по их (принципиально отображаемым) структурным описаниям. Обе названные проблемы сродни общим проблемам описания функциональных взаимодействий между частями и целым.

Особенностью системы FineReader-рукопись, является то, что в ней нет фиксированного набора отношений, а используется “конструктор”, позволяющий в процессе создания описаний строить практически любые отношения и создан аппарат, дающий возможность работать с любыми отношениями.

## **Предварительная обработка изображений**

Структурный уровень системы FineReader-рукопись и в настоящее время системы FormReader работает на векторном представлении изображения, полученном после его утоньшения. Утоньшение и векторизация изображения делаются на этапе, предшествующим распознаванию. Это нарушает основной исходный принцип – ничего не делать до распознавания и вне распознавания, то есть без “понимания”. Действительно, в результате “слепого” утоньшения и векторизации может получаться результат в определенной (иногда значительной) степени не совпадающий по форме и структуре с исходным изображением.

В описываемой системе это частично компенсируется возможностью при необходимости обращаться непосредственно к исходному растровому изображению. Кроме того, перспективным представляется рассмотрение вопросов о разработке алгоритмов целенаправленной, то есть ориентированной на гипотезу векторизации, либо целенаправленного выделения структурных элементов непосредственно на растровом изображении.

Максимально развитый верхний уровень, в пределе построение полных семантических моделей проблемной среды, является общей проблемой для систем моделирования интеллекта.

Априорная векторизация изображения это не единственный потенциальный источник ошибок. В последнее время в статьях и докладах часто используется придуманное кем-то красивое наукообразное выражение “приведение изображения к виду, удобному для распознавания”. Эти слова означают то же самое, что и давно употребляющийся термин - предварительная обработка. Однако, в предварительной обработке речь должна идти не об удобстве, а о необходимости или, как минимум, о целесообразности, т.е. об улучшении качества работы системы. При этом всегда нужно помнить, что априорная предварительная обработка изображения, состоящая в каком-то его изменении еще до распознавания, может быть источником ошибок.

К предварительной обработке относится все, что делается с изображением до его распознавания: считывание, бинаризация, фрагментация, фильтрация, векторизация, отделение объектов от фона и т.п. Для программ, распознающих тексты, это часто еще выравнивание, выделение строк, выделение слов, выделение отдельных символов.

Принятый в описываемой системе чтения рукописных символов общий принцип состоит в том, что “слепые” априорные операции над изображением нужно делать только по необходимому минимуму, поскольку они могут приводить к необратимой потере информации и ошибкам, либо нужно иметь возможность в зависимости от текущего результата возвращаться к исходному изображению.

Считывание изображения происходит с каким то пространственным разрешением. Обычно это около 10 точек на миллиметр. При распознавании маленьких объектов этого разрешения может быть недостаточно и нужно иметь возможность в зависимости от каких то результатов получать новое изображение с большим пространственным разрешением.

Бинаризация состоит в преобразовании многоградационного по яркости изображения в двоичное, то есть такое, в котором точки фона имеют значение “0”, а точки изображения “1”. Преобразование по фиксированному порогу яркости обычно дает плохой результат: на изображении появляются разрывы и заливки в таких местах, где человек воспринимает изображение правильно.

Для улучшения бинаризации применяется переменный порог, выставляемый для каждой точки по результатам анализа свойств изображения в некоторой локальной окрестности. Однако и это может приводить к ошибкам, которые может компенсировать возможность при необходимости обращения к исходному растровому изображению.

Фильтрация – это очистка изображения от помех – лишних точек и каверн (дырочек), и, в некоторых случаях, сглаживание контура. Понятно, что в процессе фильтрации может портиться изображение, например удаляться точки над ё или і. Поэтому, фильтрация должна быть осторожной и, как всегда, должна иметься возможность обращения к исходному, не испорченному изображению.

Априорное отделение объекта от других объектов и фона (для текстов - выделение строк, слов и отдельных символов) может быть неоднозначным. В некоторых случаях правильный вариант может попросту отсутствовать. Для описанного выше целенаправленного структурного распознавания выделение отдельных распознаваемых объектов в принципе не нужно. Была бы правильная гипотеза. в списке гипотез, при этом не обязательно на первом месте. Для работы признакового и растрового уровней выделение отдельных объектов нужно. Неоднозначность на этих уровнях может компенсироваться перебором вариантов.

Фрагментация – это выделение на изображении отдельных, последовательно читаемых блоков: заголовков, столбцов, подписей к рисункам, текста в графах таблицы или формы, отделение рисунков и графики от текста и т.п. Уменьшение вероятности ошибки при фрагментации может достигаться за счет использования на этом уровне целостных структурных описаний типов документов и организации целенаправленного процесса их интерпретации. В описываемой системе это в определенной степени реализовано для анализа форм, например, таких, как банковские платежные документы.

Одним из источников гипотез для структурного распознавания является предварительное распознавание на растровом уровне. Для растрового распознавания необходима предварительная нормализация размеров и ориентации растрового изображения. Кроме того, важно, чтобы обучение и распознавание проводились не только при одинаковых размерах и ориентации изображений объектов, но и при одной (стандартной) толщине линий. Все это может быть отнесено к предварительной обработке

Определение толщины линий можно делать различными способами. Если выполнена векторизация изображения, то среднюю толщину линий можно определить разделив число черных точек (площадь) на длину векторного изображения. Точность этого метода в значительной степени зависит от качества векторизации.

Если изображение состоит из линий (например, графика или текст), то вне зависимости от того, выполнялась или нет векторизация, толщину линий можно вычислять по формуле:

$$T=S/(P/2 - T+2)$$

Где T – толщина линий,

S – площадь изображения (число черных точек),

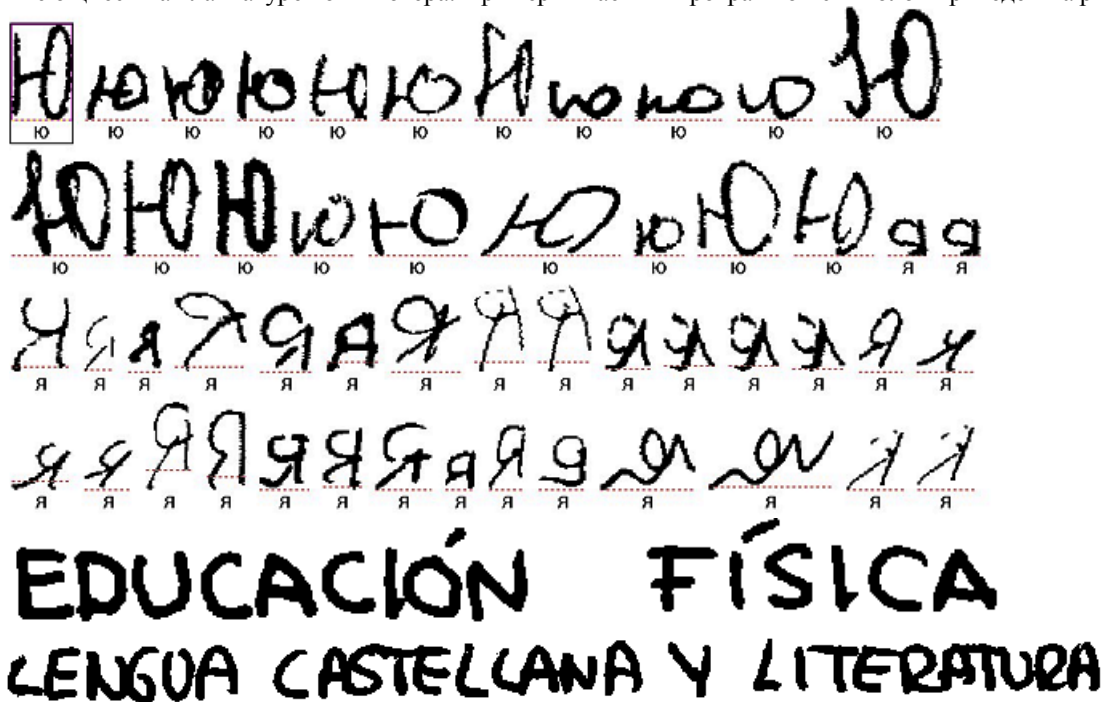
P - число точек контура изображения.

Длина контура в пикселах может определяться по числу переходов с черного на белое по горизонтали и по вертикали. При RLE – представлении изображения (кодирование сериями) длина контура может определяться по числу горизонтальных и вертикальных RLE-отрезков.

Формула дает удовлетворительную точность, особенно если была выполнена предварительная фильтрация контура.

### Практические результаты

Входящая в FormReader –программа чтения рукописных текстов была выпущена в 1998 году одновременно с системой ABBYY FineReader 4.0. В настоящее время разработана новая версия, которая входит в состав системы FormReader 6.0. Эта программа может читать все рукописные строчные и заглавные Handprint символы, имеющиеся на клавиатуре компьютера. Пример читаемых программой символов приведен на рис.



Допускаются ограниченные соприкосновения символов между собой и с графическими линиями. Читаются символы, входящие в алфавиты 15-ти языков и при этом учитываются особенности национального письма. Основное практическое применение – распознавание и ввод информации с машиночитаемых бланков.

Достоверность правильного распознавания зависит от качества распознаваемой информации. Для цифр это 99.6% – 99.8%. Для русских или английских букв с контекстной обработкой достоверность правильного распознавания в среднем превышает 99%.

Область возможного применения описанного подхода не ограничивается задачей автоматического анализа изображений, содержащих символическую и графическую информацию. Целостность и целенаправленность на основе предвидения важны при любом восприятии и тем более при восприятии более сложных изображений, например изображений трехмерных сцен. Эти же принципы важны и при восприятии информации незрительной модальности.

Возможность практической реализации подхода применительно к конкретной задаче зависит от нескольких факторов.

Во-первых, необходима возможность построения полных целостных структурно-метрических описаний классов объектов восприятия, включающих не только элементы целого, но и отношения между ними. Уже отмечалось, что принципиальная отображаемость это критерий целостности и полноты описаний зрительных образов. Описания должны быть принципиально отображаемыми в объекты своего и только своего класса. Каков критерий целостности и полноты описаний для незрительных, например речевых, образов? Повидимому, таким критерием может быть принципиальная воспроизводимость. Слово ‘принципиальная’ говорит о том, что в описаниях достаточно информации для отображения (воспроизведения), и что реальное отображение (воспроизведение) в принципе могло бы быть реализовано.

Во-вторых, для реализации целостного целенаправленного восприятия должна иметься возможность выделять структурные элементы и проверять отношения между ними непосредственно в среде восприятия.

И, наконец, желательно, чтобы в системе было максимально полное описание проблемной среды, используемое как на этапе формирования гипотез, так и на этапе их проверки. В пределе описание среды должно стать активной моделью среды. Связанные с этим вопросы подробно рассматриваются в последующих разделах.

## Лекция 8. **Организация**

Мы уже говорили о том, что для понимания феномена мышления важно определить разницу между живым и неживым и особенности задач, которые должны решаться живыми организмами для обеспечения своего существования. Представляется, что в этом плане имеют значение рассмотренные в предыдущих лекциях модель поведения на основе принципа оптимальности  $\max T$  и модель восприятия “с пониманием” на основе принципов целостности и целенаправленности. Как для решения задач поведения, так и для решения задач восприятия в мозге строится целостная семантическая модель проблемной среды, отображающая существующие в среде объекты, ситуации и процессы, организацию объектов и взаимосвязей между ними. Можно ли как-то охарактеризовать общие принципы организации существующих живых и неживых целостных систем (объектов)? Рассмотрим связанные с этим вопросы. Одним из наиболее важных в этом рассмотрении будет понятие целостности.

### **Целостность и системность.**

К понятию целостность можно идти от понятия система. В последнее время часто применяется словосочетание ‘системный подход’, причем, часто не вполне ясно, что это означает. Иногда шутят, что ‘системно’ значит попросту ‘хорошо’.

Теория систем как самостоятельная область науки несмотря на свое название еще не вышла на теоретический уровень. Более того, становление нового направления в науке начинается с появления нового предмета исследования и завершается определением адекватных предмету методов исследования. Теория систем имеет свой, хотя и довольно расплывчатый, предмет – систему. Эффективный метод исследования систем не создан. Попробуем прояснить для себя смысл термина система.

Разные исследователи трактуют понятие система по-разному. В то же время, давая определение системе, все в той или иной степени, прямо или косвенно затрагивают интересующее нас в первую очередь понятие **целостность**. Несмотря на некоторые различия в существующих многочисленных определениях понятия система, все эти определения появились не произвольно, а возникли в результате необходимости охарактеризовать появившуюся в науке новую проблематику. В чем же состоит эта новая проблематика?

Интересный анализ принадлежит Уорену Уиверу, который делит научные проблемы на проблемы линейных причинных рядов – проблемы двух переменных, проблемы большого числа слабо связанных переменных, которыми занимаются статистика и термодинамика, и проблемы большого числа сильно связанных переменных, требующие для своего решения создания новых понятийных средств и нового формального аппарата.

Для этих классов проблем Уивер вводит названия: проблемы организованной простоты, проблемы неорганизованной сложности и проблемы организованной сложности. Проблемы организованной сложности в

первую очередь возникают при исследовании биологических объектов и при исследовании сложных современных промышленных, экономических, социальных и военных организаций. Именно к проблемам организованной сложности многие авторы относят понятие “система”.

Определение Уивера не затрагивает в явном виде проблему соотношения частей и целого. Большинство авторов считают этот аспект основным и ставят его на первое место. Так Эшби говорит о том, что суть системного подхода состоит в изучении целого, не разделенного на части, поскольку разделение нарушает взаимодействие. Этот же аспект отражает наиболее распространенное определение: система – целостное множество взаимосвязанных элементов (Берталанфи). Приведенное определение даже в сочетании с определением Уивера дает не очень много, поскольку остается невыясненным, что такое целостность, и какая роль в образовании и поддержании целостности принадлежит связности элементов.

Важная сторона проблематики систем связана с взаимоотношением между целым и образующими его частями. Так, например, говорят о том, что теория систем обязана своим появлением как недостаточности механицизма, то есть объяснения целого через части, так и недостаточности организмического подхода, то есть невозможности объяснения частей через целое, и теория систем должна органически объединить оба названных подхода.

Другой стороной проблематики систем является организация. Выше уже упоминались выделяемые Уивером системы *организованной* сложности. Организация определяет взаимодействие между элементами системы и возникновение в результате этого новых качеств целого. Организация, другими словами внутреннее устройство системы, определяет и сохранение ее качеств во времени. .

Приведем в связи с этим высказывание А.Сент-Дьердьи : “Одним из основных принципов жизни является ‘организация’; мы понимаем под этим, что при объединении двух вещей рождается нечто новое, качества которого не аддитивны и не могут быть выражены через качества составляющих его компонент. Это относится ко всей гамме форм организации, к объединению электронов и ядер, образующих атом, к объединению атомов в молекулы, аминокислот в пептиды, пептидов в белки, белков и нуклеиновых кислот в нуклепротеиды и т.д.”.

С учетом типичных взглядов на проблематику системности признаки систем это:

- большое число составляющих элементов,
- связность элементов,
- целостность;
- многоуровневость – как минимум в системе есть уровень элементов (частей) и уровень целого,
- новые свойства целого, не сводимые к свойствам частей, но каким-то неизвестным образом свойствами частей определяемые,
- состав и свойства элементов каким-то неизвестным образом определяются свойствами целого,
- возникновение устойчивых структур и взаимоотношений (взаимодействий) между элементами, объединяющих элементы в целое,
- организация, как системобразующий интегральный фактор, обеспечивающий существование во времени целостного объекта как некоторой качественной определенности,
- синергия (взаимосодействие) как необходимое свойство, проявляющееся в первую очередь в функционировании живых целостных систем.

Нетрудно заметить, что перечисленными признаками обладает любой материальный, а может быть, и не материальный объект и системный подход есть просто методологический прием, требующий рассмотрения целого и составляющих его элементов в неразрывном единстве. Таким образом, слово система в рассматриваемом аспекте – лишнее. Его мог бы заменить термин целостностный объект. Однако в дальнейшем все же иногда будет использоваться и привычное слово система.

## **Организация и дезорганизация**

Целостный объект имеет организацию, определяющую его качества и существование во времени. Количественной оценке уровня разупорядочивания или дезорганизации соответствует величина энтропии. Согласно второму закону термодинамики энтропия, то есть разупорядочивание и дезорганизация, в замкнутых системах возрастает. Происходит диссипация энергии, т.е. переход энергии любого вида, или, так называемой, упорядоченной свободной энергии, в неупорядоченную тепловую.

На основе второго закона термодинамики прогнозируется печальный конец Вселенной – ее тепловая смерть, т.е. полная дезорганизация, выравнивание всех неоднородностей, хаос. Одновременно активно обсуждается якобы существующий термодинамический парадокс – с одной стороны, существование во Вселенной неоднородностей и организации, с другой стороны бесконечность времени и соответственно бесконечно длительный процесс разупорядочивания и увеличения энтропии. Парадокс решается разными способами. Один из наиболее известных и часто приводящихся - это теория флуктуаций Л.Больцмана.

Однако существует ли угроза тепловой смерти Вселенной и, соответственно, термодинамический парадокс? Во-первых, ни из чего не следует, что Вселенная является замкнутой системой, подчиняющейся в целом второму

закону термодинамики. Бесконечно не только время, но и пространство. Вселенных может быть много. Обмениваются ли вселенные материей, энергией и информацией не известно. Во-вторых, мы не очень четко представляем, что такое вечность и бесконечность, а соответственно, что такое время и пространство. В-третьих, во Вселенной действуют не объясненные пока еще наукой силы, к которым в первую очередь можно отнести гравитацию, электрические взаимодействия, а также сильные внутриатомные взаимодействия. Есть во Вселенной и другие, не объясненные пока еще физические явления, например, связанные с темной материей и темной энергией. Таким образом, у нас нет достаточных оснований для того, чтобы делать обоснованные заключения о судьбе Вселенной, основываясь на втором законе термодинамики.

Кроме того, если идут процессы дезорганизации, значит, организация и порядок существуют и как-то они возникли. Прежде чем обсуждать термодинамический парадокс нужно иметь какую-то точку зрения и на то, как организация или упорядочивание возникает. Как минимум нужно допускать саму возможность возникновения порядка, причем как на основе известных физических свойств мира, так, может быть, и на основе неизвестных. Один из распространенных вариантов это ссылка на бога. Однако, это абсолютно не обязательный вариант.

(Когда Лаплас изложил Наполеону свою концепцию образования планет, император спросил: «А как же бог?» Лаплас ответил: «Я не нуждаюсь в этой гипотезе»).

Будем считать, что на основе известных, а возможно и каких-то неизвестных нам материальных физических сил, во Вселенной идут процессы организации и упорядочивания, противоположные процессу возрастания энтропии. Одним из таких очевидных процессов организации и упорядочивания является возникновение и эволюция живых организмов. Процессом организации и упорядочивания является и развитие цивилизации или социальная эволюция.

Слово “организация” используется в двух разных смыслах. Во-первых, под организацией можно понимать процесс (от глагола организовывать), результатом которого является что-то – система, объект или что угодно, имеющее какую-то качественную определенность. Другими словами, можно сказать, что организация, понимаемая как процесс, это создание порядка или упорядочивание, т.е. антиэнтропийный процесс, противоположный процессу возрастания энтропии.

Помимо создания качественной определенности объекта результатом антиэнтропийного процесса должно быть возникновение свойств объектов, противодействующих тенденции возрастания энтропии, т.е. свойств, обеспечивающих сохранение качественной определенности, т.е. невозрастание энтропии.

Процессы упорядочивания мы будем называть антиэнтропийными процессами, или процессами создания и сохранения порядка. Результат антиэнтропийного процесса это образование упорядоченной системы (объекта), обладающей качественной определенностью и сохраняющей свою качественную определенность во времени. Энтропия (дезорганизация) в такой системе не возрастает, или возрастает медленно.

Антиэнтропийные процессы упорядочивания могут быть естественными и искусственными. Приблизительные представления о естественных антиэнтропийных процессах могут быть, например, такими: Большой взрыв, разлет материи, концентрация материи за счет гравитации в галактиках и звездах, термоядерные реакции в звездах и образование химических элементов, образование планет, образование веществ и отдельных устойчивых материальных объектов, возникновение и развитие жизни и разума. К этому можно добавить еще и социальную эволюцию, приводящую к возникновению и развитию цивилизации, т.е. социальных отношений, науки, технологии, искусства.

Искусственные антиэнтропийные процессы это специальное, в том числе и “разумное”, создание разнообразных устойчивых объектов живыми организмами. Слово разумное взято в кавычки потому, что хотя в создании некоторых (очень многих) искусственных объектов и участвует разум, получающийся результат не всегда разумен с общечеловеческой точки зрения.

Результатом естественного антиэнтропийного процесса может быть создание как устойчивых равновесных объектов (систем), изменения в которых описываются классической равновесной термодинамикой, так и создание устойчивых неравновесных систем. Мы в дальнейшем будем говорить об активных и пассивных неравновесных системах.

Пассивные неравновесные это диссипативные системы, изменения в которых описываются теорией неравновесной термодинамикой Пригожина [N]. Активные неравновесные системы это живые организмы, к которым относится принцип устойчивого неравновесия Бауэра [4].

Процесс естественного возникновения неравновесных систем принято называть словом *самоорганизация*. Это слово несет на себе отпечаток какой-то собственной активности, чего-то связанного с кибернетикой или с принципами функционирования живого. Это плохо и суть дела затуманивает. Нужно различать процессы создания активной неравновесной организации и пассивной неравновесной организации. Эти процессы описываются разными теориями и называть их одним словом не нужно.

Кроме того, приставка “само” применительно к пассивным неравновесным системам конкретного смысла не несет, а лишь вносит какой-то элемент таинственности. На самом деле в пассивных неравновесных системах все происходит без какой либо активности со стороны системы, в которой появляется упорядоченность, а на основе физических свойств объектов и их естественного и пассивного взаимодействия.

Термин “организация” можно относить не только к процессу создания объекта, но и к его устройству, т.е. к конкретному результату упорядочивания или к результату антиэнтропийного процесса. Именно в этом смысле мы и будем использовать термин “организация”. Итак, организация, понимаемая как результат антиэнтропийного процесса, с одной стороны, определяет устройство объекта и, с другой стороны, сохранение (не разрушение) объекта во времени. Соответственно будем говорить об определяющей и о сохраняющей организации.

К сожалению, конструктивно формализовать определяющую организацию, то есть, количественно оценить ее уровень, а также формально определить связь свойств и структуры взаимодействия частей с качествами целого, или переход от целого к частям, к их составу, структурной организации, свойствам и взаимодействию, в общем случае не удается. Все что можно сделать это дать обобщенное описание организованной антиэнтропийной системы (объекта) и предложить какой-то способ ее формального представления.

Центральной проблемой теории систем считается формальное представление двусторонней связи между свойствами частей и целого. Эта связь в общем случае аналитически не выражается. Это не проблема или недостаток существующей математики, а, скорее всего, принципиальный момент. Формальный переход от свойств частей и взаимодействий между ними к неизвестным свойствам целого, так же как и переход от свойств целого к составу и свойствам образующих его частей, в общем случае, по-видимому, невозможен.

### **Представление целостного объекта**

И, все-таки, можно ли как-то формально представить двусторонние связи между уровнем целого и уровнем частей, а также и взаимодействия между элементами (частями) одного уровня? Попробуем это сделать на очень общем уровне с оговоркой, что наша цель состоит лишь в том, чтобы дать язык, удобную схему, в рамках которой можно было бы обсуждать рассматриваемые явления по возможности в их существенных и общих аспектах.

Эта схема, которая может показаться чересчур абстрактной, фактически использовалась нами при рассмотрении таких конкретных вопросов, как, во-первых, формальная модель поведения и, во-вторых, алгоритмы системы машинного зрительного восприятия, разработанные для распознавания изображений рукописных текстов (Графит, FineReader – рукопись).

При построении описаний полезно использовать язык переменных, называя переменной все, что можно (хотя бы теоретически) измерить, перевести в показания приборов, или оценить количественно каким-либо другим способом.

Все объекты обладают качествами (свойствами) и тем отличаются друг от друга. Переменная отражает наличие и интенсивность отдельного качества. Все что можно сказать о целом объекте на языке переменных – это, во-первых, перечислить его внешние качества, определить их интенсивность и указать закон изменения качеств во времени и в пространстве, и во-вторых, перечислить части объекта, их качества и отношение между частями.

Таким образом, объект можно задать точкой в пространстве переменных, где каждой оси соответствует некоторое качество или свойство. Будем различать два фазовых пространства - внешних переменных, соответствующих качествам целого, и внутренних переменных, соответствующих качествам частей (элементов) и отношениям между элементами. В частности, отношения между элементами могут задавать структуру или структурное описание целого. Пример таких структур, а также способ их задания и анализа рассматривался в предыдущей лекции применительно к задаче построения программной системы чтения рукописных знаков.

Пусть внешние качества отображаются в пространстве внешних переменных вектором  $Z(z_1, z_2, \dots, z_n)$ , а внутренние качества отображаются в пространстве внутренних переменных вектором  $X(x_1, x_2, \dots, x_m)$ .

Пусть переменные  $z_i$  и  $x_i$  отражают не абсолютные значения соответствующих качеств, а разность между текущими и некоторыми оптимальными или нормальными значениями. Тогда все переменные неотрицательны, т.е.  $x_i \geq 0$  и  $z_i \geq 0$ . Положение отображающей точки в начале координат соответствует нормальным (идеальным) значениям переменных (качеств).

Внешние качества полностью определяются внутренними. Поэтому положение точки  $Z(z_1, z_2, \dots, z_n)$  в пространстве внешних переменных однозначно определяется положением точки  $X(x_1, x_2, \dots, x_m)$  в пространстве внутренних переменных (обратное в общем случае не справедливо). Зависимость  $Z$  от  $X$  обычно не может быть определена аналитически, но может быть иногда охарактеризована каким либо другим образом,

например, с помощью словесного описания или таблично.

Введем теперь две области  $D_z$  - область допустимых значений внешних переменных и  $D_x$  - область допустимых значений внутренних переменных. Объект (система) существует как некоторая качественная определенность, пока отображающие точки  $Z(z_1, z_2, \dots, z_n)$  и  $X(x_1, x_2, \dots, x_m)$  находятся соответственно внутри областей  $D_z$  и  $D_x$ .

Область  $D_x$  ограничена координатными плоскостями в первом ортанте и некоторой поверхностью, аналитическое выражение которой может не существовать. В конкретных приложениях возможны предположения о каких-то свойствах этой поверхности, например, такие, как введенные ранее при рассмотрении модели поведения. Возможны также предположения о точном аналитическом виде ограничивающей поверхности  $F(X)=0$ , как, например, введенные при рассмотрении задачи построения программной системы зрительного восприятия. То же самое относится и к области  $D_z$ .

При изменении объекта отображающие точки  $X$  и  $Z$  перемещаются в фазовых пространствах допустимых значений внутренних и внешних переменных. Выход точек  $X$  и  $Z$  на границу области допустимых значений происходит одновременно. При выходе отображающей точки за границу области допустимых значений целостность нарушается, и объект прекращает существование.

Заметим, что в рамках введенных представлений можно говорить, например, об оценке состояния объекта, о закономерностях изменения состояния, о тенденции к сохранению или разрушению, о вероятности разрушения, об алгоритмах управления, направленных на сохранение качественной определенности объекта, о существенных и несущественных переменных и др. Все это может основываться на информации о положении точек  $X$  и  $Z$  в фазовых пространствах  $D_x$  и  $D_z$ , о направлении и скоростях перемещений этих точек, а также на информации о границах областей  $D_x$  и  $D_z$ . Однако конкретное рассмотрение всех этих и других подобных вопросов можно делать лишь в отношении каких-то конкретных систем.

## Принципы организации

Как возникновение порядка, так и его сохранение в разных антиэнтропийных системах происходят по-разному. Соответственно можно выделить следующие типы организации упорядоченных антиэнтропийных систем: пассивная равновесная, пассивная неравновесная, активная неравновесная.

Пассивная равновесная антиэнтропийная организация основывается на фундаментальных свойствах материи, таких как, в первую очередь, гравитация, а также электричество, электромагнитное поле, сильные и слабые внутриатомные взаимодействия. Эти, а может быть и какие-то другие неизвестные нам свойства материи, обеспечивают пассивный в основном случайный антиэнтропийный процесс, приводящий к возникновению и длительному сохранению, т.е. существованию материальных объектов.

При пассивной равновесной антиэнтропийной организации возникают упорядоченные объекты, обладающие свободной энергией, находящиеся в равновесии со средой, длительное время сохраняющие свободную энергию и свою качественную определенность вследствие существующих кинематических ограничений и не увеличивающие либо медленно увеличивающие энтропию. Сюда относятся, например, камень в углублении на склоне горы, каменный уголь, стул, или любой другой объект, имеющий организацию, обеспечивающую его внешние качества и относительно устойчивое существование. Пассивная равновесная антиэнтропийная организация может быть не только статической, но и динамической. Примерами пассивной равновесной динамической организации являются планетная система, атом, маятник.

Окружающая среда для пассивных равновесных систем является как актуальным, так и потенциальным источником внешних воздействий, причиной нарушения равновесного состояния и, в конечном счете, причиной разрушения. Точки, отображающие положение пассивной равновесной антиэнтропийной системы в фазовых областях  $D_x$  и  $D_z$ , хотя и медленно, но с неизбежностью движутся к границам областей допустимых значений. Система пассивна и никаких действий, направленных на продление существования в ней не происходит.

Пассивным неравновесным антиэнтропийным системам посвящена теория неравновесной термодинамики И.Р.Пригожина [N], описывающая возникновение и изменение открытых неравновесных диссипативных систем.

Сейчас многие представители биологии и синергетики на основе этой теории объясняют возникновение, существование и развитие жизни. При этом считается, что живая антиэнтропийная организация состоит в том, что при подводе к системе энергии в ней может возникать процесс создания неравновесных структур со свободной энергией и эта энергия будет диссипировать, т.е. выделяться в среду в виде тепла. Подобный антиэнтропийный процесс неравновесного упорядочивания, т.е. образования неравновесных диссипативных систем, как уже отмечалось, не очень удачно принято называть самоорганизацией.

Неравновесная диссипативная система пассивна. Такая система не выполняет ни внутренней, ни внешней работы для поддержания своего неравновесия, т.е. не может считаться живой. Процесс антиэнтропийной организации в неравновесной диссипативной системе идет принудительно за счет подводимой извне энергии. Еще раз отметим, что применение термина *самоорганизация* к процессу возникновения пассивных неравновесных диссипативных систем выглядит не очень оправданным, поскольку процесс организации в этих системах происходит и поддерживается принудительно за счет внешних сил.

В природе существует также и активная антиэнтропийная организация. Это в первую очередь живые, обладающие свободной энергией неравновесные системы, устойчиво поддерживающие и увеличивающие свое неравновесие, т.е. существование с помощью внутренней работы по синтезу и восстановлению живых неравновесных структур и внешней работы, т.е. целенаправленного взаимодействия со средой. Как внешняя, так и внутренняя работа активно выполняются за счет собственной свободной энергии живых неравновесных структур [4].

Этот процесс с большим основанием можно было бы назвать самоорганизацией, обеспечивающей непрерывное активное самовосстановление, т.е. сохранение во времени обладающих свободной энергией неравновесных структур. При этом свободная энергия не диссипирует через тепловой канал, а тратится на физиологические функции и восстанавливается за счет внутренней и внешней работы живого организма.

Подводя некоторый итог можно сказать следующее. Существуют два вида пассивных упорядоченных антиэнтропийных объектов – равновесные и неравновесные. При равновесном упорядочении свободная энергия со



временем уменьшается и энергия выделяется. Степень неравновесия и свободной энергии поддерживается существующими кинематическими ограничениями, которые со временем чаще всего снижаются. Среда оказывает на систему разрушающее воздействие.

При пассивном неравновесном упорядочении в зависимости от изменений в среде степень неравновесия и свободной энергии может уменьшаться, увеличиваться, либо оставаться неизменной. Неравновесие поддерживается принудительным внешним воздействием.

Собственная свободная энергия неравновесной диссипативной системы выделяется (диссипация) и восстанавливается за счет подвода энергии извне и вынуждаемых этим структурных перестроек в системе.

В живых активных антиэнтропийных системах неравновесие поддерживается за счет внутренней работы собственных системных сил, направленной на воспроизводство упорядоченных неравновесных структур, а также за счет собственного активного целенаправленного взаимодействия со средой, направленного на извлечения вещества и энергии из среды (внешняя работа).

Общий вывод состоит в том, что важные системные принципы, которым подчиняется как создание и поддержание жизни, так и ее эволюция, наряду с принципом устойчивого неравновесия это принципы *целостности, целенаправленности и активности*. Под целостностью понимается неразрывное единство целого и образующих его частей. Целенаправленность и активность, так же как и работа по поддержанию неравновесия, это то, что наиболее четко отличает живое от не живого. (Именно к целенаправленным процессам поддержания неравновесия в живых активных неравновесных системах наиболее естественно отнесение термина самоорганизация.)

Можно ли антиэнтропийную организацию и антиэнтропийные процессы описать формально и построить соответствующие математические модели? В отношении определяющей организации ответ в настоящее время должен быть отрицательным. Формально связать элементы (части) и целое в общем случае не удастся. Это относится как к пассивным (не живым), так и тем более к активным (живым) системам. Единственное что можно сделать, это описать фазовые пространства внутренних и внешних переменных, определить связи между переменными, исследовать внутренние и внешние фазовые траектории системы и закономерности их совместного изменения в этих пространствах.

В отношении сохраняющей организации возможно построение различных частных математических моделей, ориентированных на конкретные случаи организации. Возможно, что могут быть построены и модели сохраняющей организации какого-то общего характера.

Что касается антиэнтропийных процессов, то существуют формальные описания как для процессов, направленных на создание пассивной равновесной организации, так и для процессов неравновесной пассивной организации. Строгих формальных описаний процессов создания активной антиэнтропийной неравновесной живой организации нет. Существуют качественные описания - это различные теории начального возникновения жизни и теории эволюции.

В отношении эволюции можно сказать, что прогрессивная эволюция это антиэнтропийный процесс упорядочивания, т.е. создания определяющей и сохраняющей активной организации, противоположный процессу возрастания энтропии. Количественно оценить результаты упорядочивающего антиэнтропийного процесса прогрессивной эволюции можно по уровню создаваемого термодинамического потенциала.

## **Принцип maxT в поведении живых и искусственных систем.**

Системы с активной неравновесной организацией для своего существования должны быть открытыми и получать из среды необходимые для существования вещества на основе постоянного активного поведения. Мы ввели принцип maxT, как принцип оптимальности, объясняющий закономерности поведения животных.

Представляется, что принцип maxT является всеобщим и его можно отнести не только к поведению живых организмов, и даже не только к другим активным динамическим системам, для существования которых необходимо активное взаимодействие со средой, но и вообще к любым не только активным, но и пассивным как естественным, так и искусственным системам.

Применение этого принципа для объяснения поведения или выбора управления в активных неравновесных системах, требующих для сохранения своей целостности направленного взаимодействия со средой, наиболее очевидно. В пассивных системах длительное время существования, обеспечивается пассивно, за счет начальной организации.

На основе принципа maxT мы получили алгоритм принятия решения в ситуации выбора применительно к упрощенной одноэкстремальной формальной задаче поведения животных. Были также приведены качественные обоснования распространения полученных результатов на полный случай, то есть на поведение в многоэкстремальной среде. В основе описанной модели поведения кроме принципа оптимальности maxT лежат принципы целостности, целенаправленности и активности.

Все это может быть отнесено не только к живым организмам, но и к любым, в том числе и искусственным системам, названным нами открытыми системами с активно поддерживаемым устойчивым неравновесием. Такой системой является, например, промышленное предприятие, или какая-либо иная система, имеющая следующие характеристики:

- большое число управляемых параметров,
- постоянное изменение (ухудшение) параметров во времени, например, расход сырья, горючего или запасов энергии,
- ухудшение качества системы за счет амортизации компонентов, расхода денежных средств и т.п.
- ухудшение любого из компонентов приводит с течением времени к снижению запаса устойчивости системы и в пределе к нарушению или прекращению ее нормального функционирования,
- восстановление запаса устойчивости системы происходит извне (замена или ремонт стареющих компонентов, пополнение сырьем и материалами, пополнение запасов энергии и т.п.),
- ограничения возможностей восстановления запаса устойчивости,
- наличие случайного фактора, накладываемого на возможности восстановления компонентов за счет изменений внешней среды, например, перерыв в снабжении сырьем или денежными средствами, отсутствие рабочей силы и т.п..

Если конкретная искусственная система обладает подобными характеристиками и может быть описана рассмотренной нами математической моделью, то возможно оптимальное по критерию  $\max T$  управление этой системой с использованием локального правила выбора управляющего действия. Кроме того, если допустима перестройка технологических процессов внутри системы или изменение среды, например, смена поставщиков сырья, то возможна оптимизация внутренних процессов и отношений.

Реальные производственные системы чаще всего рассмотренной формальной моделью поведения напрямую, повидимому, не описываются. Одна из причин этого состоит в том, что в модели поведения животных предполагается альтернативность между выбираемыми управляющими действиями и дискретность обменных взаимодействий со средой. В отличие от этого, в производственной системе организация по критерию  $\max T$  обычно должна учитывать вариант, когда возможна непрерывная компенсация одновременно всех или многих накапливающих неустойчивость компонентов с единственным ограничением по сумме отпускаемых денежных средств, отнесенной к единице времени. В таких системах должна решаться задача не о выборе цели поведения, а об оптимальном по критерию  $\max T$  распределению денежных средств. Повидимому, возможны и смешанные варианты.

Можно предположить, что критерий  $\max T$  не является универсальным и конкретную активную динамическую систему нужно строить, исходя из других критериев. Например, в каких то случаях, можно говорить о системе, приносящей максимальную прибыль, или о системе, приносящей прибыль, но не вообще, а на заданном отрезке времени, после чего существование системы прекращается, или, допустим, о системе, максимально надежной защищающей государственную границу и так далее – возможны какие угодно самые разнообразные формулировки, определяющие систему.

Может показаться, что в названных случаях максимизироваться должно не время, а какие то другие параметры. Однако, ни максимальная прибыль, ни максимальная прибыль на отрезке времени, ни максимально надежная защита границы не являются прямыми критериями управления и должны использоваться только при проектировании системы для определения состава параметров управления, взаимосвязей между ними и вида области существования системы.

Система должна быть так спроектирована, что задаваемые внешние критерии, например, максимальная прибыль, включаются в состав параметров, создающих качественную определенность системы. В введенных нами обозначениях это вектор внешних переменных (качеств) системы -  $Z(z_1, \dots, z_n)$ . Далее нужно решать задачу максимизации времени существования системы как заданной качественной определенности, то есть нужно выделить переменные управления - вектор  $X(x_1, \dots, x_m)$ , определить вид границы области допустимых значений переменных, определить ограничения на управление и решать задачу управления исходя из критерия  $\max T$ .

Построение конкретной математической модели во многих случаях может представлять самостоятельную непростую, а, может быть, и не решаемую формально проблему.

Рассмотрение разных видов организации позволяет сделать следующий вывод: для любых систем главный результат – это обеспечение своего существования как некоторой качественной определенности. Критерий качества организации для пассивных как равновесных, так и неравновесных систем и критерий качества организации и управления (поведения) для открытых активных неравновесных систем формулируется одинаково:  $\max T$  или с учетом недетерминированности среды  $\max M(T)$ , где  $M(T)$  – математическое ожидание времени пребывания существенных переменных системы внутри области ее существования.

## Лекция 9. **Информация**

Мышление и информация неразрывно связаны. Мышление это всегда осознанная работа с информацией, состоящая в конечном счете либо в получении, обработке, фиксации и передаче информации, либо в ее использовании.

Для существования активных открытых динамических систем с накапливающейся неустойчивостью (в частности живых организмов) нужна среда. Для организации эффективного управления (поведения) в среде необходимо знание среды. Соответствующая информация определенным образом должна структурироваться и отображаться в формирующейся в мозге модели среды. Как мы уже говорили, специфическая сложная и многоплановая информация о среде особенно требуется, если в процессе управления (поведения) необходимо решать многоэкстремальные задачи. В связи с этим сначала коснемся вопросов, связанных с понятием информация. Затем в этой и в последующих лекциях рассмотрим общие представления о структуре, функциях и принципах работы строящейся в мозге человека (животного) модели среды.

### **Подходы к определению и измерению информации.**

Описание целостного объекта или системы должно содержать информацию об этом объекте или системе. Для того, чтобы построить или просто представить себе что-то нужна информация о том, что это такое, какими обладает свойствами, как это выглядит и так далее. В двух предыдущих фразах использовано слово информация, а с начала лекции оно использовано уже шесть раз, и у читателя в связи с этим не возникает никаких затруднений при чтении. Ни у пишущего, ни у читателя как в приведенных примерах, так и практически в любом другом случае при использовании или чтении слова информация, не возникает потребности в определении и, тем более, в формализации этого понятия, если только этот термин не употребляется в каком-то специальном смысле.

Тем не менее, попытка формализации понятия информации произвела большое впечатление на представителей гуманитарных наук. Сразу после выхода классической работы Клода Шеннона, у биологов, психологов, лингвистов, историков и представителей других гуманитарных дисциплин появилась надежда на то, что эти описательные эмпирические дисциплины, используя предложенную формализацию понятия информация, станут науками более точными и теоретическими. Этим надеждам не суждено было сбыться, поскольку теория информации Шеннона относится лишь к одному узкому аспекту проблемы и не имеет отношения к решению основных вопросов, представляющих интерес для исследований, проводящихся в рамках гуманитарных наук.

После Шеннона представителями математики, кибернетики и даже философии предпринимались многочисленные попытки формализации понятия информации при его более широкой или, наоборот, более узкой и в каком либо смысле более полезной трактовке. Однако, до сих пор достаточно полной, законченной, устраивающей всех строгой теории информации не придумано. Оказалось, что интуитивно понятный термин поддается формализации с большим трудом. Интересно понять почему. Рассмотрим связанные с этим вопросы несколько подробнее.

Для определения понятия, часто достаточно уметь отвечать на два вопроса: Что это такое? и Как это измерить? Ответ на первый вопрос дает качественную определенность, ответ на второй – количественную. Ответ на первый вопрос необходим всегда. Второй вопрос можно поставить не всегда, а лишь тогда, когда определяемое понятие относится к маломерной в принципе измеряемой сущности (величине, качеству, переменной). В случае с информацией это не так. Вопрос - как это измерить? применительно к информации можно поставить далеко не всегда.

Первый вопрос – что это такое? , часто особых затруднений не вызывает. Обычно всех на бытовом уровне удовлетворяет примерно такой ответ: информация это сведения о чем-то, содержащие или описывающие знание чего-то. Однако, в век интенсивного развития информационных технологий такое нестрогое определение многих не удовлетворяет. Было предложено очень много разных определений, в том числе и определений, ориентированных на построение разных вариантов теории информации. Общепринятого строгого определения нет.

Выделяются два основных научных подхода к определению того, что такое информация. Согласно первому, информация есть внешнее описание объекта, содержащееся в передаваемом сообщении. Количество информации в этом случае определяется как мера неопределенности, снимаемая полученным сообщением (Шеннон [18]). Согласно второму подходу, информация не связывается с передаваемым сообщением. Она относится к объекту и есть мера разнообразия объекта. Странником такой точки зрения является Эшби..

Для первого подхода важен процесс передачи информации и включение в рассмотрение не только объекта – носителя информации, но и субъекта – приемника информации. Теория информации Шеннона возникла из статистической теории связи и имеет прямое отношение лишь к таким операциям с информацией как ее кодирование, передача по техническим линиям связи и хранение в памяти технического устройства, например в памяти электронной вычислительной машины.

Основная формула Шеннона позволяет определить количество информации в сообщении  $X$ , вероятность которого  $P(X)$ .

$$I(X) = -\log P(X).$$

Другая формула Шеннона позволяет определить количество информации, содержащейся в сообщении из  $N$  символов в зависимости от частоты их встречаемости:

$$I = -N \sum_i P_i \log P_i.$$

Еще одна формула Шеннона такова:

$$I(x,y) = \log(P(x/y) / P(x)),$$

где  $I(x,y)$  — информация, которую содержит событие  $y$  о событии  $x$ ,  $P(x)$  вероятность события  $x$  до наступления события  $y$ ,  $P(x/y)$  — вероятность события  $x$  при условии наступления события  $y$ .

Таким образом вводится количественная мера, которая может быть использована при проектировании технических устройств для оценки количества передаваемой или запоминаемой информации, а также при решении ряда сопряженных проблем, например таких, как кодирование, фильтрация, помехоустойчивость и т.п.

Для процесса обработки информации, осуществляемого мозгом человека, или сложной искусственной системой наибольшее значение имеют совсем другие аспекты. Игнорирование теорией Шеннона содержательной смысловой стороны информации послужило поводом для утверждений о недостаточности теории, хотя сама теория имеет четко ограниченный предмет и на всеобщую полноту не претендует.

Итак, при первом подходе к определению того, что это такое, информация понимается как мера изменения знаний субъекта об объекте в процессе их взаимодействия. Субъект получает информацию не вообще, а лишь в некотором заданном или выбранном отношении.

Такой подход с необходимостью вносит субъективность. Одно и то же сообщение может иметь различное информационное содержание для разных субъектов в зависимости от того, что они хотят и что могут узнать. Например, один субъект избирательно узнает из радионовостей как сигналы Спартак, второй не знает, что такое Спартак и узнает какая будет завтра погода, а третий хочет и может узнать и то и другое.

В общем случае и чаще всего, субъект может получить информацию не вообще, а лишь как дополнение к тому, что он уже знает.

Если поступило информационное сообщение о том, что слон белых стоит на поле  $e5$ , то для того, чтобы получить из этого сообщения информацию нужно иметь определенные необходимые знания, на которые эта информация может быть наложена. Нужно понять, что речь идет о шахматах, нужно знать шахматную нотацию и то, что слон может находиться на одном из 32 полей шахматной доски, в том числе и на поле  $e5$ . Нужно понять, что поле  $e5$  черное. Только зная все это, а также то, какая часть черных полей шахматной доски кроме поля  $e5$  свободна, можно подсчитать, по формуле Шеннона, сколько бит информации содержится в сообщении. Так если каким-то образом определить, что все черные поля кроме поля  $e5$  свободны, то можно подсчитать, что в сообщении содержится 5 бит информации.

В более сложном случае, когда в сообщении приводится позиция с большим числом фигур, подсчет количества информации, содержащейся в этом сообщении по формуле Шеннона практически невыполним. Правда, если речь идет о записи позиции, как и в любом другом случае, этот подсчет никакого значения для шахматиста не имеет. Точно также, для человека, получившего сообщение о том, как сигналы Спартак, никакого значения не имеет содержащееся в этом сообщении количество бит информации.

При другом подходе считается, что информация является объективной характеристикой объекта, такой же, например, как энергия, то есть информация заключена в самом объекте. Чем сложнее организован объект, тем больше он содержит информации, которая является объективной мерой разнообразия.

Однако, если принять тезис о неисчерпаемости материи вглубь, то при таком подходе неожиданно оказывается, что все объекты вне зависимости от их сложности обладают одинаковым количеством информации, равным бесконечности. В любом случае, сделать количество информации о конкретном объекте обозримым можно только одним способом – вводя уровень рассмотрения.

Количество информации, которое можно получить об объекте в этом случае зависит от выбранного уровня рассмотрения и характеризует объект не полностью, а лишь в определенном отношении. Информации вообще как бы не существует. Есть лишь информация о чем то. Вводя уровень рассмотрения нужно ввести и рассматривающего субъекта. В этой точке два выделенных выше основных подхода к определению того, что такое информация смыкаются.

Итак, для получения информации нужны информационный срез (уровень рассмотрения) и, в общем случае, априорная информация. Как выделяется уровень рассмотрения? Обычно рассматривают два различных варианта.

Часто говорят, что объект может содержать информацию и исследоваться в отношении целевой функции субъекта. При достаточно общем понимании термина “управление” можно говорить, что информационный срез в этом случае определяется собственными задачами управления субъекта.

Такой подход позволяет достаточно естественно, правда, не всегда успешно, вводить количественные меры не только для определения количества информации в Шенноновском или каком-либо ином смысле, но также и для определения содержательности и ценности получаемой информации.

В другом варианте говорят, что объект может исследоваться вне зависимости от целевой функции субъекта, то есть получаемые сведения не используются для управления. В этом случае строятся информационные срезы в

отношении организации самого объекта, в частном случае, в отношении его устройства или целевой функции, если она определена.

Большое внимание в работах по теории информации уделяется проблеме измерения ее количества. После того, как было выяснено, что предложенный Шенноном статистический подход к измерению количества информации не учитывает ее смысловую, содержательную сторону и в связи с этим во многих случаях, очевидно, неприменим, было предложено много альтернативных подходов, например: алгоритмический, семантический, прагматический, топологический, термодинамический.

Эта классификация сделана на основе работ, относящихся, главным образом, ко времени кибернетического бума шестидесятых годов. Последующие работы практически ничего принципиального к приведенной классификации не добавляют.

Обычно во всех работах по определению количественной меры информации вводится какой-то дополнительный аспект путем постановки вопросов типа: как это используется (объектом или субъектом) или что это дает (объекту или субъекту). Правда, на такие дополнительные вопросы не всегда возможен точный ответ. Например, количество информации может определяться как мера устранения неопределенности, имеющейся у субъекта, по выбору действий, ведущих к достижению субъектом каких-либо целей.

Такой подход к определению количественной меры информации типичен. Схема Шеннона, в центре которой между объектом и субъектом находится сообщение, по сути сохраняется, но вводится определенная семантика. То есть для вычисления количества передаваемой от объекта к субъекту информации нужно вычислять вероятности или меры неопределенности или энтропии и тому подобное, но не вообще, а при некоторых дополнительных условиях, что в определенном смысле эквивалентно введению уровня рассмотрения. Так, для того, чтобы приведенное выше определение практически работало, для измерения количества информации нужно еще знать о возможных действиях субъекта и о способе их выбора, а также о целях субъекта, правилах постановки и выбора целей и о способах их достижения.

Таким образом, в классической теории информации, так же как и в упомянутых ее модификациях, на первое место ставится оценка количества информации. При этом в подавляющем большинстве случаев, интересных для гуманитарных исследований, эта оценка либо практически невыполнима, либо попросту не представляет интереса.

Поэтому, оставаясь в рамках схемы Шеннона, подходящей для оценки количества простой хорошо структурированной информации, передаваемой от одного устройства (субъекта) к другому, оправдать надежды биологов, психологов, историков, лингвистов и представителей других гуманитарных дисциплин на то, что эти дисциплины, используя теорию информации, станут науками более точными и теоретическими, невозможно. Все эти науки по-прежнему остаются описательными и эмпирическими. Надежда на помощь этим гуманитарным дисциплинам со стороны более точных наук может снова возникнуть, по-видимому, лишь при другом взгляде на информацию и при разработке на этой основе семантической теории информации.

Итак, можно ли и нужно ли дать строгое формальное определение термину “информация”? Общего строгого определения, по-видимому, давать не нужно. Это слово просто и понятно. Информация это информация или сведения о чем-то.

Тем не менее, существует очень много самых разных определений.

От общих и простых типа информация это сведения, или информация это знания, до специальных и более сложных, например: “информация есть случайный и запомненный выбор одного варианта из нескольких возможных”. Часто отождествляют информацию и отражение, например: “информация есть отражение в сознании человека причинно-следственных связей в окружающем мире”. Иногда, говоря об информации как об отражении, считают, что информация имеет отношение не только к живой природе, но в какой-то форме существует и передается при взаимодействии объектов неживой природы.

Иногда говорят даже, что информация имеет тот же статус, что вещество и энергия и все в мире состоит из вещества, энергии и информации. Правда, с этим согласиться труднее всего. Если что-то и добавлять к веществу и энергии, то это естественней было бы назвать не информацией, а организацией, обеспечивающей качественную определенность и существование во времени материальных объектов (систем). При этом конкретной организации может соответствовать и конкретная, определенным способом зафиксированная и, в принципе, передаваемая информация об этой организации.

Иногда в некоторых современных и не только современных теориях сопоставляя статус вещества, энергии и информации идут еще дальше, утверждая, что информация первична, а материя вторична. При этом ссылаются на Библию – вначале было Слово, или на то, что перед постройкой дома должен существовать его план, хотя бы в голове строителя. Обсуждать вопросы религии мы здесь не будем. Что касается плана в голове строителя дома, то этот план ни что иное, как информация об организации задуманной постройки, зафиксированная на материальном носителе – мозге. Кроме того, план появляется не на пустом месте, а на основе разнообразной существующей в мозге информации об организации и законах взаимодействия объектов окружающей материальной среды.

Еще раз подчеркнем, что организацию системы (объекта), определяющую его качественную индивидуальность и сохранение качеств во времени, отождествлять с информацией не нужно. Организация может быть (к сожалению, не всегда) записана на каком-то языке, передана или запомнена. Таким образом, организация,

как и все в мире несет в себе лишь потенциальную информацию. Кроме того, в мозге возможно возникновение информации об организации воображаемых не существующих реально объектов.

Итак, есть очень много различных трактовок понятия информация. В конце концов, это дело вкуса или удобства построения каких либо теорий. Чтобы не конкурировать со многими, в том числе и с появляющимися в последнее время новыми определениями, вводить еще одно определение понятия информации мы не будем. Тем более что общее удовлетворяющее всех строгое определение, по-видимому, дать невозможно.

Попробуем, не давая строгого определения, уточнить наши качественные представления об информации. В мире есть материальные и не материальные объекты, явления, ситуации и процессы. Все это на самых разных уровнях рассмотрения может быть описано (зафиксировано, отражено) в виде информации трех разных типов - как описание, как алгоритм построения или как активная модель. Для создания описания или алгоритма нужен язык. Это может быть обычный язык общения, либо специальный язык, в частности язык математики. Для фиксации информации в виде активной модели, могут использоваться какие-то другие средства, в частности и главным образом нейронные механизмы мозга.

Мне кажется, что информацию, понимаемую как описание, нужно (удобно) связывать с фиксацией, хранением или передачей, и, следовательно, с фиксирующим, передающим, принимающим или запоминающим информацию субъектом. Кроме того, информацию всегда нужно связывать с используемым информационным срезом. Информация первого вида, т.е. языковое описание создается человеком. Активная модель среды, или модель мира создается в мозге человека или в мозге животного в процессе обучения и приобретения жизненного опыта.

Таким образом, будем привязывать понятие информации, понимаемой как модель среды, к мозгу, в первую очередь к мозгу человека, и считать, что информация о среде возникает и фиксируется в виде модели в процессе активного, субъективного отображения материального и не материального мира и самого субъекта отображения (человека или животного) в его мозгу. Информация, поступает в мозг либо как отражение мира при взаимодействии с ним, либо как описание. Эта информация используется для построения субъективной модели среды и себя в среде, или для непосредственной организации восприятия и поведения в конкретной текущей ситуации. При понимании информации как знаний, фиксирующихся в виде активной модели среды, в рассмотрение вводятся объект – среда, субъект человек (животное), получение и фиксация – отображение в мозгу в виде целостной модели и ее актуализируемых фрагментов, информационный срез – например, проблемная среда в задаче управления поведением и ее специальное частное отображение, определяемое текущими первичными и вторичными потребностями.

Таким образом, результатом исследования мира является строящаяся в мозгу человека информационная модель. Информационное содержание этой модели в какой-то форме и в каком-то объеме может быть переносимо на внешний носитель информации в виде описания. Информационная модель всегда не полна, поскольку относится лишь к части мира и к каким-то информационным срезам. Модель чаще всего не точна, являясь лишь некоторым приближением к действительности. Модель всегда субъективна. Так же как при строительстве дома очередной кирпич кладется на предыдущий или стыкуется с предыдущим, при построении модели новые знания сочетаются, стыкуются, согласуются или каким то иным образом учитывают предыдущие. Важной особенностью строящейся в мозге модели среды является ее иерархичность не только по линии общее-частное, но и по линии часть-целое.

Добавим еще, что информационная модель также как и описания могут быть просто неправильными, то есть строиться на основе и в сочетании с ложными знаниями, заблуждениями или ошибками. Следует заметить, что и так называемая научная картина мира, фиксирующаяся и передающаяся в виде описаний, не только не полна и не точна, но и в каких-то аспектах может быть далека от объективной реальности и, попросту, ложна. Вспомним хотя бы астрономию Птолемея, теорию эволюции Ламарка или агрессивную агробиологию Лысенко.

В последующих лекциях, при ориентации в первую очередь на задачи поведения и восприятия, будет говориться не о модели мира, а о модели проблемной среды. Что такое проблемная среда? Проблемная среда животных это среда, в которой решается задача поведения, оптимизируемого по критерию  $\max T$ , то есть задача удержания регулируемых переменных – первичных физиологических потребностей внутри области допустимых значений. Проблемная среда человека это тоже среда, в которой решается та же задача  $\max T$  с добавлением к первичным физиологическим потребностям вторичных потребностей, определяемых, в первую очередь, общественным способом существования и развитыми социальными отношениями.

Итак, информация, фиксирующаяся в виде модели проблемной среды, это субъективное отображение мира, которое строится в мозге для решения задач удовлетворения первичных и вторичных потребностей обладающего мозгом организма. Это отображение определяется в каждый момент целью (задачей), средствами и возможностями организма, а также накопленной к данному моменту информацией. Поэтому, нужно еще раз подчеркнуть, что информационное отображение среды в мозге не является простым зеркальным отражением. Информационное отображение всегда неполно и субъективно и обычно неточно.

Чем отличается активная модель от описания? Описание статично и пассивно. Строящаяся в мозге модель среды динамична и активна. На этой модели разворачивается активный процесс мышления, включающий предвидение, умозрительное моделирование и умозрительное конструирование. Модель должна “жить по законам

среды". В построенной в мозге модели должен не просто актуализироваться фрагмент, соответствующий текущей воспринимаемой ситуации. По модели должно осуществляться расширение этого фрагмента как в пространстве, так и во времени. Должно происходить соотнесение себя со средой как субъекта восприятия и поведения в отношении прошлого, настоящего и будущего. Возможно, что на этом основывается функция сознания. В основе активности лежит возникновение в модели очагов возбуждения и потребностей их гашения. Эти вопросы еще будут затрагиваться в дальнейшем.

Говоря об информации часто разделяют данные и знания. Данными считается совокупность сведений о независимых или слабо зависимых фактах или объектах. Данные чаще всего представляются как описания в виде списков, таблиц, реестров и т.п. Знаниями считается информация о целостной системе взаимосвязанных фактов или объектов. Знания могут представляться как с помощью текста, так и специальных языковых средств, например таких как логика предикатов, графы понятий и отношений, семантические сети или фреймы. Эти сложные в том числе и иерархические информационные системы не являются моделями. Они остаются описаниями поскольку не обладают свойством активности.

Наиболее адекватным способом представления знаний является строящаяся в мозге активная модель проблемной среды. При изучении и попытках конкретизации процессов мышления отдельный интерес будет представлять перевод информации из формы описания в форму модели среды и обратный процесс – перевод из формы модели в форму описания.

Что дает введенное представление об особом виде информации, создающейся не как описание, а как модель проблемной среды? Для целей измерения ее количества практически ничего. В то же время, определение информации как модели проблемной среды дает основание и возможность поговорить о том, что такое модель проблемной среды, сформулировать гипотезы о том, как эта модель устроена в мозгу человека, как она строится и используется, чем отличается фиксация информации и работа с ней в мозге и в вычислительной машине. Это, в свою очередь, может дать возможность связать представления о модели среды с проблемами моделирования нейронных механизмов мозга и сформулировать качественные требования к этим моделям. Большинство из этих вопросов в той или иной мере будет затронуто в следующих лекциях..

Итак, единого, устраивающего всех определения информации не существует. Проблема по-видимому состоит в том, что информация может существовать в трех разных видах – в виде описания, в виде алгоритма построения и в виде строящейся в мозге активной целостной модели среды. Дать определение информации, имеющей вид описания или алгоритма проще. Дать строгое определение информации, существующей в виде активной модели среды сложно. Может быть так же сложно, как дать строгое определение процессу мышления.

## **Ценность информации.**

Наряду с количеством часто рассматривается вопрос о ценности получаемой информации. Мы уже рассматривали примеры когда одна и та же информация может иметь разную ценность для разных людей. Человек обычно уверенно отбирает информацию по ее субъективной ценности. Интересно понять как он это делает. Интерес этот не является чисто теоретическим. Проблема отбора информации по ее ценности из часто очень большого входного потока имеет важное значение для оперативного управления сложными системами.

Определения ценности информации вводились многими авторами. Обычно эти определения строятся на основе сопоставления по каким-то критериям вариантов достижения заданных целей. Например, предлагается сравнивать вероятности достижения определенной цели с учетом или без учета пришедшей информации (см., например, [5]), либо сравнивать варианты достижения цели до и после получения информации по расходу каких-либо полезных ресурсов, например таких, как материалы, энергия, денежные средства и т.п.. Такие определения ценности информации естественно назвать "поведенческими", поскольку в них всегда присутствует цель поведения или управления. Отличия состоят в разных критериях сравнения вариантов достижения поведенческой (управленческой) цели.

Приведенные подходы к определению ценности информации имеют серьезные недостатки. Во-первых, поведенческий аспект, связанный с задачей достижения заданной цели, не является единственным при получении и использовании информации. Во-вторых, предлагаемые формулы оценки вариантов чаще всего являются чисто теоретическими и их практическое применение может быть невозможно, либо связано со значительными трудностями. Есть и другие недостатки, на которых мы остановимся ниже.

Можно выделить два варианта использования поступающей в мозг информации. В первом варианте информация принимается и используется для построения в мозге отображения целостной системы знаний о проблемной среде. Информационные срезы, которые в этом случае делает принимающий информацию субъект, конечно определяются задачей поведения, но не оперативной задачей, а общей задачей удовлетворения первичных и вторичных потребностей, т.е. задачей максимально долгого удержания потребностей внутри условной субъективной области существования (принцип maxT). При этом какие-то вторичные цели, с учетом которых строилась в мозге субъективная область существования, могут быть значительно отнесены во времени от текущей ситуации принятия решения.

Человек целенаправленно отбирает нужную ему информацию, определяя ее ценность по связи со сформированной в его мозге субъективной областью существования. Так студент математического факультета и

студент консерватории выберут в библиотеке или в магазине разные учебники. Информация, содержащаяся в учебнике, выбранном математиком, как правило, будет иметь нулевую ценность для музыканта, так же как и информация, содержащаяся в учебнике, выбранном музыкантом, для математика. В обоих случаях могут иметься очень общие отдаленные поведенческие цели – получить музыкальное или математическое образование, т.е. построить специальные отображения проблемной среды. Количественное определение ценности информации в этих случаях вряд ли возможно.

Во втором варианте поступающая в мозг информация принимается и используется для организации оперативного поведения. На первый взгляд, этот случай подходит для применения формул подсчета ценности информации типа предложенной М.М.Бонгардом [5]:

$$W = \log P/p \quad (1)$$

или

$$W = (P-p) / (1-p) \quad (2)$$

где  $p$  – вероятность достижения цели до получения оцениваемой информации, а  $P$  – вероятность достижения цели после получения информации.

Однако, практически воспользоваться приведенными формулами в реальных случаях управления чаще всего будет невозможно. Во-первых, трудности могут возникнуть при подсчете вероятностей. Во-вторых, вероятности достижения цели до и после получения информации могут быть одинаковыми и оказаться равными, например, 1. В этом, как и в других случаях, сравнивать варианты достижения цели может оказаться полезным не по вероятностям, а по каким-то иным критериям, например по времени. В общем случае этот выбор может быть многокритериальным, т.е. сравниваться должны не отдельные параметры, а изменения целевой функции многих параметров. В-третьих, для работы с приведенными формулами должна рассматриваться фиксированная цель, в то время, как информация может использоваться для оценки и выбора целей. И, наконец, недостатком является и то, что при использовании приведенных формул задача должна просчитываться до конца (до цели), что не всегда возможно. В принципе, те же недостатки относятся и к случаю, когда предлагается сравнивать не вероятности достижения цели, а какие-то расходуемые ресурсы.

Представляется, что во многих случаях ценность информации, используемой для оперативного поведения (управления), может определяться на основе рассмотрения задачи максимизации времени удержания системы внутри области допустимых значений регулируемых переменных (принцип  $\max T$ ). Информация, используемая для принятия решения в задаче поведения может быть двух типов. Во-первых, это базовая информация о среде. Источник этой информации - построенное в мозге отображение или модель среды. Во-вторых, это информация о текущей ситуации. Эта информация поступает по сенсорным каналам и актуализирует соответствующий ей фрагмент модели среды. И та и другая информация используется для выбора варианта поведения на основе оценки вариантов приращения целевой функции на рассматриваемых локальных шагах поведения.

Если задача описывается рассмотренной нами ранее упрощенной формальной схемой то выбор варианта поведения осуществляется по формуле:

$$\min \Delta Z_j = \min \left( \sum_i \frac{\partial F(X)}{\partial x_i} U_i - \frac{\partial F(X)}{\partial x_j} V_j \right) \quad (3)$$

j

Эта же формула лежит в основе “эмоциональной” оценки сравниваемых вариантов поведения. Общие принципы сведения более сложной реальной задачи поведения или управления к этой упрощенной схеме рассматриваются как в лекции 4, так и в последующих лекциях. В реальной задаче управления формула подсчета приращения целевой функции при управлении по критерию  $\max T$  может быть иной.

Формула для выбора варианта поведения, приведена здесь для иллюстрации того, каким многоплановым может быть процесс принятия решения в сложной динамической системе. Получение дополнительной информации может изменить количественную оценку как любого имеющего значение параметра  $\{i\}$ , так и сравниваемых вариантов управления  $\{j\}$ . Соответствующее изменение значения целевой функции, или эмоциональной оценки в случае поведения, может происходить как в положительную, так и в отрицательную сторону. Абсолютное нормированное значение этого изменения может считаться ценностью полученной информации. Если информация истинна, то ценность всегда имеет положительное значение независимо от того ухудшает или улучшает введение новой информации оценку целевой функции, в том числе независимо от того, увеличивается или уменьшается апостериорная вероятность достижения цели. (Представляется, что слово значимость здесь было бы лучше, чем принятый термин ценность.)

Все сказанное выше относится к случаю, когда получаемая информация истинна. Получение неверной информации может приводить к неверным управляющим решениям, ухудшающим состояние управляемой системы, оцениваемое по значению целевой функции В таких случаях иногда говорят об отрицательной ценности информации. Представляется, что в этом случае правильнее говорить не об отрицательной ценности, а о



дезинформации. Значимость дезинформации, в принципе, можно измерять так же как и ценность информации, например по изменению целевой функции.

Для определения ценности как в случае применения формулы (3), так и в случае применения формул (1) или (2), информацию нужно получить и решить задачу с ее использованием либо на локальном шаге поведения либо до конца. В обоих случаях определение значимости полученной и уже использованной информации особого смысла не имеет.

Полезной была бы какая-то быстрая априорная предварительная оценка и отбор важной информации из большого входного потока. Например, для оперативного управления полезно по каким-то простым признакам отсеивать ненужные сообщения, отбирать нужные документы, назначать нужные встречи и т.п. Также по каким-то внешним признакам можно устанавливать градации важности (ценности) входной информации и соответствующие приоритеты ее приема и использования. В такой постановке задача определения ценности информации для ее предварительного "просеивания" близка задаче распознавания. Точного решения такая задача не имеет. Использование построенных на основе опыта эвристик и решающих правил сопряжено с риском ошибиться.

## **Познание и информационное отображение мира**

Информационное отображение мира в мозгу животного или человека мы называем моделью мира. Применение термина модель основано на том, что строящееся в мозгу отображение мира не должно быть простым описанием. Это отображение должно обладать свойствами модели - быть активным и динамичным, т.е. его фрагменты должны актуализироваться на основе выявленных в предыдущем опыте и отраженных в модели законов среды. Опережающее моделирование ситуаций взаимодействия со средой и сравнение результатов моделирования с реальными результатами взаимодействия или восприятия происходит в мозге постоянно и непрерывно. Этот процесс получил в физиологии название акцептор действия. Вопросы, связанные с построением в мозге и работой модели среды будут еще затрагиваться в последующих лекциях.

Формирование модели мира и, в частности, модели окружающей среды может происходить с использованием первичных или вторичных знаний. Вторичные знания возникают при передаче части модели одного человека другому. Источником знания в этом случае может быть книга, преподаватель, проповедник, актер или любой другой человек. Передача вторичных знаний в основном осуществляется при помощи передачи информации в виде устных или письменных текстов.

Большую часть первичных конкретных знаний о той части мира, которая называется средой обитания, каждый человек получает сам в процессе поведенческого эксперимента главным образом методом проб и ошибок. Первичные знания возникают при прямом взаимодействии человека с миром. Это взаимодействие может происходить при помощи органов чувств или при помощи измерительных приборов. Ограниченность и того и другого определяет то, что человек взаимодействует лишь с частью мира и соответственно, получает знания только о его части. Так радиации как бы не существовало до ее открытия и создания соответствующих датчиков. Не исключено, что в мире и непосредственно вокруг нас есть много такого, чего мы не знаем и поэтому не замечаем.

Находясь на материалистических позициях будем считать, что мир реален. Тогда строящаяся в мозгу человека модель есть результат частичного, приблизительного, субъективного отражения объективной реальности. Кроме того, человек строит в своем мозгу "модель себя" и себя в мире. Возможно, что это и является основой сознания. Сознание и самосознание если и не синонимы, то очень близкие сущности.

Оставаясь на материалистических позициях, будем считать, что в мире наряду с материальными существуют и не материальные объекты, явления и даже системы. Примеры: дружба, любовь, мировоззрение, искусство. Конечно, все это не существует само по себе без опоры на что-то материальное. Однако понятно, что в картине есть что-то гораздо большее, чем краски и холст, на который краски нанесены, а любовь не сводится к физическому контакту.

Познание человеком мира ограничено выбираемыми информационными срезами, нужными или полезными в каком-то отношении, и существующими экспериментальными возможностями.

Мир как-то устроен. Есть общие принципы оптимальности, есть физические, химические, социальные и другие законы, есть существующие в мире материальные и не материальные системы и объекты, организованные каким-то образом для сохранения во времени своей качественной определенности, и находящиеся во взаимодействии и каких то других отношениях между собой. К сожалению, о значительной части всего этого человек ничего не знает.

Обычно меньше всего человек знает о внутрисистемных взаимодействиях и отношениях, определяющих появление и сохранение во времени качеств целого, не сводимых к качествам составляющих целое элементов. Если считать, что все объекты и явления в принципе могут быть описаны и потенциально содержат в себе полную информацию о своих свойствах, своем устройстве и организации, то можно сказать, что мир является своим собственным полным "описанием".

Может ли человек прочесть до конца и понять это описание неизвестно. Может ли человек понять бесконечность времени, и бесконечность пространства? Скорее всего нет. Может ли кролик получить хоть какое-нибудь представление о сонетах Шекспира? Нет, не может. Строящаяся в мозгу кролика модель мира не имеет и

не может иметь никаких пересечений с сонетами Шекспира. Если не разделять представлений древних евреев о возникновении и устройстве мира и не считать землю центром мироздания, то естественно допустить, что в бесконечном и вечном мире может иметься что-то принципиально не имеющее никаких не только временных и пространственных, но и смысловых пересечений со строящейся в мозгу человека моделью мира. Прочсть и понять какую-то часть “описания” бесконечного мира, касающуюся относительно небольшой конечной и обозримой для человека его части, стремятся наука, религия, искусство и каждый человек в процессе поведенческого взаимодействия со средой.

Правда, не известно может ли человек понять все, имеющее отношение хотя бы только к его Вселенной и даже только к обозримой ее части. Например, неплохо было бы для начала понять природу сильных внутриатомных взаимодействий, или природу электричества, а также гравитационного и электромагнитного полей. Важнейшими и труднейшими загадками остаются загадки жизни и разума. И возможно, что главная загадка бытия это вопрос – что такое и как устроено время.

Касающиеся всего этого физические и математические модели, в том числе и так называемые “превосходные” [12], описывают только внешнюю сторону проявлений фундаментальных свойств материи и принципов организации мира на уровне взаимодействия между объектами макро или микро уровней и чаще всего не затрагивают их сущности. (К “превосходным” физическим теориям Р.Пенроуз [12] относит такие, которые достигают очень высокую точность описания существующих явлений).

Например, только внешнюю сторону объектов и явлений мира описывают превосходные физические теории Галилея, Ньютона или Эйнштейна. Так теория гравитации Ньютона определяет только зависимость силы притяжения от масс притягивающихся объектов и расстояния между ними и ничего не говорит о том почему эти силы такие, а не другие, как эти силы возникают и как они передаются в пустом пространстве. Внешняя сторона тех же самых явлений совсем по-другому описывается теорией гравитации Эйнштейна с заменой сил притяжения понятием искривления пространства-времени. Однако, и это, оставаясь только способом внешнего описания явления на некотором информационном уровне, не затрагивает глубинных вопросов “почему” и “как”. К сожалению, построить более глубокие теории очень трудно, а может быть во многих случаях и невозможно. Таким образом, “мечты об окончательной теории” [7], объясняющей все в нашем мире, еще очень долго, а может быть и всегда, будут оставаться только мечтами.

В качестве определенной иллюстрирующей аналогии рассмотрим следующий пример. Пусть кто-то не знакомый с программированием и вычислительной техникой хочет экспериментальным путем понять как работает компьютерная программа, например текстовый редактор. Проведя многочисленные эксперименты можно выяснить какие функции выполняет программа, как эти функции вызывать с помощью клавиатуры или мышки и как этими функциями пользоваться. Однако, скорее всего, всех возможностей текстового редактора экспериментальным путем выяснить не удастся. Например, может не удастся выяснить, можно ли с помощью программы вставлять в текст математические формулы. Тем не менее, на основе экспериментов можно построить достаточно хорошее описание текстового редактора, причем не только на уровне таблицы, связывающей возможные действия и их результаты, но и на уровне каких-то обобщений. Все в совокупности экспериментатор сможет назвать теорией конкретного текстового редактора и эта теория, даже не отвечая на все возможные вопросы, как и любая хорошая теория, будет обладать значительной предсказательной силой, может быть отнесена в разряд “превосходных” и с успехом использоваться самим экспериментатором и другими людьми в практической работе.

В то же время, построенная теория будет относиться только к внешнему информационному уровню исследуемого объекта. Идя вглубь проблемы с целью получить ответы на вопросы “как” и “почему” можно было бы разбираться с компьютерной программой редактора. При этом нужно было бы каким-то образом получить текст программы, нужно понять общую теорию программирования, понять язык, на котором написана программа, выяснить функционально-логическую схему программы и назначение отдельных операторов.

Переходя к более глубокому информационному уровню можно было бы пытаться исследовать систему команд компьютера, т.е. исходные компьютерные (машинные) коды, а заодно разобраться и с программой трансляции текста программы в эти коды с языка более высокого уровня. Можно, в принципе, опуститься и на более глубокие информационные уровни, такие как общие принципы вычислительной техники и их конкретная реализация на логическом уровне и на уровне физических электронных элементов, из которых строится память, логико-арифметическое устройство и устройство управления компьютера.

Идя еще дальше можно прийти от полупроводниковых элементов к электрону и к фундаментальной физической проблеме – что такое электричество? Исследуя и эту проблему можно придти к следующему, наверняка не лучшему и не полному, но типичному определению: электричество это электродвижущая сила, возникающая в электрической цепи при определенном пересечении ее проводниками силовых линий магнитного поля и приводящая к направленному движению электронов в замкнутой цепи, или к возникновению разности потенциалов между концами разомкнутой цепи.

Это внешнее определение, построенное по обычной схеме “если-то”, не отвечает на вопросы “почему” и “как”, т.е. не решает проблемы понимания природы электричества. Опускаясь еще глубже можно прийти, например, к заменяющей представления об эфире идее черной материи, заполняющей все пространство вселенной,

не взаимодействующей с известными физическими объектами, но ответственной за полевые взаимодействия и передачу гравитационных и магнитных сил. Но даже при экспериментальном подтверждении подобной теории вопросы “почему” и “как” останутся, так же как останутся и стремления исследователей опускаться на еще более глубокие уровни познания.

Возвращаясь к началу поставленной экспериментальной проблемы – исследованию программы текстового редактора, можно сказать, что условный экспериментатор, ничего не знающий о программировании и вычислительной технике, вынужденно остановится на первом внешнем информационном уровне описания. Единственно что он сможет сделать дальше – это поставить вопрос: как все это происходит? Получить ответ на этот вопрос, идя в рассмотренном случае обычным путем науки: эксперимент – теория – эксперимент, т.е. от эксперимента с программой к теории более глубокого уровня, нельзя.

Точно так же, модель мира, включающая разнообразные научные теории, относится только к внешнему информационному уровню, т.е. к уровню, на котором возможно проведение прямых наблюдений или специально поставленных экспериментов. Любая научная теория должна нести в себе информацию, которую можно было бы хотя бы в принципе проверить. То же самое можно сказать и о той части модели мира, которую каждый человек строит сам в процессе прямого экспериментального взаимодействия со своей средой обитания.

### **Информация и потребности.**

Выше говорилось о целенаправленности и непрерывной активности на уровне внешнего поведения и восприятия (человека, животного). Целенаправленность и активность создаются и поддерживаются отражающимися в мозге живого организма возрастающими потребностями, которые удовлетворяются в целевых ситуациях взаимодействия со средой. Кроме того, животные и человек входя в контакт со средой активно и целенаправленно строят и корректируют ее отображение в мозге. Что заставляет мозг делать и эту работу?

Естественно пока, до рассмотрения активных моделей нейронных механизмов мозга, условно предположить, что одной из потребностей животного, в том числе и человека, является потребность в получении информации для построения в мозге модели проблемной среды, (В дальнейшем это предположение будет несколько уточнено). Без модели проблемной среды невозможно не только свободное поведение, но даже и упрощенное поведение, например, связанное с жизнью в клетке или тюремной камере. Для построения модели проблемной среды необходимо получение информации, причем не информации вообще, а информации конкретной, определяемой первичными и вторичными потребностями и целями поведения.

Многие исследователи считают получение информации (часто сводимое к простому любопытству) одной из генетически заданных первичных потребностей живого организма, такой же как потребности в пище, воде, сне и т.п.. Приведем в связи с этим описание эксперимента В.С.Ротенберга и В.В.Аршавского, цитируемое по [8].

“ Крыс помещали в комфортабельную камеру, где им создавались “санаторные” условия, возможность удовлетворения всех потребностей, жизнь безо всяких забот. В одной из стенок камеры была дверь, которая вела в необжитое и не исследованное темное помещение, таившее в себе опасность самой неизведанностью. После относительно короткого периода освоения комфортабельной камеры крысы одна за другой начинали предпринимать попытки проникнуть в это необследованное помещение. Это было отнюдь не праздное спокойное любопытство. Крысы осторожно продвигались по темному коридору, проявляли все признаки страха, - у них дыбилась шерсть, усиливалось мочеиспускание, учащался пульс. Они эпизодически в быстром темпе возвращались назад и тем не менее вновь и вновь пускались в свое рискованное и ничем не спровоцированное путешествие.”

Мы связали информацию со строящейся в мозге моделью проблемной среды. С точки зрения задач поведения и восприятия первичной потребностью, в принципе, может быть не потребность в информации как таковой, а потребность в специфической информации, необходимой для построения модели проблемной среды. Живой организм направленно и активно входит во взаимодействие со средой для удовлетворения всех своих потребностей, в том числе и для получения информации, необходимой для построения и уточнения модели среды. При этом нет информации вообще. Информация выделяется на ограниченных информационных срезах, и служит для построения или уточнения модели проблемной среды. Слово “проблемной” здесь имеет важное значение, поскольку именно проблемы, определяемые первичными и вторичными потребностями животного (в том числе и в первую очередь человека), определяют исследуемые им информационные срезы.

Какой вывод можно сделать из всего этого? Являются ли генетически заданными потребность в получении информации или потребность в построении модели проблемной среды? Получение информации (чистое любопытство) само по себе первичной потребностью животного или человека не является. Постоянная необходимость в построении уточнении и коррекции модели проблемной среды внешне очень похожа на генетически заданную первичную потребность. Однако, необходимость построения модели - это слишком общая формулировка. Все потребности живого организма абсолютно конкретны и удовлетворяются в результате конкретного целенаправленного поведения.. Должна быть конкретной и потребность, связанная с получением и обработкой информации о среде.

Забегая далеко вперед, сформулируем следующее гипотетическое представление: одна из генетически заданных потребностей - это оптимизация функционального состояния нейронов и нейронной сети в целом. Одним

из средств оптимизации является гашение очагов возбуждения, образующихся в мозге как при возникновении первичных и вторичных потребностей, так и при недостаточности информации для решения задач поведения и восприятия. Возникновение очагов возбуждения приводит к нарушению нормальной работы мозга по оптимизации его функционального состояния. Гашение таких очагов возбуждения происходит, в частности, при построении, достройке или коррекции фрагментов нейронной модели проблемной среды, необходимой для целенаправленного поведения в процессе удовлетворения первичных и вторичных потребностей.

## Лекция 10. **Нейронные модели**

Если придерживаться высказанной ранее точки зрения, состоящей в том, что мышление это не только способы решения специфических задач, но и, в первую очередь, специфическая нейронная реализации этих способов решения, включающая способность строить и использовать активную динамическую модель среды, то для того, чтобы понять что такое мышление нужно понять не только функциональное содержание мышления, но также устройство и принципы работы мозга.

Первые модели нейронных сетей Мак-Каллока и Питса, также как и перцептрон Розенблата, разрабатывались и представлялись авторами как модели механизмов мозга. В современных формальных распознающих нейронных сетях теоретический аспект часто оттесняется на второй план. На первый план вышла прикладная ориентация – распознавание образов. Однако и об этих распознающих нейронных сетях часто говорят как о моделях мозга. Попробуем разобраться в том, насколько оправданы такие представления.

### **Свойства и функции нейронных моделей, необходимые для решения базовых задач мышления**

Итак, для того, чтобы моделировать мозг нужно понять не только то, что мозг делает, но и как он это делает. И то и другое одинаково важно.

Ранее мы сформулировали предположения о том, что делает мозг для решения многоэкстремальной задачи поведения – строит модели среды, имеющие уровни обобщений и укрупнений. Это позволяет сводить многоэкстремальные задачи поведения к одноэкстремальным и обеспечивает тем самым возможность пользоваться локальными правилами принятия решений, основанными на эмоциональной оценке вариантов поведения в ситуациях альтернативного выбора.

Также ранее на примере системы машинного зрительного восприятия мы сформулировали предположения об устройстве целостных моделей объектов среды и основных элементов этих моделей, то есть укрупненных и обобщенных структурных описаний. Были также высказаны предположения о роли предвидения в процессе восприятия и описана программная реализация целостного активного зрительного восприятия на основе выдвигания и целенаправленного подтверждения гипотез об объектах восприятия, представляемых полными, целостными, отображаемыми структурно-метрическими описаниями.

Кроме того, отмечались проблемы, связанные с необходимостью описания взаимодействия между частями и целым и сформулированы общие необходимые свойства модели проблемной среды применительно к задачам управления поведением и зрительного восприятия. Были также упомянуты некоторые результаты классических работ в области нейрофизиологии, относящиеся, главным образом, к поведению животных.

На основании всего этого с учетом представлений о накапливающейся неустойчивости и принципа устойчивого неравновесия Э.С.Бауера можно определить свойства и функции, которыми должны обладать моделируемые нейронные механизмы мозга. Представляется, что приводимые ниже свойства, как минимум, являются необходимыми, хотя и возможно, что часть этих свойств может оказаться только полезными, но необязательными. Вопрос о достаточных свойствах остается пока еще открытым.

Итак, сформулируем представляющиеся необходимыми свойства и функции нейронных механизмов мозга. В мозгу должна строиться и использоваться при поведении и восприятии модель проблемной среды. Поэтому, естественно, список необходимых свойств и функций механизмов мозга в значительной степени совпадает с представлениями об основных необходимых свойствах и функциях модели проблемной среды.

1. Для решения задач восприятия и поведения мозг должен при обучении строить нейронное отображение среды. Нейронная модель среды субъективна, то есть она строится с позиций субъекта восприятия и поведения и включает информацию, необходимую для решения задач восприятия и поведения.
2. Нейронная модель среды должна быть активной и моделировать изменения в среде, как зависящие, так и не зависящие от субъекта восприятия и поведения. Должно осуществляться опережающее моделирование, т.е. предвидение.
3. В нейронной модели должна содержаться и использоваться информация, необходимая для реализации поведения, оптимального по критерию  $\max T$  и основывающегося на предвидении, построении в воображении и эмоциональной оценке вариантов поведения.

4. В нейронной модели механизмов мозга должны быть средства для построения целостных структурных обобщенных отображений классов объектов, определяющих, с одной стороны, свойства целого, и с другой стороны, состав и свойства частей целого и отношения между ними. Объект в нейронной модели это не обязательно реальный физический объект среды. Объектами в нейронной модели могут быть также ситуации в среде и действия, изменяющие ситуации.
5. Нейронная модель механизмов мозга должна иметь средства для формирования иерархии, то есть должна быть возможность построения объектов разного уровня укрупнения и обобщения. Частями (элементами) объектов каждого уровня укрупнения должны являться объекты более низкого уровня. Основная цель построения уровней укрупнения – это сведение многоэкстремальных задач поведения к одноэкстремальным.
6. В нейронной модели должно осуществляться моделирование переходов как по вертикали – между уровнями укрупнения, так и по горизонтали – между ситуациями и объектами одного уровня.
7. В нейронной модели механизмов мозга должны быть средства для формирования ассоциаций по смежности во времени и реализации на этой основе предвидения и преадаптации.
8. Среда не является статической, поэтому нейронная модель среды также не должна быть статической. Нейронная модель должна быть активной и “жить” по законам среды.
9. В нейронной модели среды должно строиться отображение себя и эмоционально окрашенное отображение взаимодействия себя со средой.
10. В нейронной модели в пределе должно реализовываться абстрактное мышление. Для моделирования мышления человека необходимо использование вербального уровня, то есть необходимо построение понятийной словестно-логической модели;
11. Наконец, будет очень хорошо, если в модели будут реализовываться подсознательные процессы и элементы творчества.
12. Кроме того, было бы хорошо, а может быть даже и принципиально, если бы нейронные модели объясняли эмоциональные реакции, в том числе и такие неспецифические, как реакции на музыку, абстрактную живопись, свет костра или морской прибой.

Эти предположения о необходимых свойствах и функциях нейронных моделей механизмов мышления, по-видимому, могут быть использованы как при построении определения мышления, так и в работах по моделированию механизмов мозга. В то же время, сами по себе эти предположения недостаточны для построения исчерпывающего определения мышления, поскольку не определены достаточные свойства механизмов мышления. Недостаточны эти предположения и для создания нейронных моделей механизмов мозга. Нехватает предположений о том, как мозг все это делает. Однако, по всей видимости, без учета этих предположений строить полные нейронные модели механизмов мозга нельзя.

Уже говорилось о том, что не все перечисленные свойства могут оказаться необходимыми. Кроме того, первичная нейронная модель может не включать всех, даже необходимых, свойств полной модели. В то же время, представляется, что нейронная модель механизмов мозга, как минимум, должна отвечать на вопросы: как строится включающее целостные обобщенные и укрупненные представления объектов иерархическое многоуровневое отображение среды, и как на этой модели среды разворачиваются процессы предвидения и оценки предвидимых ситуаций.

Модель мозга не должна быть простым пассивным дешифратором, формирующем выход, как функцию входа. Модель мозга должна быть активна и обеспечивать “умозрительное моделирование” процессов в среде и результатов взаимодействий субъекта поведения со средой. Именно это позволит говорить, что мышление это не только способы решения специфических задач, но и, в первую очередь, специфическая нейронная реализации этих способов решения.

Таким образом, одно из выделенных свойств нейронных механизмов мозга, а именно активность, является, по-видимому, важнейшим и, безусловно, необходимым для моделей мозга. Ниже в последующих лекциях это свойство рассматривается более подробно.

### **Минимальные сведения об элементах мозга**

Как подойти к построению моделей нейронных механизмов мозга? Естественнее и проще всего было бы делать это, имея достаточную нейрофизиологическую информацию. Однако, такой информации нет и, к сожалению, по-видимому, никогда не будет. Нейрофизиология как и 100 лет назад остается преимущественно эмпирической наукой. О том, как работает отдельный нейрон известно все или почти все. Однако, в нейрофизиологии нет законченной теории, а следовательно и понимания того, как работает целый мозг в процессе мышления.

Ситуацию отчасти иллюстрирует бывшее популярным в шестидесятые годы прошлого века и остающееся актуальным до сих пор следующее шуточное сравнение. Представим себе, что на земле есть изолированная цивилизация, незнакомя с телевидением. и что живущие там люди пытаются понять как работает попавший к ним телевизор. Проводя всевозможные эксперименты они поймут как включать телевизор, определяют назначение клавиш панели управления, научатся переключать программы и регулировать громкость. Затем, изучая устройство телевизора, экспериментаторы проследят и установят все электрические связи, начнут делать послойные срезы телевизора и рассматривать их в микроскоп и так далее. Можно придумать и еще много других тонких исследований, однако ясно, что таким путем понять как работает телевизор нельзя. Нужен целостный подход и построение на этой основе теории работы телевизора.

Придумать и экспериментально проверить теорию работы мозга, повидимому, должны и могут не физиологи, а, скорее, математики и инженеры. При этом, естественно, придумывая теорию работы мозга нужно максимально использовать существующую обширную экспериментальную нейрофизиологическую и психологическую информация.

Основная минимальная информация об устройстве головного мозга, которую должен знать инженер, пытающийся понять и смоделировать мышление, была известна еще в первой половине прошлого века. Состоит эта информация в следующем.

Подавляющее большинство физиологов считает, что мышление реализуется объединенными в сеть нервными клетками – нейронами, которых в мозге человека около 14 миллиардов. Крайне незначительное меньшинство предполагает, что в мышлении помимо нейронов участвуют и нейроглиальные клетки, которых на порядок больше и в среду которых погружены нейроны. По основной версии нейроглиальные элементы обеспечивают для мозга только опорную и трофическую функции.

Обычно пишут, что нейрон имеет много входов (до нескольких сотен) и один выход. В отношении выходов это неточно. Отходящая от нейрона одна связь ( аксон ) чаще всего разветвляется и эти разветвления ( коллатерали ) заканчиваются на разных нейронах. Это полностью эквивалентно тому, что от нейрона отходит много связей, передающих информацию о возбуждении нейрона одновременно по многим адресам.

По каждому входу на нейрон может приходиться дискретный возбуждающий или тормозящий электрический импульс от другого нейрона или рецепторной (сенсорной) клетки, например, от клетки сетчатки глаза. Амплитуда и полярность электрических импульсов, передаваемых по конкретной связи постоянны. Может меняться частота импульсов. Окружающая клетку мембрана нейрона имеет исходный потенциал, поддерживаемый “натриевым насосом”. Приходящие на нейрон импульсы создают возбуждающие или тормозящие потенциалы, которые суммируются (со знаком) на мембране нейрона увеличивая или уменьшая ее исходный потенциал, т.е. производя поляризацию или деполяризацию мембраны. Чаще всего считается, что происходит как пространственная, так и временная суммация накапливающихся на мембране дополнительных потенциалов. С течением времени если не происходит возбуждения и разряда нейрона, то накопленный на мембране нейрона дополнительный потенциал поляризации или деполяризации уменьшается ( по-видимому экспоненциально ).

Полярность всех импульсов при их передаче по связям одинакова. Импульс приобретает знак, т.е. возбуждающую или тормозящую полярность проходя через соединение между входной связью и телом нейрона. Это соединение называется синапс. Большинство физиологов считает, что синапс определяет не только знак но и величину передаваемого нейрону потенциала. Это связывается с изменением проводимости синапса. Часть нейрона, на которой расположены синапсы, называется дендрит.

Нейрон является дискретным пороговым элементом. Когда суммарный накопленный на мембране нейрона потенциал поляризации достигает некоторого порога, в нейроне происходит разряд. В результате разряда по отходящей от нейрона связи ( аксону ) передается незатухающий по длине аксона импульс.

Различные аксоны могут иметь разную толщину и существенно разную длину. Скорость передачи импульса по аксону составляет в зависимости от толщины волокна примерно от 1 до 100 метров в секунду. Аксон на конце обычно разветвляется на коллатерали, заканчивающиеся синапсами на дендритах других нейронов. В результате разряда нейрона накопленный на нем потенциал обнуляется.

После разряда нейрона наступает период абсолютной рефрактерности длительностью около 1 миллисекунды, когда нейрон невозбудим. Затем следует период относительной рефрактерности длительностью несколько миллисекунд, в течение которого порог возбудимости нейрона снижается до нормального уровня, после чего следует экзальтация – кратковременное снижение порога возбудимости ниже нормального уровня. Следует заметить, что мозг включает нейроны многих типов, отличающиеся по своим свойствам. Поэтому периоды рефрактерности и экзальтации у каких то нейронов могут иметь свою специфику.

Существуют не имеющие значительного распространения и поддержки гипотезы состоящие в том, что взаимодействие между нейронами осуществляется не только по связям, но и посредством передачи полей возбуждения и торможения.

Вот, собственно и все. Можно пробовать строить нейронные модели. Для того, чтобы попытаться как то по минимуму представить себе как работает мозг и попробовать смоделировать эти представления информации достаточно. Можно конечно предположить, что в мозгу есть еще что то, не открытое пока еще физиологами. Допуская это и строя нейронную модель можно, конечно, придумывать недостающие необходимые свойства, но

делать это нужно только в том случае, если эти свойства действительно необходимы для работы модели. Ну и, конечно же, нужно стараться, чтобы нефизиологичность вводящихся таким образом новых свойств не была чрезмерной.

Ограничиваясь при моделировании мышления приведенной информацией о нейронах и нейронной сети можно, главным образом, варьировать всего четыре вещи. Первое – это связи между нейронами, их топология, а также законы или правила формирования связей и их изменения. Второе – это правила формирования и изменения синаптических проводимостей связей. Третье – это правила изменения порога срабатывания нейрона. И четвертое – правила формирования результирующего потенциала нейрона. Точными сведениями обо всем этом нейрофизиология не располагает.

Конечно, профессионалы физиологи могут сказать, что приведенная информация о мозге неполна и в чем-то неточна. Это не страшно. Приведенный перечень свойств элементов мозга конечно же является открытым. В то же время, вряд ли к приведенным свойствам можно добавить многое, что будет иметь решающее значение. Кроме того, если не придерживаться тезиса, что мышление – это функция мозга и только мозга, то, изобретая мышление, не обязательно, ориентируясь в целом на мозг, пытаться его слепо копировать. Можно добавить элемент фантазии.

Многочисленные известные, а может быть и еще неизвестные, свойства нейрона часто разбивают на две группы. Первая группа – это свойства, определяющие внешнюю логику работы нейрона и логику работы нейронной сети. Вторая группа – это свойства, обслуживающие “собственные нужды” нейрона как живой клетки. Разделить эти группы можно попытаться с позиций целостных представлений на основе теории работы мозга. Правда, может оказаться и то, что с позиций целостных представлений о работе мозга эти группы свойств во многом неразделимы. Одновременно теория и успешные нейронные модели механизмов мозга могут стимулировать целенаправленные нейрофизиологические эксперименты. Таким образом, целостность и целенаправленность так же важны при проведении научных экспериментов, как и при распознавании образов.

Приблизительно в рамках приведенной информации о нейронах и нейронной сети для объяснения работы мозга были предложены различные модели, являющиеся в подавляющем большинстве производными от модели логических нейронных сетей Мак-Каллока и Питса, либо модификациями перцептрона Розенблата. Вам уже читались лекции о формальных нейронных сетях. Теперь еще раз посмотрим на искусственные нейронные сети и попробуем оценить их свойства и возможности на основе информации о необходимых свойствах нейронных механизмов мышления и сведений о работе нейронов мозга.

### **Логические нейронные сети Мак-Каллока и Питса**

В 1943 году Мак-Каллоком и Питсом была опубликована статья - A logical calculus of the ideas immanent in nervous activity ( Логическое исчисление идей, относящихся к нервной активности ). В этой статье была предпринята попытка описать работу нейронных механизмов мозга, используя аппарат математической логики.

Мак-Каллок и Питс не ставили своей целью решать какие-либо прикладные задачи. Их модель была чисто теоретической попыткой представить мозг как совокупность связанных в сеть логических элементов. К этому подталкивало то, что живой нейрон похож, а при определенных ограничивающих предположениях становится логическим пороговым элементом, имеющим фиксированное число двоичных входов и двоичный выход.

Свойства нейронов и нейронной сети определялись следующими аксиомами.

1. Возбуждение нейрона соответствует принципу ‘все или ничего’.
2. Время делится на дискретные моменты - такты. Возбуждение нейрона в какой-то момент времени происходит, если в предшествующий момент времени произошли возбуждения определенного фиксированного числа синапсов. Это число не зависит ни от предыдущей активности, ни от расположения синапсов на нейроне.
3. Возбуждение по связи от одного нейрона к любому другому происходит без задержки ( за один такт ).
4. Синапсы могут быть как возбуждающими, так и тормозящими. Входной сигнал, прошедший через тормозящий синапс, абсолютно исключает возбуждение данного нейрона в рассматриваемый момент времени.
5. С течением времени структура сети не изменяется.

Удовлетворяющие приведенной аксиоматике формальные нейроны и нейронные сети Мак-Каллока и Питса обладают следующими свойствами. Во-первых, доказано, что эти формальные нейроны могут реализовать любую двоичную логическую функцию. Во-вторых, доказано, что из этих формальных нейронов можно построить сеть, реализующую любое высказывание логики высказываний. Что из этого следует? Пересказывая фон Неймана, можно сказать, что если функционирование системы заключается в формировании реакций на ее выходе в зависимости от входов, то с помощью формальной нейронной сети Мак-Каллока и Питса в этой системе возможна реализация любого такого функционирования, которое можно точно и однозначно описать конечным числом слов.

Модель мозга Мак-Каллока и Питса вызвала массу возражений, как со стороны физиологов и психологов, так и со стороны инженеров. Мы не будем подробно рассматривать эти возражения. Приведем лишь слова

Ф.Розенблатта: “Требовать точного знания логической структуры нервной сети для предсказания ее поведения равносильно тому, чтобы требовать знания точного положения и скорости каждой молекулы находящегося в сосуде газа для предсказания его температуры ”

Предложенная Мак-Каллоком и Питсом модель, конечно же, не является моделью мозга. В этой модели не реализуется ни одно из сформулированных нами ранее необходимых свойств моделей мозга. Даже если предположить, что большинство из этих свойств не являются необходимыми, то остается необходимым, по крайней мере, одно – активность. Модель мозга не может быть пассивным логическим дешифратором, жестко связывающим вход с выходом. Кроме того, вызывает, по меньшей мере, сомнение физиологичность большинства принятых Мак-Каллоком и Питсом аксиом о свойствах нейронов и нейронной сети.

Тем не менее, работа Мак-Каллока и Питса была очень полезна. Во-первых, эта работа стимулировала многие целенаправленные нейрофизиологические исследования, касающиеся как свойств отдельного нейрона, так и сети в целом, например, хотя бы, работы по поиску энграмм памяти.

Во-вторых, эта работа послужила определенным толчком к развитию аппарата математической логики. Очевидная недостаточность классической математической логики для описания работы мозга потребовала разработки различных неклассических логик, например таких, в которых делаются попытки формализации ассоциаций, аналогий, индукции, учет времени и так далее. Стимулировались и работы по созданию различных формальных систем представления знаний.

И, наконец, в-третьих, нейронная модель Мак-Каллока и Питса вызвала как разработку ее различных модификаций, так и создание иных нейронных моделей, например, таких как перцептрон Розенблатта или современные формальные нейронные сети, широко применяющиеся на практике для распознавания образов.

## **Перцептрон Ф. Розенблатта**

В 1958 году Ф.Розенблатт предложил нейронную сетевую модель механизмов зрительного анализатора мозга, которую он назвал “Перцептрон”. Слово перцептрон является производным от английского perception, или латинского perceptio – восприятие.

Также как и нейронные сети Мак-Каллока и Питса, перцептрон, во всяком случае в начале, был чисто теоретической моделью, хотя и строился как автомат, реализующий в перспективе такую практически значимую функцию как распознавание зрительных образов. Однако, возможность эффективного практического применения перцептрона в начальных авторских вариантах всерьез не рассматривалась.

Исходная задача состояла в том, чтобы объяснить как сеть из нейроноподобных элементов может обучаться и распознавать подаваемые на вход сети изображения. В подавляющем большинстве случаев все физические предположения о свойствах нейронов и нейронной сети реализовывались программно, и все выводы основывались на теоретических построениях и моделировании на компьютере. Как Розенблаттом так и другими авторами рассматривались многие разные варианты перцептронов. Один из вариантов простого трехслойного дискретного перцептрона устроен следующим образом.

Изображение фиксируется на входном прямоугольном дискретном поле - растре  $m \times n$  (аналогия – сетчатка глаза). В каждой точке входного поля расположен дискретный чувствительный (сенсорный) элемент, передающий дискретный сигнал по отходящим от него связям, если яркость изображения в соответствующей точке входного поля больше некоторого порога. Обычно этот сигнал может принимать только два значения 0 или 1. Входной уровень перцептрона называется сенсорным или уровнем  $S$  – элементов.

Сигналы от  $S$  – элементов передаются по связям на входы  $A$  – элементов или элементов ассоциативного уровня. Связей, отходящих от каждого  $S$ -элемента много. Распределяются эти связи на уровне  $A$ -элементов случайно.  $A$  – элемент это формальный нейрон, то есть дискретный элемент, имеющий порог срабатывания, входы, связанные с  $S$  – элементами, и выходные связи, идущие к  $R$  – элементам (см. ниже). В отличие от формального нейрона Мак-Каллока и Питса формальный нейрон в некоторых вариантах перцептрона имеет собственный вес. Время дискретно, т.е. делится на моменты, синхронизирующие работу всех элементов. Ассоциативный нейрон срабатывает, если сумма сигналов на его входах превышает порог. Пороги срабатывания всех нейронов равны и постоянны. В варианте, когда нейрон имеет свой вес, величина выходного сигнала прямо зависит от веса нейрона. В этом случае вес нейрона – величина переменная, изменяющаяся при обучении. В ином варианте нейрон не имеет вес и величины сигналов, передаваемых нейроном по связям на другие нейроны, зависят от проводимостей связей. В этом случае при обучении меняются коэффициенты передачи или проводимости связей.

Сигналы от нейрона ассоциативного уровня передаются по связям на входы  $R$  – элементов или элементов уровня реакций.  $R$  – элемент это формальный нейрон, то есть дискретный элемент, имеющий порог срабатывания, входы, на которые поступают сигналы от  $A$  – элементов, и два состояния 0 и 1. Порог  $R$  – элемента обычно равен 0. Если алгебраическая сумма сигналов на входе  $R$  –элемента больше 0, то элемент находится в состоянии 1, в противном случае элемент находится в состоянии 0. Каждый  $R$  –элемент имеет имя, соответствующее имени одного из классов распознаваемых объектов. Имена присваиваются  $R$ -элементам при обучении. На выходе системы как результат ее работы выбирается имя того из возбужденных  $R$ -элементов, сумма входных возбуждений которого, поступающих от  $A$ -элементов максимальна.



Связи от S – элементов к A – элементам и от A – элементов к R – элементам распределены случайно. Законы распределения варьируются в экспериментах. В простом дискретном перцептроне, ассоциативные элементы которого имеют свои индивидуальные веса, коэффициенты усиления (веса) всех связей могут принимать только два значения + 1 и – 1. Соответственно, величины воздействий, передаваемые по связям на нейроны, пропорциональны весам передающих нейронов и могут быть возбуждающими либо тормозящими. В варианте простого дискретного перцептрона, когда нейроны не имеют индивидуальных весов, связи между нейронами могут иметь различные изменяющиеся возбуждающие (положительные) или тормозящие (отрицательные) индивидуальные веса. Вариант перцептрона, имеющего изменяющиеся при обучении проводимости связей является основным.

Исследовавшиеся варианты перцептрона существенно отличались друг от друга. Эти отличия определяются, во-первых, принятой топологией связей между S-элементами и нейронами ассоциативного слоя (A-элементами), топологией связей между A-элементами и нейронами результата (R-элементами), а также взаимодействием между нейронами внутри ассоциативного слоя. Это взаимодействие реализуется в перцептроне с перекрестными связями, то есть в перцептроне, имеющем связи между A-элементами. Рассматривались также четырехслойные перцептроны, имеющие два слоя ассоциативных элементов.

Кроме того, варианты перцептронов отличались правилами изменения весов проводимостей связей и/или весов нейронов. Обучение классического перцептрона организуется как обучение с учителем и состоит в следующем. На сенсорный слой проектируется изображение объекта, принадлежащего какому либо классу из обучающей выборки. Если известно имя соответствующего этому классу R-элемента и на него приходят возбуждающие воздействия, то поощряются активные связи, идущие от S-элементов к A-элементам, участвующим в возбуждении этого R-элемента и связи, идущие от этих A-элементов к R-элементу. Одновременно может делаться ‘наказание’ остальных связей. (В другом варианте перцептрона “поощрялись” и “наказывались” не связи, а A-элементы).

Поощрение состоит в добавлении фиксированной положительной величины к весу связи или к весу нейрона, в зависимости от модификации перцептрона. Если происходит наказание элементов и связей, не участвующих в возбуждении нужного R-элемента, то из их весов вычитается постоянная величина. Если R-элементы еще не поименованы, то при обучении предварительно имя класса, соответствующее имени распознаваемого объекта, присваивается тому R-элементу, который имеет в данном такте наибольшую алгебраическую сумму входных воздействий.

При обучении в классическом перцептроне происходит приблизительно следующее. A-элементы разделяются на группы связанные преимущественно возбуждающими связями со “своим” R-элементом и тормозящими с остальными. S-элементы, возбуждающиеся разными изображениями одного класса связываются возбуждающими связями с одной группой A-элементов. Таким образом происходит как бы объединение на уровне A-элементов разных изображений одного класса, в том числе и изображений, имеющих значительные различия или вообще не пересекающихся на сенсорном уровне. Одновременно на уровне A-элементов происходит разделение изображений разных классов, в том числе и изображений, пересекающихся на сенсорном уровне. Описанные объединения и разделения не имеют жесткого однозначного характера. Это определяется начальной случайной топологией связей, продолжительностью обучения и представительностью обучающей выборки. Поэтому при воспроизведении опыта может возникнуть элемент случайности, усиливающийся при подаче на сеть случайных воздействий.

Перцептрон разрабатывался для непосредственного распознавания изображений объектов, поступающих в зрительный анализатор в виде растра. Используемые для такого распознавания эталоны также должны быть изображениями или их препаратами. Формируемые при обучении перцептрона эталоны распознаваемых классов являются подмножествами ассоциативных элементов, соответствующих точкам входных изображений. При этом проекция изображения с сенсорного уровня на уровень A-элементов не имеет геометрического смысла, поскольку связи между этими уровнями распределены случайно.

Инвариантность растрового распознавания по отношению к масштабу, сдвигу или повороту изображения может достигаться подгонкой изображения к эталонам при помощи их нормализации путем масштабирования, сдвига или поворота. Этих операций нейронная сеть перцептрона не делает. Вместо этого в перцептроне, так же как в multifont системах распознавания печатных символов, для каждого класса изображений может создаваться много эталонов – по одному на каждый значительно отличающийся пример изображения из обучающей выборки этого класса. В этом случае отдельным эталоном можно считать группу усиленных связей, идущих от группы S-элементов к группе A-элементов. При объединении этих эталонов на одном подмножестве A-элементов может возникать “проблема разделяющего ИЛИ”.

Поясним это на примере. Обучим перцептрон, предъявляя ему одну горизонтальную линию, а затем несколько сдвинутую вниз или вверх другую. В режиме экзамена при подаче на вход любой из этих линий будет возбуждаться одна и та же группа A-элементов и один и тот же R-элемент, соответствующий ответу “горизонтальная линия”. Затем подадим на вход обе эти линии и обучим перцептрон знаку =. Теперь для правильной работы нужно, чтобы ответ “горизонтальная линия” возникал при подаче на вход первой горизонтальной линии ИЛИ второй, но не двух одновременно.

В перцептроне с перекрестными связями выделенные при обучении группы А-элементов могут образовывать ансамбли нейронов, объединенных взаимовозбуждающими связями В таких системах возможно распознавание по сходству, когда распознаваемое изображение не совпадает в точности ни с одним из эталонов, но с каким-то из них имеет не очень большие допустимые различия.

Перцептрон Розенблатта проверялся, главным образом, на изображениях горизонтальных и вертикальных линий. Делались так же попытки отличить изображения треугольника и квадрата.

Мы называем перцептроном распознаватель, на входной растр которого проектируются точки, соответствующие распознаваемым объектам. Это не обязательно должны быть точки изображения. Это могут быть и описывающие объект независимые двоичные признаки.

Как теоретический анализ, так и анализ результатов экспериментов позволяет сделать вывод, что предложенный Ф.Розенблаттом нейронный анализатор (классификатор) растровых изображений не обладает свойством обобщения по примерам, то есть распознает недостаточно хорошо. По оценкам многих исследователей перцептрон не является и удовлетворительной моделью мозга. С подробным анализом возможностей в первую очередь трехслойных перцептронов можно ознакомиться по работе М.Минского и С.Пайперта [11].

Перцептрон разрабатывался как модель мозга. Модель, являясь каким-то приближением, в то же время, не описывает объект в точности. Однако, модель, как минимум, должна давать язык или схему, в рамках которой было бы можно обсуждать моделируемое явление в каких то по возможности общих и существенных его аспектах. С этой точки зрения с определенной значительной натяжкой перцептрон, так же как и нейронную сеть У.Мак-Каллока и В.Питса можно считать начальными моделями мозга, однако, нужно отметить, что эти модели не обладают ни одним из выделенных нами необходимых свойств. В первую очередь, эти автоматы являются пассивными дешифраторами вход – выход, то есть не обладают активностью.

В то же время, перцептрон Ф.Розенблатта, так же как и работа У.Мак-Каллока и В.Питса, дал значительный толчок к развитию многочисленных теорий построения распознающих автоматов, строящихся как сети из формальных нейронов. Современные ансамблевые растровые распознаватели (перцептроны) принципиально отличаются от классического перцептрона только топологией связей между нейронами и принятым способом обучения нейронной сети. Один из наиболее распространенных алгоритмов обучения предложен Хопфилдом [22].

Следующим шагом стали современные формальные нейронные сети, реализующие признаковое распознавание.

## **Лекция 11. Признаковые распознающие нейронные сети**

Современные признаковые распознающие нейронные сети (ПРНС), также как и растровые нейронные распознаватели, можно считать развитием перцептрона Ф.Розенблатта. Часто между ними вообще не делают различия, называя перцептроном любой распознаватель, построенный с использованием формальных нейронов. Будем проводить условную границу между перцептроном и ПРНС, называя перцептроном растровый распознаватель, а ПРНС – признаковый. (С оговоркой, что фиксированное множество независимых двоичных признаков фактически является растром). Можно считать, что предпринятая Розенблатом попытка получить обобщения и эффективное распознавание изображений в формальной нейронной сети с растровым входом без предварительных преобразований входной информации не удалась. Это послужило толчком к развитию признаковых нейронных распознавателей, в том числе и распознавателей, работающих с многоградационными зависимыми признаками.

Необходимость или, по крайней мере, целесообразность изменения модели Розенблатта путем перехода от задач распознавания растровых изображений к задачам распознавания по признакам была понятна многим исследователям. При разработке перцептрона первичная цель состояла в том, чтобы понять как работает мозг в процессе распознавания зрительных образов, спроектированных на входной растр зрительного анализатора. При создании ПРНС на первое место вышла практическая цель – распознавание.

Формальную распознающую нейронную сеть часто рассматривают как черный ящик, имеющий вход, выход и какое-то почти таинственное, строго не определяемое внутреннее функционирование. Попробуем настолько, насколько это возможно, заглянуть внутрь черного ящика. О логике работы классического перцептрона мы говорили в предыдущей лекции. Приблизительно такая же логика должна работать во всех растровых ансамблевых распознавателях. Теперь рассмотрим совсем другую логику, а именно логику работы признакового нейронного распознавателя.

### **Расознавание с помощью разделяющих гиперплоскостей**

Итак, на вход ПРНС подается не поточечный растр, а в общем случае многоградационные признаки, причем признаки любой модальности, то есть распознаваться, как и в любой признаковой распознающей системе, могут не только изображения, а все что угодно, например, звуки, болезни по их симптомам, геологические объекты и так далее. Поэтому в публикациях часто приводят очень широкие списки областей применения ПРНС. При этом нужно понимать, что широта области применения относится не к распознающим нейронным сетям, а к распознаванию по признакам. Одна область применения отличается от другой не какой-то принципиальной

спецификой работы нейронной сети, а, в первую очередь, алфавитом, особенностями и способом получения используемых для распознавания признаков.

Что касается живого мозга, то информация в зрительный анализатор коры головного мозга передается и как растровая проекция и как признаки. Первичное получение простейших зрительных признаков происходит в рецептивных полях, - связанных с сетчаткой глаза специальных нейронных сетях. Кроме того, зрительный анализатор мозга устроен, по-видимому, иерархически с последовательным усложнением анализируемых признаков. Проблема получения признаков, тем более формирования их иерархии, в современных распознающих нейронных сетях обычно не рассматривается.

Формальный нейрон – это пороговый элемент, имеющий число входов, равное числу участвующих в распознавании признаков и один выход. Признаки на входах могут иметь много градаций. Каждая входная связь имеет свой вес. Веса связей могут быть как положительными, так и отрицательными. Классический формальный нейрон считает взвешенную алгебраическую сумму значений входных признаков и сравнивает ее с фиксированным порогом. Если эта сумма меньше порога, то выход нейрона равен – 1, если больше порога, то + 1. Это иллюстрирует сплошная линия на рис. 1

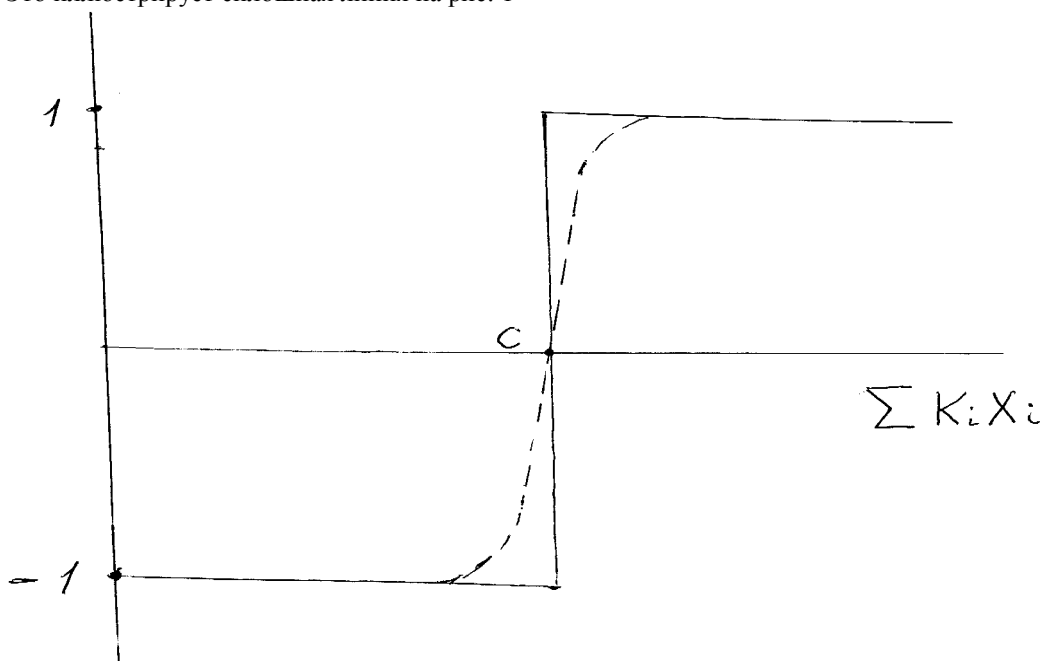


Рис.1

Условие срабатывания нейрона:

$$\sum K_i X_i > C \quad (1)$$

где  $X_i$  значение признака,  $K_i$  – весовой коэффициент,  $C$  – порог срабатывания. Величины признаков и коэффициентов могут иметь как положительные, так и отрицательные значения. Уравнение (1) с заменой знака  $>$  на знак  $=$  определяет гиперплоскость в пространстве признаков (при двух признаках – прямую линию).

Каждому распознаваемому объекту соответствует вектор в пространстве признаков. В соответствии с уравнением (1), нейрон строит гиперплоскость в пространстве признаков и определяет по какую сторону от этой плоскости лежит точка, соответствующая вектору признаков распознаваемого объекта. Как проходит гиперплоскость зависит от коэффициентов  $K$  и порога  $C$ .

Часто, рассматривая условие срабатывания нейрона - уравнение (1), забывают о том, что формальный нейрон просто строит разделяющую гиперплоскость в пространстве признаков и говорят о весе  $i$ -го признака, о вкладе признака в сумму  $K_i X_i$  о различной значимости признаков в зависимости от их веса и тому подобное. По-видимому, может вводить в заблуждение термин вес признака. В действительности  $K_i$  это не вес, а просто коэффициент в уравнении разделяющей гиперплоскости, определяющий ее положение в пространстве признаков.

Для распознавания объектов, относящихся к двум классам, может оказаться достаточным один нейрон. Это возможно в случае, когда в пространстве признаков существует такая гиперплоскость, что все точки, соответствующие векторам признаков объектов первого класса, лежат по одну сторону от этой гиперплоскости, а все точки, отображающие объекты второго класса по другую. Этот случай иллюстрирует рис.2А

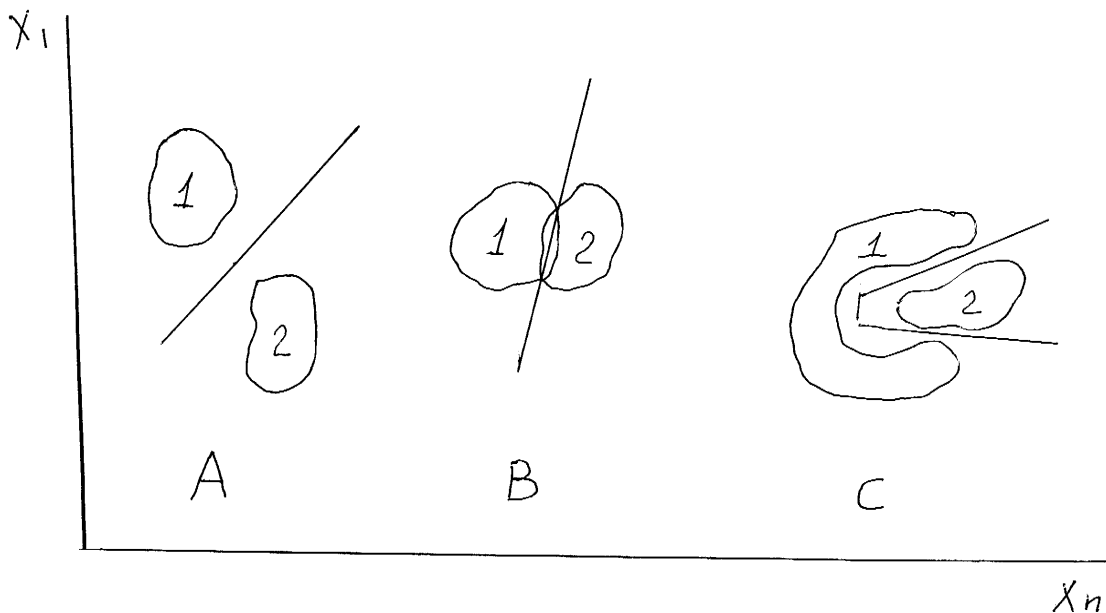


Рис.2

Качественно определить подобный случай можно следующим образом: точки, отображающие объекты каждого класса в пространстве признаков, расположены компактно, поверхности, огибающие множество точек каждого класса допускают, в принципе, выпуклые не пересекающиеся аппроксимации.

Задача построения разделяющей гиперплоскости состоит в определении коэффициентов  $K$  и порога  $C$ . Решается эта задача при обучении с использованием большого числа примеров – обучающих выборок. Поскольку при обучении могут использоваться только конечные обучающие выборки, то даже в простом случае, подобном приведенному выше, решается не задача абсолютного разделения, а лишь задача максимизации вероятности правильного отнесения неизвестного распознаваемого объекта к одному из классов или, что то же самое, задача минимизации вероятности ошибки. Эта задача решается не строго и тем лучше, чем представительней обучающая выборка и чем выше ее статистическая достоверность.

Проведение разделяющей гиперплоскости, то есть нахождение необходимых для наилучшего разделения классов параметров  $K$  и  $C$  нейрона, усложняется, если поверхности, огибающие множество точек каждого класса пересекаются. Такой случай показан на Рис. 2.В.

Как видно из иллюстрируемого рисунком примера, провести разделяющую плоскость так, чтобы точки, представляющие в обучающей выборке объекты первого класса, лежали по одну сторону от плоскости, а второго класса по другую, в этом случае невозможно.

Тем не менее, задача решается эвристически с использованием, по-прежнему, только одного нейрона. При этом при нахождении величин  $K$  и  $C$  можно строить целевую функцию, вычисляемую на всех точках обучающей выборки и зависящую, например, от числа ошибок на обучающей выборке, либо от каких то функций расстояний от точек обучающей выборки до разделяющей плоскости.

Еще усложним задачу. Рис. 2.С иллюстрирует случай, когда поверхности, огибающие множество точек обучающей выборки каждого из двух распознаваемых классов, не выпуклы и расположены так, что проведение разделяющей плоскости невозможно. Однако, в этом случае существует разделяющая поверхность, которую можно попытаться аппроксимировать плоскостями. Для аппроксимации нелинейной разделяющей поверхности с использованием нейронов нужно несколько нейронов, по числу аппроксимирующих плоскостей.

Предположим, что для обучающей выборки, состоящей по-прежнему из объектов только двух классов и ‘невыпуклого’ случая, иллюстрируемого рис.2.С удалось при обучении получить значения  $K_i$  и  $C_i$  для всех аппроксимирующих гиперплоскостей (нейронов). Назовем эти нейроны нейронами первого уровня. Теперь для получения результата распознавания необходим еще один - второй уровень, объединяющий результаты первого уровня. В случае распознавания объектов двух классов этот уровень должен состоять из одного нейрона, выход которого  $+1$  или  $-1$  указывает к какому из двух классов относится распознаваемый объект. Нейрон второго выходного уровня строит разделяющую плоскость в пространстве значений выходов нейронов первого уровня.

На втором нейронном уровне могут возникнуть проблемы, связанные с необходимостью реализации в некоторых случаях ‘исключающего ИЛИ’. Снять эту проблему можно обычным программным путем, либо путем предшествующей обучению кластеризации обучающего множества. При этом число распознаваемых на первом нейронном уровне классов увеличивается на число дополнительно вводимых кластеров.

Классический формальный нейрон, реализуя уравнение (1), строит разделяющую плоскость в  $n$  – мерном пространстве признаков (при  $n = 2$  прямую линию). Уйти от необходимости аппроксимировать плоскостями

нелинейную разделяющую поверхность и вводить дополнительный результирующий или собирающий нейронный слой можно изменив уравнение нейрона. Например, нейрон может строить не уравнение гиперплоскости, а уравнение  $n$  – мерного эллипсоида ( на плоскости – эллипса ), то есть считать взвешенную сумму квадратов значений признаков. В принципе, на формальном нейроне, имеющем  $n$  входов, можно строить любую разделяющую поверхность, определяемую любой вычислимой функцией  $n$  – переменных. Правда, в этом случае может значительно усложниться задача обучения, т.е. построения уравнений разделяющих поверхностей, и усилиться сомнение в правомочности использования слова нейрон.

До сих пор мы рассматривали задачу распознавания объектов, относящихся к двум классам. Задача распознавания в случае, когда число классов больше двух, например  $m$ , сводится к  $m$  задачам с двумя классами. В каждой из этих  $m$  задач должна рассматриваться задача распознавания с двумя классами, когда первый класс - это один из  $m$  классов основной решаемой задачи распознавания, а второй класс – это объединение всех остальных  $m-1$  классов.

В рамках приведенной выше логики с использованием в явном виде разделяющих гиперплоскостей в пространстве признаков можно строить распознающие системы, не апеллируя к нейронам и нейронным сетям. Особенно просто и естественно это делать при распознавании двух классов объектов. Именно так организован и работает дифференциальный уровень системы распознавания символов печатного текста – FineReader, а также дифференциальные уровни систем распознавания рукописных текстов FineReader – рукопись и FormReader . На этих уровнях распознавателей текстов при обучении ищутся аналитические зависимости, задающие разделяющие гиперплоскости в пространстве признаков. Это же можно реализовать и на формальных нейронах. При этом в распознавании ровным счетом ничего не изменится.

Отметим, что работая с разделяющими гиперплоскостями или, что то же самое, с формальными нейронами, нужно иметь в виду, что мерность пространства признаков должна быть фиксирована и не меняться как при обучении, так и при распознавании разных объектов. Например, нельзя, построив при обучении разделяющую плоскость в трехмерном пространстве признаков, уменьшить мерность пространства признаков и пытаться распознавать, используя не плоскость, а разделяющую прямую линию. В связи с этим может возникать определенная проблема, поскольку отдельные признаки могут быть разрушены и невычислимы на конкретных объектах распознавания. При обучении такие случаи должны идентифицироваться, а такие объекты можно попросту пропускать. При распознавании такие объекты тоже, конечно, можно пропускать, то есть не распознавать этим методом. Но можно и пытаться их распознавать, делая с целью сохранения мерности разделяющей гиперплоскости определенные специальные действия для замещения отсутствующих признаков некоторыми фиктивными “безвредными” значениями.

На рисунке 1. сплошной линией показана выходная характеристика классического формального нейрона. В современных ПРНС обычно используются нейроны, у которых пороговый скачок от  $-1$  к  $+1$  заменяется плавным, дифференцируемым, чаще всего сигма-образным переходом, показанным на рис.1 пунктиром.

Такой нейрон по-прежнему строит разделяющую гиперплоскость. Во всех точках пространства признаков, лежащих по одну сторону от разделяющей плоскости выход нейрона равен положительной величине, по другую сторону – отрицательной.

Введение в формальном нейроне вместо порогового скачка плавного дифференцируемого перехода ничего принципиально в распознавании не меняет, а лишь позволяет использовать различные градиентные алгоритмы обучения сети, то есть алгоритмы нахождения коэффициентов в уравнениях разделяющих гиперплоскостей. Кроме того, плавная выходная характеристика нейрона дает возможность получать оценку уверенности распознающей системы в результате распознавания. Эта оценка используется в некоторых системах распознавания при формировании результата. В то же время, нужно понимать, что эта оценка, строго говоря, не является ни вероятностью правильного распознавания, ни расстоянием до разделяющей гиперплоскости. Правда, ни из чего не следует, что евклидово расстояние до разделяющей гиперплоскости в пространстве признаков всегда является наилучшим или просто хорошим критерием для получения количественной оценки “уверенности” системы в правильности результата распознавания. Не исключено, что в каких-то конкретных случаях строить разделяющие поверхности в пространстве признаков и определять в этом пространстве расстояния лучше в какой-нибудь другой метрике.

Существующие многочисленные варианты признаковых распознающих нейронных сетей в подавляющем большинстве случаев принципиально отличаются друг от друга только используемыми градиентными алгоритмами обучения, то есть алгоритмами нахождения коэффициентов (весов входов нейронов) и свободных членов в уравнениях разделяющих гиперплоскостей.

## **Формальная нейронная сеть как модель механизмов мозга**

Таким образом, признаковая распознающая нейронная сеть – это пассивный признаковый распознаватель, обладающий всеми отмеченными нами недостатками признаковых распознавателей. Напомним, что эти недостатки, в первую очередь, определяются последовательной организацией процесса распознавания, отсутствием целостности восприятия, целенаправленности и активности. Еще одним важнейшим недостатком

признаковых распознавателей, и в частности ПРНС, является то, что эти системы работают только в фиксированном пространстве признаков, одинаковых для всех объектов распознавания.

В то же время, признаковая распознающая нейронная сеть не является и серьезной нейронной моделью мозга. Не является она и удовлетворительной моделью специфических нейронных механизмов зрительного анализатора мозга. В этом плане признаковая распознающая нейронная сеть ничем не отличается от логической нейронной сети Мак-Каллока и Питса и перцептрона Розенблатта. Все эти нейронные модели работают по схеме стимул – реакция. При этой схеме невозможна организация целенаправленного (двунаправленного) процесса восприятия, идущего одновременно снизу-вверх, т.е. от изображения к пониманию, и сверху вниз, т.е. от понимания к изображению. Эти модели не имеют и других приведенных нами функций и свойств, необходимых для построения модели среды, эффективной в отношении задач восприятия и управления поведением.

Иногда говорят о принципиальной возможности построения в признаковой нейронной модели описаний целостных объектов с использованием разделяющих поверхностей в пространстве каких-то признаков или характеристик. Эта концепция представляется малопродуктивной и не физиологичной по трем причинам. Первая причина состоит в том, что описания целостных объектов в модели среды должны быть независимы и описывать объекты как таковые, а не строиться на функции разделения. Независимость описаний позволяет отличать объект от любых других, как известных (описанных в модели), так и от неизвестных.

Вторая причина состоит в том, что разделяющие поверхности можно строить только в фиксированном пространстве характеристик, одинаковых для всех объектов восприятия.. При этом все характеристики (признаки) должны использоваться при построении описаний (разделяющих поверхностей) всех объектов. Понятно, что это абсолютно нефизиологично и принципиально отличается от рассмотренного нами способа представления в модели объектов восприятия на основе индивидуальных целостных независимых описаний. И, наконец, главная причина состоит в том, что описания объектов с использованием строящихся в сетях из формальных нейронов разделяющих гиперплоскостей не подходят для реализации целенаправленного процесса “распознавания с пониманием”.

Остановимся на вопросе о физиологичности и некоторых других функций, решаемых сетью из формальных нейронов. Как уже отмечалось, в основе работы ПРНС лежат две функции. Первая это построение на нейронах разделяющих гиперплоскостей в фиксированном пространстве признаков. Вторая - построение иерархии распознавателей, где распознаватель более высокого уровня строит разделяющие плоскости в пространстве некоторых характеристик, являющихся производными от результатов работы распознавателя более низкого уровня. Реализуются ли подобные свойства в мозге?

Предположение о том, что реальные нейроны строят разделяющие гиперплоскости в фиксированном, одинаковым для всех распознаваемых объектов пространстве каких-то признаков или характеристик представляется весьма сомнительным, или, как минимум, требующим подробного рассмотрения и определения места этой если и существующей, то безусловно частной функции в общей картине работы мозга. Весьма сомнительной является возможность формирования величин синаптических проводимостей отдельных нейронов (коэффициентов в уравнениях разделяющих гиперплоскостей) путем решения на сети общей (градиентной ?) задачи, типа задачи поиска этих коэффициентов методом обратного распространения ошибки. Более чем сомнительной является привязка каких либо логических функций к отдельному нейрону, в чем убеждают результаты большого количества нейрофизиологических работ, посвященных поиску энграмм памяти.

О необходимости иерархической организации строящейся в мозге модели среды мы говорили неоднократно. Однако, в ПРНС возникает очень специфическая иерархия, определяемая первым свойством, то есть тем, что нейроны строят разделяющие гиперплоскости в пространстве признаков или каких-то характеристик и необходимостью объединения этих результатов.. Эта иерархия не является иерархией типа часть-целое. Не является она, скорее всего, и иерархией типа частное-общее. Переход от частного к общему реализуется в ПРНС еще в первом слое при построении разделяющих гиперплоскостей. В многослойных ПРНС на верхнем уровне к этому, по-видимому, добавляется только объединение результатов нижнего уровня за счет реализации простых функций, близких по смыслу логическим функциям И и ИЛИ. Отличие иерархии, реализуемой в ПРНС от иерархии в строящейся в мозге модели проблемной среды состоит еще и в том, что при распознавании процессы в ПРНС разворачиваются только в одну сторону – снизу вверх. В формальных распознающих нейронных сетях не просматривается возможность создания иерархической модели среды, включающей действия и процессы, определяющие переходы между ситуациями. Последнее представляется принципиально необходимым функциональным свойством моделей мышления

В добавление к отмеченной функциональной нефизиологичности уточним в чем состоит нефизиологичность и некоторых конкретных свойств приведенных выше формальных нейронных моделей. В этих моделях носителями информации являются отдельные нейроны. На отдельных нейронах реализуются логические функции в сети Мак-Каллока и Питса, отдельные нейроны формируют выход в перцептроне и строят разделяющие плоскости в признаковых распознающих нейронных сетях. Разрушение одного единственного нейрона в этих системах должно приводить к полному нарушению их работы. В живом мозге ничего подобного не происходит.

Уже отмечалось, что разработка формальных моделей нейронных сетей подтолкнула физиологов к проведению многих направленных экспериментов и, в частности, многочисленных экспериментов по поиску так

называемых энграмм памяти. В этих работах ставилась задача соотнесения конкретной запоминаемой информации или функции с конкретным нейроном или местом в коре головного мозга. Эксперименты проводились на больных с частичными повреждениями мозга и на животных с экспериментальным разрушением отдельных нейронов или компактных областей коры головного мозга. Найти энграммы памяти не удалось. Локальные разрушения приводят не к забыванию конкретного объекта или к нарушению конкретной функции, а, например, к ухудшению качества распознавания многих объектов или к снижению качества выполнения разных функций.

Вывод, делающийся во многих работах, состоит в том, что единицей переработки и хранения информации в мозге является не отдельный нейрон, а ансамбль нейронов, объединенных взаимно возбуждающими связями. При этом нейроны одного ансамбля распределены не локально, а в некоторой области пространства коры головного мозга. К аналогичным выводам приходят и при построении некоторых растровых нейронных моделей.

Приведем и другие отличия. В формальных признаковых распознающих нейронных сетях нейрон подсчитывает взвешенную сумму входов и сравнивает ее с порогом. Это значит, что при решении задачи признакового распознавания на каждый вход нейрона поступает значение соответствующего этому входу признака, которое умножается на некоторый определенный при обучении постоянный коэффициент. Значение признака зависит от того, что поступило на вход распознающей системы и может быть различным. В реальном нейроне в качестве коэффициента связи (входа) может рассматриваться величина синаптической проводимости. Второй сомножитель для конкретной связи является постоянным по амплитуде и как значение признака интерпретироваться не может. Сигналы, приходящие на нейрон по конкретной связи, могут иметь разную частоту. Это конечно может использоваться для построения логики какой-то распознающей нейронной схемы, но это будет уже совсем другая логика и совсем другие нейроны, т.е. не формальные нейроны, подсчитывающие взвешенную сумму значений признаков (входов), а что-то иное.

В формальных распознающих нейронных сетях, так же как в перцептроне и логических сетях Мак-Каллока и Питса, время разбивается на такты, задержки при передаче сигналов по связям отсутствуют. Отсутствует временная суммация возбуждающих и тормозящих воздействий. Живой нейрон это дискретный пороговый элемент, работающий по принципу “все или ничего”. Введение в формальном нейроне плавной, например сигма-образной выходной характеристики делает элемент непрерывным. Понятия порог возбуждения (срабатывания), разряд, выходной импульс, серия импульсов для такого формального нейрона теряет смысл.

Порог возбуждения (срабатывания) живого нейрона не является постоянным, а изменяется во времени в зависимости от предыстории срабатываний. Сразу после разряда нейрона возникают динамические составляющие порога – рефрактерность и экзальтация. Благодаря этому результатом возбуждения нейрона может быть как одиночный импульс, так и серия импульсов. Есть данные о том, что в зависимости от предыстории срабатываний изменяется и статическая составляющая порога возбуждения нейрона.

И, наконец, необходимо еще раз отметить главное отличие описанных выше формальных нейронных моделей от живого мозга, отличие, состоящее в отсутствии синергичности и активности. Об этих свойствах живого мозга мы будем подробно говорить в дальнейшем.

## **Нейрокомпьютер**

До сих пор мы говорили о перцептронах и признаковых распознающих нейронных сетях. К этому часто добавляется понятие нейрокомпьютер. Термин нейрокомпьютер не очень удачен. Системы, к которым он обычно применяется, вычислительными машинами не являются.

Нейрокомпьютеры в каком-то приближении можно разделить на четыре группы. Первая группа это просто уже упоминавшиеся формальные признаковые распознающие нейронные сети.

Вторая группа это (по нашей классификации) перцептроны – растровые или ансамблевые системы. Возможности этих систем, как в части распознавания образов, так и в части решения каких-то других задач пока еще весьма ограничены.

Третья группа это системы, построенные на формальных нейронах, но решающие специальные задачи. Это могут быть задачи, близкие по своему смыслу задаче распознавания. Например, кластеризация. Это могут быть и совсем другие задачи. Общим в системах этой группы является то, что в них используются формальные нейроны, строящие, чаще всего, разделяющие гиперплоскости в многомерном пространстве каких либо векторов, характеристик или признаков. С помощью гиперплоскостей и их объединения в конкретных задачах специальным образом организуется отделение или выделение каких-то входных или промежуточных векторов. На основе этого выделения и формируется результат. Специфика в этих системах и их особенности определяются как самой задачей, т.е. тем, что нужно выделить, так и способом выделения нужных объектов. В общем случае специфическими решаемыми проблемами в этих системах являются задача автоматического нахождения коэффициентов в реализуемых нейронами уравнениях и задача формирования общего результата на основе использования результатов работы нескольких (многих) нейронов. Последнее, так же как и при распознавании решается введением дополнительных нейронных слоев и специфической организацией связей между нейронами.

При большом числе классов входных объектов и/или при их сложном устройстве и соответственно при их сложном отображении в пространстве признаков современная формальная нейронная признаковая система строится без представления о разделяющих гиперплоскостях и, вообще, без какого-либо представления об

алгоритме ее работы, а как “черный ящик”. Работа такой системы вписывается в схему стимул-реакция. Это определяется с самого начала алгоритмом обучения формальной нейронной сети.

При обучении итерационным градиентным методом, например методом обратного распространения ошибки, настраиваются изменяемые параметры сети. Обычно это величины синаптических проводимостей связей и параметры переходных функций нейронов. Обучение проводится по формируемым учителем обучающим выборкам. Критерий итерационных изменений параметров сети это величина и направление изменений несоответствия между ответом системы и необходимым задаваемым результатом. Таким образом, с помощью итерационной часто очень длительной настройки нейронной сети достигается задаваемое соответствие между классами входных объектов и требуемым выходом системы. (Трудно представить, что нечто подобное может происходить в реальной нейронной сети.)

Что может демонстрировать такая система? Только распознавание сложных входных ситуаций и формирование реакций на эти ситуации. Иногда говорят: вот сделаем большую сеть, например из миллиарда нейронов, и она начнет проявлять элементы мышления. На самом деле такая, по сути автоматная система, сможет только обрабатывать большое число обучающих выборок и, соответственно, отвечать на большое число вопросов. В такой системе при обучении могут устанавливаться связи между сложными многопараметрическими классами входных стимулов и сложными многопараметрическими ответными реакциями. Однако, в такой системе нельзя реализовать активного целенаправленного процесса мышления и таких его видов как перцептивное мышление, когнитивное мышление, решение многоэкстремальных задач поведения, постановка и решение абстрактных задач, творческое мышление.

Четвертая группа нейрокомпьютеров – это специальные частные системы, строящиеся на особых элементах. Перечислять и пытаться обобщать эти изобретения не представляется необходимым, поскольку это могло бы увести нас слишком далеко в сторону от основной рассматриваемой темы. Отметим лишь, что чаще всего в основе специальных нейрокомпьютеров лежит какой-то особый способ распознавания образов. Безусловно, интересный оригинальный, тем более физиологичный, способ распознавания может быть шагом в направлении понимания работы, как отдельных механизмов мозга, так и мышления в целом. Но только первым, или одним из первых шагов. К сожалению, часто такой шаг объявляется последним, или одним из последних.

Оценивая современные нейрокомпьютеры как модели механизмов мозга можно повторить все то же самое, что уже было сказано при оценке формальных распознающих нейронных сетей. Таким образом, в отличие от распространенного мнения, создание как признаковых, так и идущих от перцептрона растровых (ансамблевых) распознавателей и строящихся на их основе устройств или программ, называемых нейрокомпьютерами, не является пока еще заметным шагом в моделировании нейронных механизмов мозга.

## Лекция 12. **Активные нейронные модели (А-сети)**

В этой лекции будет говориться еще о некоторых вариантах нейронных сетевых моделей. Среди этих вариантов есть как запрограммированные, так и умозрительные гипотетические. Объединяет все эти варианты общие многократно уже упоминавшийся свойства – **неустойчивость, активность и синергичность**. Представляется, что эти свойства определяют основное отличие активных живых систем от пассивных неживых.

### **Активность**

Полная модель нейронных механизмов мозга должна, в конечном счете, обладать всеми приведенными в предыдущей лекции свойствами. Однако, как уже отмечалось, наиболее принципиальным и самым важным из этих свойств представляется свойство активности. Другие важнейшие необходимые функциональные или рабочие свойства мозга могут оказаться в значительной степени производными от свойства активности. В связи с этим нужно ответить на два вопроса: что такое активность мозга и зачем мозгу активность.

Как уже говорилось, активность внутренней работы живой системы (ассимиляции и синтеза), внутреннего регулирования (гомеостаза), а также и активность внешней работы (поведения) нужны для существования и являются необходимым следствием организации живого. Особый интерес представляет очевидная активность процессов мышления. Каковы ее особенности, для чего она нужна и чем определяется на уровне механизмов работы мозга?

Мы думаем всегда. Бодрствующий мозг не может не думать. В мозге при решении каких-то поведенческих или абстрактных задач разворачивается непрерывный управляемый волей целенаправленный процесс мышления. При отсутствии решаемых задач и полном отдыхе во время бодрствования происходит непрерывный ненаправленный поток сознания, состоящий из ассоциируемых образов и ассоциируемых мыслей, не связанных ни с какой решаемой проблемой. Направление ненаправленного в целом потока сознания может изменяться волевым усилием.



Мыслящая система в отличие от пассивных автоматов, работающих по схеме вход-преобразователь-выход, постоянно активна. Активность должна входить в определение мышления. Можно сказать, что активность является необходимым свойством мышления. Активное действие - это действие, которое инициируется не извне, а собственными внутренними причинами. На уровне поведения причина активности живого организма определяется необходимостью непрерывного поддержания неравновесия.

Непрерывная активность мозга, причем, не только в процессе решения задач восприятия и поведения, но и при отдыхе требует объяснения. Возможно, что именно это поможет понять особенности живого мышления, определяющие его эффективность. Создание модели активной мыслящей системы, как на алгоритмическом уровне, так и на уровне нейронного моделирования, это большая и очень далекая от решения проблема.

Активность не просто сопутствует человеческому восприятию, обучению, управлению поведением и решению формальных задач, а является основным определяющим свойством этих процессов. Еще раз вспомним, что мы не видим и слышим, а смотрим и слушаем. Восприятие целенаправленно и активно. Оно основано на выдвижении и проверке гипотез и на акцепторе восприятия, что делает его сходным с целенаправленным поведением. При обучении мыслящая система целенаправленно входит во взаимодействие со средой для построения и проверки гипотез о свойствах среды и для построения на этой основе модели среды. При этом строящиеся информационные срезы определяются целями системы и решаемыми ею задачами. В свою очередь, модель среды не должна быть мертвым описанием, а должна "жить" по законам среды.

Активным является и процесс практического мышления при планировании и управлении поведением. Этот процесс активен не только потому, что он вызывается и управляется первичными или вторичными потребностями организма. Процесс практического мышления активен, поскольку он, в соответствии с теорией доминанты А. Ухтомского [26], вызывается стабильными очагами возбуждения в коре головного мозга и непосредственно направлен на их гашение. Точно так же, на гашение очагов возбуждения направлены процессы в рамках перцептивного и когнитивного мышления.

Относительно работы очагов возбуждения возможна следующая приблизительная интерпретация. Информационная единица в нейронной сети это ансамбль нейронов. Очаг возбуждения это постоянная или временная активация нейронов, входящих в нейронный ансамбль. Постоянная активация может вызываться постоянным внешним информационным воздействием или потребностью и соответствующей этой потребности доминантой и стабильным внутренним мотивационным возбуждением. Временная активация нейронов ансамбля может поддерживаться в течение какого-то времени и без внешних по отношению к нейронному ансамблю воздействий за счет положительных обратных связей между нейронами ансамбля.

Существование очага возбуждения сопровождается сначала слабой, но усиливающейся со временем отрицательной эмоциональной оценкой. Гашение очага возбуждения сопровождается сильной, а иногда и очень сильной положительной эмоциональной оценкой. Можно вспомнить удовольствие, получаемое от решения трудной задачи. Большой эмоциональный подъем происходит при завершении работы над художественным произведением.

Созданием и гашением временных очагов возбуждения можно, в частности, объяснить эмоциональное восприятие музыки и поэзии, когда в процессе восприятия произведения на основе предвидения возникают, фиксируются и гасятся при подтверждении временные очаги возбуждения. Модельные гипотетические предположения о возможных нейронных механизмах возникновения, поддержания и гашения очагов возбуждения могут быть самыми разными.

Мозг активен всегда. Нейроны постоянно возбуждаются даже при полном отдыхе и отсутствии внешней информации, т.е. возбуждений, поступающих от органов чувств. Нейроны мозга работают (возбуждаются) и во время сна. Естественно предположить, что постоянная работа нейронов, не только при решении каких-то задач восприятия или поведения, но и при полном отдыхе, а также и во время сна имеет какой-то приспособительный смысл.

## **Гипотеза Емельянова-Ярославского**

Попытка решить проблему активности нейронной сети на теоретическом уровне и на уровне программного моделирования нейронной сети была предпринята Л.Б.Емельяновым-Ярославским [9]. Примерный смысл основной гипотезы состоит в следующем. Каждый нейрон имеет изменяемый параметр, который характеризует его функциональное состояние. Назовем его  $Q$ . Значение этого параметра увеличивается при возбуждениях нейрона и уменьшается, если нейрон не возбуждается.

Параметр  $Q$  имеет фиксированное одинаковое у всех нейронов оптимальное значение  $Q_{opt}$ . Поддерживаться значение параметра  $Q$  нейрона на уровне  $Q_{opt}$  может только при оптимальной частоте возбуждений нейрона, а следовательно при оптимальной интенсивности приходящих на нейрон возбуждающих воздействий.

Таким образом, для поддержания оптимального состояния нейрон должен быть постоянно активен, т.е. возбуждаться. Возбуждение нейронов возможно только при поступлении на их входы сигналов от других нейронов или рецепторов. То есть необходимо взаимодействие (синергия). От параметра  $Q$  нейрона зависит изменение его порога срабатывания и изменение величин синаптических проводимостей связей.

Поддерживаться функциональное состояние сети на оптимальном уровне может только при равномерной работе всех нейронов. Отклонение функционального состояния нейронной сети от оптимального уровня происходит при нарушении равномерной работы сети, происходящем за счет возникновения очагов возбуждения, за счет активации нейронных ансамблей в процессе решения поведенческих или формальных задач, а также за счет внешних информационных воздействий. Нормализация функционального состояния сети происходит при гашении очагов возбуждения, при запоминании регулярных внешних воздействий (память), при отдыхе и, главным образом, во время сна за счет ритмического распространения автоволн возбуждения и вызываемой этим процессом равномерной активности всех нейронов сети.

( Последующий текст в скобках заменяет некоторые предположения Емельянова-Ярославского об эмоциональном центре. Среднее значение параметров  $Q$  нейронов определяет функциональное состояние сети в целом и является эмоциональной переменной, которая формируется в эмоциональном центре и влияет на его работу. Если в нервной системе действительно происходит передача в эмоциональный центр информации о функциональном состоянии всех нейронов коры, то эта передача возможна от каждого нейрона либо гумарально, т.е. через кровь, либо посредством полевой передачи с последующем суммировании в эмоциональном центре. При отклонении этой эмоциональной переменной от оптимума эмоциональный центр формирует потребность в ее нормализации. Способ удовлетворения этой потребности – гашение очагов возбуждения. Радикальный способ полного удовлетворения этой потребности – сон, при котором, как уже отмечалось, происходит распространение автоволн возбуждения по сети и равномерная работа всех нейронов).

Это довольно приблизительное, значительно сокращенное, дополненное и несколько откорректированное описание гипотезы Емельянова-Ярославского, но представляется, что основная идея, а именно существование эмоционально оцениваемого функционального состояния нейронов, изменяющегося в зависимости от интенсивности их работы, и необходимость постоянной совместной активности нейронов для поддержания функционального состояния на оптимальном уровне, в этой интерпретации сохраняется.

Гипотеза Емельянова-Ярославского не нашла ни поддержки ни просто понимания. Виноват сам автор, который заявлял, что эта гипотеза полностью объясняет работу мозга в процессе мышления, в том числе все интеллектуальные функции. Убедительно подтвердить это ему не удалось ни логически, ни на уровне экспериментальной модели. Это связано, во-первых, с тем, что предложенная гипотеза о свойствах нейронов и сети не является, как утверждал автор, достаточным условием возникновения мышления. В лучшем случае эта гипотеза характеризует лишь необходимое условие возникновения свойства активности головного мозга. Переход от свойства активности к рабочим функциям мозга может быть возможен, но гипотезой Емельянова-Ярославского напрямую не определен. Во-вторых, авторское изложение очень сильно загромождено выдвигаемыми на первый план и затрудняющими понимание сложными описаниями второстепенных, необязательных и иногда спорных частных механизмов.

Тем не менее, гипотеза Емельянова-Ярославского заслуживает серьезного внимания. Это единственная существующая идея, на основе которой можно пытаться объяснять и моделировать свойство активности нейронной сети. При этом надо понимать, что необходимое свойство систему в целом не объясняет. К необходимому свойству нужно еще что-то добавлять для того, чтобы все в совокупности стало достаточным как для объяснения, так и для моделирования.

На основе общих идей Емельянова-Ярославского строились и программно реализовывались разные экспериментальные модели. Аналитически описать работу нейронной модели нельзя. Для моделирования нужно задать свойства нейронов и связей между ними в виде рекуррентных соотношений, позволяющих пошагово вычислять текущие состояния всех нейронов и связей. Таким образом возможна программная реализация физической модели.

### **Вариант свойств нейронов экспериментальной модели (А-сети).**

Нейрон, как и любой целостный объект, в принципе, может полностью характеризоваться набором переменных, их текущими значениями, зависимостями между переменными и зависимостями переменных от времени и внешних воздействий. Применительно к модели нейронной сети формальное задание всех этих зависимостей и аналитическое описание на этой основе работы системы в целом не представляется возможным. Поэтому все переменные в описываемой модели определяются пошаговыми рекуррентными соотношениями. Время дискретно, т.е. разбивается на такты. В каждом такте просчитываются и изменяются все переменные (параметры) всех нейронов и связей. Таким образом строится программная модель нейронной сети, рассматриваемой как физический объект.

Принимается (постулируется), что для моделирования нейронных механизмов мышления достаточны объекты двух типов: нервная клетка (нейрон) и нервная связь. Переменные, характеризующие нейрон, это возбудимость, потенциал поляризации мембраны и функциональное состояние. Переменные, характеризующие связь, это ее проводимость, топология, знак и временная задержка передаваемого сигнала. Сведение всех свойств нейрона к этому ограниченному множеству упрощает и делает практически возможным компьютерное моделирование. В то же время это сопряжено с опасностью игнорирования каких-то свойств нейрона как живой

клетки, что в конечном счете может отразиться на принципиальных свойствах модели. Однако, представляется, что этот риск не очень велик, если в модели сохраняются главные свойства живого – неустойчивость и активность.

Приведем общие свойства нейронов экспериментальной А-модели. Точные количественные характеристики широко варьировались при проведении экспериментов.

Функциональное состояние нервной клетки зависит от предыстории ее возбуждений и выражается в модели некоторым параметром  $Q$ . Этот параметр имеет общее для всех нейронов оптимальное значение  $Q_{opt}$ . Параметр  $Q$  дискретно увеличивается при каждом разряде нейрона и монотонно уменьшается во времени у не срабатывающего нейрона. Из этого следует, что для оптимизации функциональных состояний нервных клеток необходимы срабатывания (разряды) нейронов с некоторой оптимальной частотой. Нейрон может сработать при поступлении на его входы внешнего возбуждения. Достижение оптимального в среднем функционального состояния нейронов по всей сети достигается только при их взаимодействии – взаимном возбуждении или торможении.

Нейрон представляет собой пороговый элемент с временной и пространственной алгебраической (т.е. с учетом знака) суммацией возбуждающих и тормозящих воздействий, переменной возбудимостью, рефрактерностью и экзальтацией после срабатывания.

Условие срабатывания нейрона  $W > U$ , где  $U$  – порог срабатывания,  $W$  – накопленный потенциал.

На нейрон может приходиться любое количество возбуждающих и тормозящих связей от других нейронов или рецепторов. От нейрона отходит один аксон, разветвляющийся на любое количество возбуждающих и тормозящих коллатералей, заканчивающихся на разных нейронах..

Накопленный на нейроне потенциал  $W$  определяется алгебраической суммой пришедших на нейрон возбуждающих (+) и тормозящих (-) воздействий. Накопленный потенциал у несрабатывающего нейрона экспоненциально уменьшается во времени.

$$W_{n+1} = W_n + \sum S_i a_i - w(W_n), \quad (13.1)$$

где:

$W_n$  – значение накопленного потенциала в  $n$ -ом такте,

$W_{n+1}$  – значение накопленного потенциала в  $n+1$ -ом такте,

$w(W_n)$  – слагаемое, определяющее уменьшение накопленного потенциала во времени,

$S_i$  – величина синаптической проводимости  $i$ -ой связи,

$a_i$  – признак активности  $i$ -ой связи в  $n+1$ -ом такте.

После срабатывания нейрона накопленный на нем потенциал становится равным нулю.

Порог срабатывания имеет две составляющие- статическую и динамическую:  $U = U_s + U_d$ . В рассматриваемом промоделированном варианте статическая составляющая порога срабатывания монотонно снижается у несрабатывающего нейрона (возбудимость повышается) и увеличивается на малую дискретную величину при каждом срабатывании (возбудимость снижается). Статическая составляющая порога срабатывания связана прямой линейной зависимостью с параметром  $Q$ , т.е.  $U_s = KQ$ .

Динамическую составляющую порога срабатывания иллюстрирует рис.1.

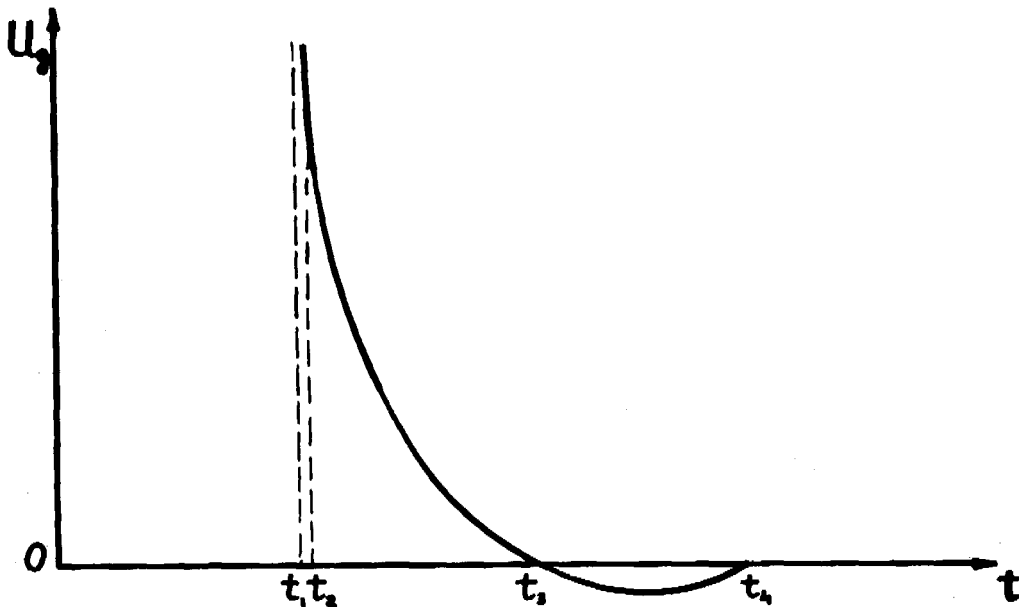


Рис..1.

Сразу после разряда порог срабатывания нейрона резко повышается, затем экспоненциально снижается до величины, меньшей чем начальное значение статической составляющей порога срабатывания ( экзальтация ) и затем повышается до величины, несколько превышающей начальное значение статической составляющей. Интервал  $t_1 - t_2$  соответствует абсолютной рефрактерности, когда разряд нейрона невозможен при любых внешних воздействиях. Интервал  $t_2 - t_3$  соответствует относительной рефрактерности, интервал  $t_3 - t_4$  соответствует экзальтации.

Связи между нейронами нужны для осуществления возбуждающих и тормозящих взаимодействий. От нейрона может отходить любое в пределах технических ограничений количество связей. Величина и знак передаваемого по связи потенциала определяются переменной величиной и знаком синаптической проводимости связи. Правила изменения синаптических проводимостей связей между нейронами наряду с правилами изменения порога возбудимости нейрона являются основой, определяющей работу нейронов и сети.

Синаптическая проводимость связи между  $i$ -м и  $j$ -м нейронами  $S_{ij}$  может “активно” изменяться при прохождении по этой связи сигнала. Приращение возможно только при условии срабатывания  $j$ -го нейрона. Если условий для “активного” изменения нет, синаптическая проводимость экспоненциально уменьшается, приближаясь к величине  $S_{min}$ . Знак и величина “активного” изменения синаптической проводимости связи зависит от сочетаний значений величин  $Q_i$  и  $Q_j$  (функциональных состояний) пары нейронов в момент прохождения сигнала по связи между ними.

Общий качественный смысл зависимостей, определяющих приращения синаптических проводимостей состоит в следующем. Величина положительного приращения возбуждающей синаптической проводимости связи от нейрона  $i$  к нейрону  $j$  при прохождении по ней сигнала прямо зависит от разности  $Q_{opt} - Q_j$ . Кроме того, это приращение связано прямой зависимостью с разностью  $Q_i - Q_j$ . Наоборот, приращение проводимости тормозящего синапса увеличивается при увеличении  $Q_j$  в прямой зависимости от разности  $Q_j - Q_i$ .

Таким образом, при прохождении сигнала по связи ее синаптическая проводимость может увеличиваться, уменьшаться или оставаться неизменной в зависимости от сочетания значений  $Q_{opt}$ ,  $Q_i$  и  $Q_j$ . При прохождении значения величины синаптической проводимости связи через нуль связь меняет знак, превращаясь из возбуждающей в тормозящую или наоборот. При моделировании рассматривался также вариант, в котором уменьшения значений синаптических проводимостей как возбуждающих, так и тормозящих связей ограничивались некоторой минимальной величиной  $S_{min}$ . В этом случае при  $S_{min}=0$  связь может уничтожаться, но менять знак не может.

Изменение величин синаптических проводимостей связей между нейронами направлены на оптимизацию функционального состояния сети в целом и объединение во взаимовозбуждающую группу одновременно срабатывающих нейронов.

Все идущие от одного нейрона связи передают сигнал с одинаковой задержкой после срабатывания нейрона. ( Вариант работы сети, в котором функциональную роль играют разные задержки сигналов, передаваемых по разным связям, мы рассмотрим в следующей лекции ).

Между одновременно возбуждающимися нейронами могут образовываться новые связи. Образование новых связей между нейронами происходит в соответствии с зависимостями, качественно совпадающими с зависимостями, по которым изменяются синаптические проводимости. Связи наиболее интенсивно образуются на

те нейроны, величины  $Q$  которых, т.е. функциональные состояния наиболее далеки от оптимальных значений. В области  $Q_{opt}$  условий для образования новых связей нет.

В модели как постулат вводится принципиальное понятие - “функциональное состояние нервной клетки” и соответствующий ему параметр –  $Q$ , имеющий оптимальное значение –  $Q_{opt}$ . Можно вслед за Емельяновым - Ярославским давать этому параметру гипотетическую не подкрепленную никакими экспериментами физиологическую интерпретацию, например, считая, что со значением параметра  $Q$  связана эффективность метаболизма (коэффициент полезного действия обменных процессов) нервной клетки.

Однако, это не обязательно. Например, для целей моделирования достаточно считать, что средняя величина параметров  $Q$  нервных клеток является одной из физиологических констант живого организма, таких как, например, температура тела или артериальное давление. Удерживается эта величина в нужных пределах как средствами внутренней регуляции (взаимовозбуждение нейронов), так и средствами удовлетворения потребностей путем внешнего поведения.

Приведенные общие свойства нейронов и нейронной сети и различные модификации этих свойств послужили основой для разработки, компьютерной реализации и экспериментальной проверки нескольких приблизительно одноплановых вариантов программных моделей. Разными в этих вариантах были выбираемые для моделирования конкретные свойства нейронов и связей между ними, а также исходная структура сети. Принципиальных отличий между этими вариантами нет. Эксперименты начинались еще в 60х годах прошлого века. Большая часть этих работ была прекращена по разным причинам. Главной причиной была недостаточная мощность существовавших тогда вычислительных средств. Полученные экспериментальные результаты могут претендовать только на иллюстрации некоторых особенностей активных нейронных сетей. Во-первых, была проиллюстрирована возможность формирования в нейронной сети на основе взаимодействия нейронов режима саморегуляции. Во-вторых, иллюстрируется возможность образования памяти внешних воздействий в виде нейронных ансамблей. .

Были получены разные экспериментальные варианты реализации режима общей регуляции на основе распространения равномерных волн возбуждения в нейронной сети. Приход внешних воздействий на сеть нарушал этот режим и при этом либо создавался ансамбль нейронов с взаимными возбуждающими связями (память), либо активировался существующий ансамбль.

При снятии внешних возбуждающих воздействий и восстановлении в сети режима общей регуляции в экспериментах возникали две проблемы.

Во-первых, по общей принятой логике нейроны существующих ансамблей должны были участвовать в процессе общей регуляции, т.е. возбуждаться при распространении по сети автоволн и при этом ансамбль как целое работать не должен, иначе эта работа будет мешать общему режиму равномерного распространения волн возбуждения. Во-вторых, режим общей регуляции не должен стирать следы памяти, т.е. изменять величины синаптических проводимостей связей между нейронами ансамбля.

И то и другое оказалось проблематичным. Устойчивое экспериментальное решение проблемы длительного сохранения следов памяти (нейронных ансамблей), не разрушающих режим авторегуляции и не разрушающихся этим режимом, получено не было. Радикальное решение этих проблем может быть получено в сети с возвратным возбуждением и возвратным торможением. Одновременно возникает возможность гипотетической интерпретации некоторых аспектов работы сознания.

### **А-сеть с возвратным возбуждением и возвратным торможением.**

Говоря о сознании, отмечаются многие важные вопросы, которые ставятся этой проблемой. На этих вопросах мы еще будем останавливаться. Интересно пока ответить и на более или менее второстепенный вопрос, почему сознание исчезает во время сна, хотя нейроны при этом интенсивно работают, что фиксируется как электроэнцефалограммой, так и прямой регистрацией импульсов. Каким образом сознание частично восстанавливается во время сновидений и полностью восстанавливается при пробуждении. Этот вопрос имеет отношение и к рассматриваемой проблеме активности.

И.П.Павлов говорил о сне как о разлитом торможении. Если это действительно так, то интересно понять что тормозится во время сна.

Вариант ответа на эти вопросы может дать гипотетическое представление о том, что смысловая единица в нейронной сети это не просто однослойный нейронный ансамбль, а двухслойная структура, в которой к нейронам основного слоя добавляется расположенные в смежном нейронном слое структуры возвратного возбуждения и возвратного торможения. Возможно, что этот слой может затормаживаться отдельно от основного нейронного слоя. Здесь можно предположить существование разных структурных вариантов.

Например, можно предположить, что внутри нейронного ансамбля специальные прямые взаимовозбуждающие связи между нейронами основного слоя не усиливаются и не образуются новые. При этом, необходимые положительные обратные связи между нейронами ансамбля реализуются через интернейроны обратного возбуждения. В этом случае если заторможен слой интернейронов,

нейронные ансамбли как целостные единицы не возбуждаются, соответственно нет и сознания, а регулирующие оптимальные состояния нейронов распространение автоволн возбуждения в основном слое нейронов во время сна может проходить равномерно, так, как будто ансамблей нет. В то же время распространение автоволн возбуждения в основном слое не стирает следов памяти, т.е. не разрушает образовавшиеся ансамбли.

Синхронное возбуждение нейронов ансамбля с участием возбуждающих интернейронов будет приводить к возникновению сознания и одновременно к нарушению процесса распространения волн возбуждения в основном слое нейронов и, тем самым, к нарушению процесса общей регуляции. Можно предположить, что это и происходит при внешнем возбуждении, при образовании очагов возбуждения и просто в процессе мышления. Можно также предположить, что тормозящие интернейроны обеспечивают прекращение работы нейронного ансамбля, т.е. гашение очага возбуждения.

Описанный гипотетический процесс в полном объеме не моделировался. Безусловно, возможны и различные другие схемные варианты реализации режима регуляции функционального состояния нейронов сети, а также и режимов создания и гашения очагов возбуждения. Это может быть очень не просто при переходе от плоской одноуровневой, в том числе и двухслойной, нейронной сети к многоуровневой иерархической нейронной модели.

Подводя некоторый итог, еще раз отметим, что в основе активности мышления может лежать возникновение и необходимость гашения очагов возбуждения как при обучении, так и в процессе работы. При этом гашение очага возбуждения становится потребностью. Однако, реализация свойства целенаправленной активности в искусственной нейронной сети даже с учетом гипотезы Емельянова-Ярославского и ее модификаций, в том числе и описанной выше схемы с возвратным возбуждением является значительной и еще весьма далекой от решения проблемой. Тем более, что как в существующих теоретических работах, так и в практически используемых формальных моделях нейронных сетей, эта проблема, также как и проблемы сознания и воли, не только не решается, но даже и не рассматривается.

Тем не менее, исходной идеей для моделирования активных механизмов мозга может стать представление о мозге как о непрерывно активной системе, построенной из неустойчивых элементов и обладающей внешней поведенческой и внутренней регуляционной активностью. Внешняя активность необходима для решения непрерывной задачи поведения и поддержания устойчивого неравновесия организма в целом. Внутренняя активность мозга – это не только активность метаболизма, но и активность синергического взаимного возбуждения нервных клеток. Эта активность необходима для поддержания оптимального состояния составляющих систему неустойчивых элементов – нейронов. Можно предположить, что с необходимостью поддержания оптимального состояния нейронов, а значит и с активностью, связаны и функциональные свойства мозга, в том числе построение модели проблемной среды, восприятие, управление поведением, решение практических и абстрактных, в том числе творческих задач.

Помимо основных функциональных или рабочих свойств мозг проявляет и специальные свойства и функции, присущие только живому мозгу: это интуиция, озарение или инсайт, творчество, эмоциональные реакции на музыку, абстрактную живопись ( т.е. живопись без смысловой нагрузки ), стихотворный ритм, красоту природных явлений. К этому перечню можно добавить еще любопытство, юмор, навязчивую идею, навязчивый мотив. Было бы хорошо, если бы нейронная модель объясняла и такие свойства. Не исключено, что именно активное устойчивое неравновесие нейронов и синергическая активность нейронной сети могут дать ключ к пониманию и этих свойств мозга. Некоторые интерпретации приводятся в дальнейшем изложении. Однако, подробная проработка и тем более экспериментальная проверка логики работы активных нейронных сетей, соответствующих принципу устойчивого неравновесия не проводились.

Однако, реализация свойства целенаправленной активности в искусственной нейронной сети в том числе с учетом гипотезы Емельянова-Ярославского и ее модификаций, например такой, как описанная выше схема с возвратным возбуждением, является значительной и еще весьма далекой от решения проблемой.

## **Преактивация и предвидение.**

В нейронной модели, обладающей свойством активности, должны происходить процессы преактивации нейронных структур, основанные на предвидении. В гипотетической полной иерархической многоуровневой модели предвидение может основываться на моделировании процессов в среде и процессов взаимодействия организма (автомата) со средой. В одноуровневой трехслойной сети предвидение, т.е. преактивация вторичных нейронных сетей, должно основываться, в первую очередь, на ассоциациях по смежности во времени поступления информационных воздействий.

В упрощенной системе, когда рассматривается только один уровень нейронной сети, эти ассоциации строятся на возбуждающих и тормозящих взаимодействиях между активируемыми в повторяющихся последовательностях вторичными сетями (ансамблями) . В прямом направлении времени возбуждающие связи между ансамблями нейронов увеличиваются, а тормозящие уменьшаются. В обратном направлении времени наоборот возбуждающие связи между ансамблями нейронов уменьшаются, а тормозящие увеличиваются.

В режиме “покоя” - при отсутствии внешних информационных воздействий в сети, подверженной равномерному случайному внешнему воздействию, если уровень возбудимости нейронов достаточно высок должен идти процесс поочередной активации вторичных сетей (ансамблей), направленный на поддержание оптимального функционального состояния нейронов. На необученной сети последовательность активации ансамблей случайна. После обучения, при котором на сеть подавались регулярные последовательности входных информационных воздействий, в режиме ‘покоя’ процесс становится не случайным – возникает тенденция возбуждения вторичных сетей в последовательности, определяемой обучением. .

Описываемая модель активной нейронной сети (А-сети) имеет общие черты с моделью Хебба [21]. В качественной модели Хебба постулируется, что одновременно возбуждающиеся нейроны связываются в функциональные структуры – нейронные ансамбли. Последовательно возбуждающиеся структуры связываются в “фазовые последовательности”. Эти процессы основываются на изменении связей между нейронами. Как спонтанная активация нейронного ансамбля или “фазовой последовательности”, так и их активация внешним стимулом вызывают соответствующий образ или ощущение.

Все это может быть отнесено и к А-сетям, но есть и важные принципиальные отличия. Нейроны в А-сетях это неустойчивые элементы, у которых изменяется функциональное состояние, а соответственно и возбудимость, в зависимости от интенсивности срабатываний. Это свойство нейронов является определяющим. Все что происходит в А-сетях, как на стадии обучения, т.е. в процессе изменения связей между нейронами, формирования нейронных ансамблей, формирования связей между ансамблями и т.д. , так и на стадии работы, включающей воспроизведение результатов обучения, все это должно подчиняться одному критерию – оптимизации функционального состояния нейронной сети. Из этого вытекают важнейшие свойства нейронной сети – синергичность и активность.

Логическая емкость приведенной выше схемы ассоциативных временных взаимодействий невысока, поскольку в этой схеме учитываются только смежные парные взаимодействия. Например, если входные информационные воздействия соответствуют буквам алфавита и система обучалась путем подачи на вход последовательностей букв, из разных текстов, то почти все парные вероятности очень быстро выравниваются и временных ассоциаций практически не будет. И, тем не менее, парные ассоциации по смежности во времени могут играть свою важную роль для предвидения повторяющихся в определенной фиксированной последовательности событий или ситуаций.

В гипотетической полной нейронной модели среды должны быть уровни иерархии. Предполагается, что на каждом уровне иерархии должны строиться укрупнения и обобщения объектов (ситуаций), представленных в виде нейронных ансамблей более низкого (более конкретного) уровня. На каждом уровне укрупнения нейронной модели помимо объектов, и ситуаций должны представляться в виде нейронных ансамблей также и действия, определяющие переходы между ситуациями. Действия, так же как и объекты, должны иметь иерархически организованные степени обобщения и укрупнения.

Предполагается, что возбуждения, приходящие на нейроны каждого уровня иерархии порождаются другими нейронами этого уровня либо объектами другого уровня, либо информацией, поступающей от сенсорного входа. Принципиальных отличий между нейронами и структурой сети разных уровней иерархии быть не должно, хотя конечно могут быть и , скорее всего, должны быть различия в количественных характеристиках предполагаемых свойств и функциональных зависимостей.

Следует отметить, что входной слой любого уровня иерархии не отождествляется с рецепторным, например с сетчаткой глаза, хотя в частном случае такое отождествление возможно. В общем случае, точки на входном слое какого то уровня иерархии возбуждаются при активации информационных элементов другого уровня иерархии. Например, в такой модели кроме уровня букв будет и уровень слогов, уровень слов, а возможно и/или другие уровни иерархии, отражающие семантику проблемной среды.

Экстраполяция регулярных временных последовательностей в иерархической модели среды, содержащей уровни различной степени укрупнения, должна быть значительно лучше, чем в отдельном конкретном уровне, поскольку поднимаясь в модели по иерархии часть-целое чаще можно приходиться к эффективным парным ассоциациям. Так, если временные ассоциации между почти любыми парами входящих в слова букв приблизительно равновероятны, то временные ассоциации между парами разных слов, тем более с учетом синтаксиса, а может быть и семантики более высокого уровня, чаще всего сильно отличаются.

Однако, в каких-то случаях парные ассоциации по смежности во времени даже в многоуровневой модели среды могут быть недостаточны.

Все формальные модели нейронных сетей – модель Мак-Каллока и Питса, перцептрон Розенблатта, признаковые распознающие нейронные сети и рассмотренные модели активной нейронной сети обладают общим свойством – задержки при передаче возбуждающих и тормозящих сигналов по связям отсутствуют. Точнее, задержки по всем связям одинаковы и равны одному такту. Такое свойство нейронных моделей плохо согласуется с данными нейрофизиологии.

В реальных нейронных сетях время передачи сигналов по связям варьируется в широких пределах. Скорость передачи сигналов по различным нервным волокнам может составлять от 1 до 100 метров в секунду. Точно также широко варьируются и длины связей. Естественно предположить, что время учитывается в логике работы

нейронной сети. Более того, можно предположить, что на временном аспекте строится логика работы некоторых нейронных механизмов.

Схема распознавания и воспроизведения временных последовательностей, обладающая существенно большими возможностями, чем связь по смежности во времени появления между парами событий, описана в следующей лекции. В основе схемы лежит включение в логику работы сети временных задержек при передаче сигнала. Задержки используются как при обучении, так и при формировании опережающего возбуждения т.е. предвидения в зависимости от временной предыстории внешних воздействий.

Рассмотренные в лекции “активные” нейронные сети имеют, в основном, иллюстративный смысл. Иллюстрируют они возможность построения активных синергичных ‘живых’, моделей нейронных механизмов мозга. Эта возможность возникает, во-первых, в результате введения в модели свойства накопления неустойчивости, выражающегося в ухудшении функционального состояния и изменении порога возбудимости у несрабатывающего нейрона, а также в ухудшении функционального состояния и противоположном изменении порога возбудимости у срабатывающего нейрона. Во-вторых, свойства нейронов и связей между ними выбираются так, чтобы обеспечивать оптимизацию функциональных состояний отдельных нейронов во взаимодействии и как следствие оптимизацию функционального состояния сети в целом.

Все остальные постулаты о структуре сети, конкретных свойствах нейронов и связей между нейронами не имеют обязательного характера. Варьируя эти постулаты и зависимости можно получить другие и, может быть, гораздо более интересные экспериментальные результаты.

В сетях с изменяющимся функциональным состоянием нейронов были экспериментально получены и некоторые не рассмотренные нами иллюстративные режимы. Это, например, модель режима образования и разрушения “сонных веретен” электроэнцефалограммы млекопитающих, и модель режима управляемой залповой авторитмической активности.

### **Лекция 13. Автомат для запоминания и воспроизведения временных последовательностей**

В этой лекции мы рассмотрим результаты исследования одной из возможных формальных логических схем распознавания и воспроизведения временных последовательностей. Иногда, применительно к мозгу или каким-то моделям, говорят об ассоциациях по смежности во времени. Смежность во времени относится к двум событиям и предполагает появление этих событий последовательно и непосредственно одно за другим. В частности, ассоциации по смежности во времени, так называемые “фазовые последовательности”, рассматриваются в модели Хебба [21]. Такого же рода ассоциации происходят в описанной выше А-сети. Об определенной логической ограниченности и, в то же время, безусловной полезности парных ассоциаций по времени уже говорилось.

В дальнейшем рассматривается другая схема, когда возникает несколько событий и каждое из них в качестве одной из характеристик имеет время своего появления. Задачей является распознавание и “экстраполяция” во времени совокупности таких событий с учетом информации о времени их появления. Ниже приводится упрощенный вариант, соответствующий часто встречающемуся случаю, при котором события появляются одно за другим равномерно во времени. Этот вариант соответствует, например, случаю, когда на входе системы появляются и должны распознаваться состоящие из произносимых фонем слова.

Рассматривается также возможность реализации этой схемы в двухслойной нейронной А-сети. В дальнейшем говорится об одноуровневых временных последовательностях потому, что описывается процесс, основанный на взаимодействии между элементами только одного любого уровня иерархической модели, например, уровня букв (фонем), уровня слогов или уровня слов. Такое упрощение позволяет сделать некоторые количественные оценки, которые могут оказаться полезными при переходе к более сложной многоуровневой схеме. Кроме того, возможно, что некоторые результаты могут иметь самостоятельное значение. Приводимые ниже результаты относятся к некоторому формальному автомату, логическая связь которого с описанными процессами в моделируемых нейронных А-сетях будет рассмотрена ниже.

#### **Задача**

Пусть имеется алфавит  $\{a\}$  с конечным числом символов  $a_1, a_2, \dots, a_n$ . В экспериментах использовались буквы русского алфавита и цифры. Однако, это не принципиально – алфавит может включать любые символы.

При обучении на вход автомата последовательно подаются символы алфавита из произвольных фиксированных выборок. Это могут быть как отрывки текстов, так и случайные последовательности символов. После обучения на вход автомата последовательно подаются эти же или другие выборки символов. Автомат в каждый момент должен предсказывать, какой символ появится на его входе, т.е. оценивать вероятности появления на входе различных символов.



Вероятность “предсказания” автоматом появления на его входе символа  $a_i$  должна находиться в прямой зависимости от того, насколько близко совпадает текущая предыстория работы автомата и предыстория его работы, предшествовавшая появлению символа  $a_i$  в обучающей последовательности. Например, если в обучающей последовательности выделить отрезок в несколько символов, после которого следует обычно символ  $a_i$ , и подать этот отрезок на вход обученного автомата, то вероятность “вспомнить” символ  $a_i$  в автомате должна быть больше, чем вероятность вспомнить любой другой символ.

Ниже описывается логическая схема такого автомата, даются оценки соответствующих вероятностей и приводятся результаты некоторых экспериментов.

Конечно, поставленную задачу можно более или менее просто решать на вычислительной машине “в лоб” - запомнить все обучающие последовательности и сравнивать путем перебора текущий отрезок входной последовательности со всеми аналогичными отрезками обучающих выборок. Например, если речь идет о словах, можно, используя словарь, предсказывать варианты окончания слова по его началу. Такое предсказание может оказаться полезным в распознающем читающем автомате. Однако подобная схема нас не устраивает по двум причинам. Во – первых, не хочется формулировать какие-либо обязательные требования к точности совпадения предысторий. Во – вторых, во многом определяющей является возможность реализации автомата в нейронной модели.

### Схема автомата.

На вход автомата могут подаваться символы из алфавита  $a_1, a_2 \dots a_n$ . Автомат имеет  $n$  элементов по числу символов алфавита  $b_1, b_2 \dots b_n$ . Все изменения в автомате происходят в дискретные моменты времени - такты. Интервал между тактами равен постоянной величине задержки между подачей на вход автомата смежных символов входного алфавита. Элементы  $b_i$  могут находиться в одном из двух состояний – возбужденном и невозбужденном. В каждом такте в возбужденном состоянии может находиться только один элемент. Между символами  $a_i$  и элементами  $b_i$  имеется однозначное соответствие.

Элемент автомата  $b_i$  возбуждается в двух случаях: если на вход автомата подан соответствующий символ входного алфавита  $a_i$ , либо тогда, когда на входе автомата нет символов и некоторая величина  $\sum q_i$  у элемента  $b_i$  больше чем аналогичные величины у других элементов. Если несколько элементов имеют равные максимальные значения величины  $\sum q_i$ , то возбуждается один из них случайным выбором.

От каждого элемента  $b_i$  на каждый  $b_j$ , в том числе и на себя ( $i=j$ ), имеется по  $m$  связей с задержками -  $S_{ij1}, S_{ij2}, \dots, S_{ijm}$ . (Рис.1.) Величина времени задержки связи в тактах  $t_k$  равна номеру связи. Каждая связь имеет свой вес –  $q$ .

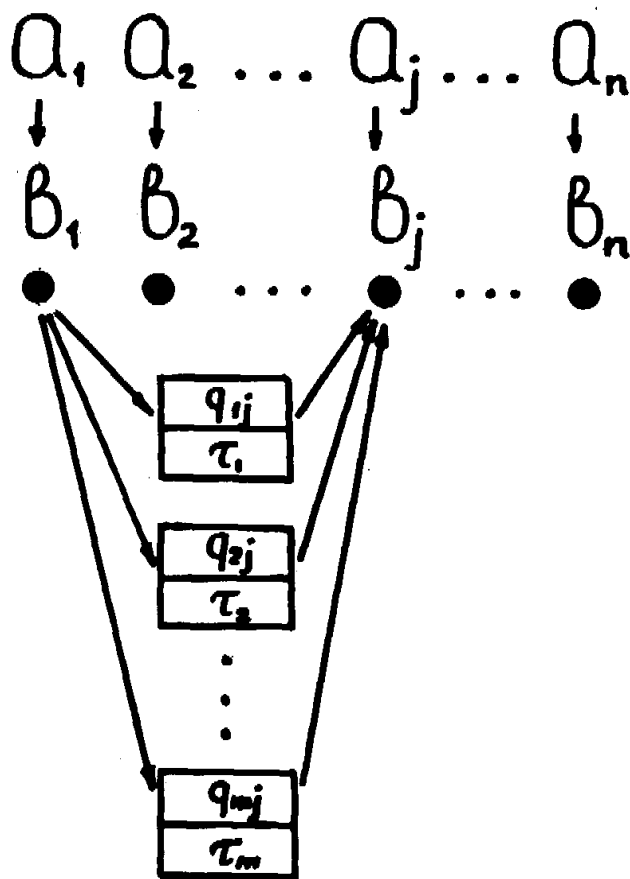


Рис .1.

Возможны два варианта задания весов связей. В основном рассматриваемом ниже варианте вес каждой связи может принимать только два значения – 0 и 1. Во втором варианте веса связей могут изменяться от 0 до  $Q_{\max}$ . Первый вариант будем в дальнейшем называть детерминированным, второй – вероятностным.

### Обучение и воспроизведение опыта

Рассмотрим обучение автомата при детерминированном способе задания весов связей. При обучении на вход автомата подается обучающая последовательность символов. При появлении на входе автомата символа  $a_1$  в момент  $t_1$  возбуждается элемент  $b_i$ . Связи, идущие от элемента  $b_i$ , возбуждаются в моменты  $t_1 + t_k$ ,  $k=1,2,\dots,m$ . Поскольку за  $m$  тактов на входе автомата появляется  $m$  символов и возбуждается  $m$  элементов, от каждого из которых отходит  $n$  связей, в каждый момент  $t_1 + t_k$  в автомате возбуждается  $nm$  связей. До обучения веса всех связей равны нулю. При обучении вес связи  $q_{kj}$  становится равным единице, если в момент ее возбуждения (с учетом задержки) на входе автомата появился символ  $a_j$  и возбудился элемент  $b_j$ , на который идет связь. Таким образом, опыт автомата отражается в распределении весов связей. Поскольку величина максимальной задержки связи равняется  $m$  тактов, то этой величиной ограничивается длина отражающейся в опыте автомата предыстории его возбуждений.

Воспроизведение запомненного при обучении происходит следующим образом. На вход автомата подается часть обучающей последовательности, под действием которой возбуждаются элементы, соответствующие символам на входе и связи, идущие от этих элементов. В каждом такте для каждого элемента производится суммирование весов приходящих на него и возбужденных в этом такте связей. В такте, следующем за

предъявлением последнего входного символа, т.е. тогда, когда на входе автомата символов нет, возбуждается тот элемент  $b_i$ , для которого сумма весов приходящих на него возбужденных в этом такте связей  $\sum q_i$  максимальна.

Начиная с этого момента дальнейшее возбуждение элементов автомата происходит при отсутствии входных символов. Теперь при правильном воспроизведении последовательность возбуждений элементов автомата соответствует части обучающей последовательности входных символов. Обучающая последовательность может быть воспроизведена с любого места. Этот процесс может быть условно назван процессом ассоциаций по смежности во времени. Первая ассоциация является внешней и зависит от запускающей последовательности символов на входе автомата. Дальнейшие ассоциации внутренние, зависящие от воспроизведения, т.е. от хода процесса в самом автомате.

Очевидно, что максимальная длина предыстории, влияющая на возбуждение элементов в ассоциативном режиме равна  $m$ , т.е. времени максимальной задержки по связи и, в конечном счете, определяется числом связей между элементами и величинами временных задержек.

Заметим, что человек более или менее успешно запоминает последовательности из 6 – 7 символов. При попытках запоминания и воспроизведения более длительных последовательностей чаще всего возникают проблемы.

Описанный процесс воспроизведения опыта может быть назван ассоциациями по времени появления входных элементов (символов) на фиксированном временном отрезке. Возможен случай, когда поданная на вход автомата последовательность входных элементов не совпадает в точности ни с одним отрезком обучающей последовательности. В этом случае максимальное значение величины  $\sum q_i$  будет у того выходного элемента, который обучался наиболее похожей последовательности. Процесс, происходящий в этом случае, может быть назван ассоциацией по сходству временных последовательностей.

### **Вариант с опорными кодами**

На вход автомата подается непрерывная обучающая последовательность. Входные символы из этой последовательности запоминаются не по следованию друг за другом, а по следованию за некоторым “опорным кодом”. Опорный код представляет собой алгоритмически получаемый набор последовательностей символов входного алфавита длины  $m$ , равной длине учитываемой предыстории. Каждый входной символ запоминается независимо от других по следованию за своей опорной последовательностью.

При обучении изменяются веса связей, идущих от элементов, соответствующих символам опорного кода, к элементам, соответствующим поступающим на вход автомата символам из обучающей выборки. При помощи опорных кодов реализуется ассоциативно – адресный способ запоминания, при котором адресом является опорный код, а связь запоминаемой информации с адресом является ассоциативной.

Организация процессов обучения и воспроизведения в автомате может отличаться от описанной. Некоторые варианты приводятся ниже.

### **Эксперимент**

Возможность ошибки при ассоциативном воспроизведении связана с тем, что по рассмотренной схеме при обучении учитывается взаимодействие во времени каждой пары элементов независимо от возбуждения остальных элементов. Так, если при обучении подавались последовательности (слова)  $abcd$ ,  $akre$ ,  $kbre$ ,  $krce$ , то в режиме воспроизведения при подаче на вход последовательности  $abc$  на элементах  $d$  и  $e$  возникнет одинаковая величина  $\sum q$ , равная 3, и символы  $d$  и  $e$  будут воспроизведены с равной вероятностью.

Вероятность однозначного правильного воспроизведения растет при увеличении длины учитываемой предыстории и числа элементов в автомате. И то и другое требует увеличения объема памяти автомата. С другой стороны, вероятность правильного воспроизведения уменьшается при увеличении объема запоминаемого материала. В соответствии с этим, экспериментально оценивалась вероятность правильного воспроизведения в зависимости от соотношения между объемом памяти автомата и объемом запоминаемого материала. Кроме того определялось наилучшее соотношение между числом элементов и числом связей для фиксированного объема памяти автомата.

Проведенные эксперименты отличались режимами обучения и воспроизведения.

Вариант 1. На вход автомата подавались непрерывные обучающие последовательности разной длины. Определялась максимальная длина обучающей последовательности, воспроизводящаяся без ошибок, в зависимости от длины учитываемой предыстории, т.е. от количества связей на элемент. Число элементов фиксировалось равным 33. Входной алфавит – буквы. Пробел между словами рассматривался как входной символ. Длина учитываемой предыстории -  $m$  изменялась в пределах от 4 до 16 тактов. Длины обучающих выборок  $L_{max}$  изменялись от 150 до 580. В экспериментах использовались обучающие последовательности трех типов: произвольный текст, наборы, составленные из простых существительных в именительном падеже, случайные наборы букв. Кривые, соответствующие усредненным результатам, приведены на рис.2.

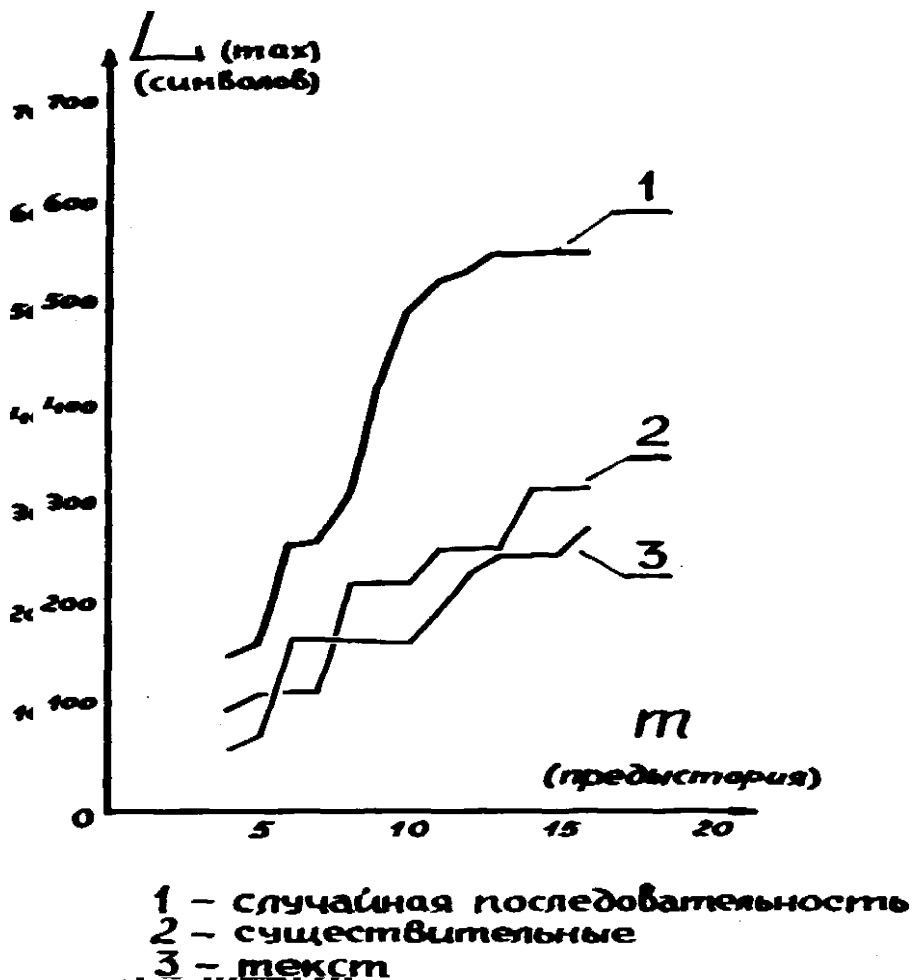


Рис. 2.

Из кривых видно, что при увеличении длины учитываемой предыстории автомат воспроизводит без ошибок более длинные последовательности. Лучшие результаты получены, когда на вход автомата подавались случайные последовательности.

Вариант 2. На вход автомата подавались отдельные слова, запоминавшиеся независимо. Автомат воспроизводил конец слова по заданному началу. Определялось среднее количества начальных букв, необходимое для правильного воспроизведения конца слова, а также количество правильно воспроизведенных концов слов в зависимости от общего числа слов, которым обучался автомат.

На рис 3. приведены кривые, показывающие изменение количества правильно воспроизведенных окончаний слов -  $L_1$  при увеличении общего объема запоминаемой информации  $L$ . Отношение  $L_1/L$  при увеличении  $L$  снижается (рис.7.4.Б). Одновременно снижается среднее число правильно воспроизведенных символов в конце слова (рис.7.4.А))

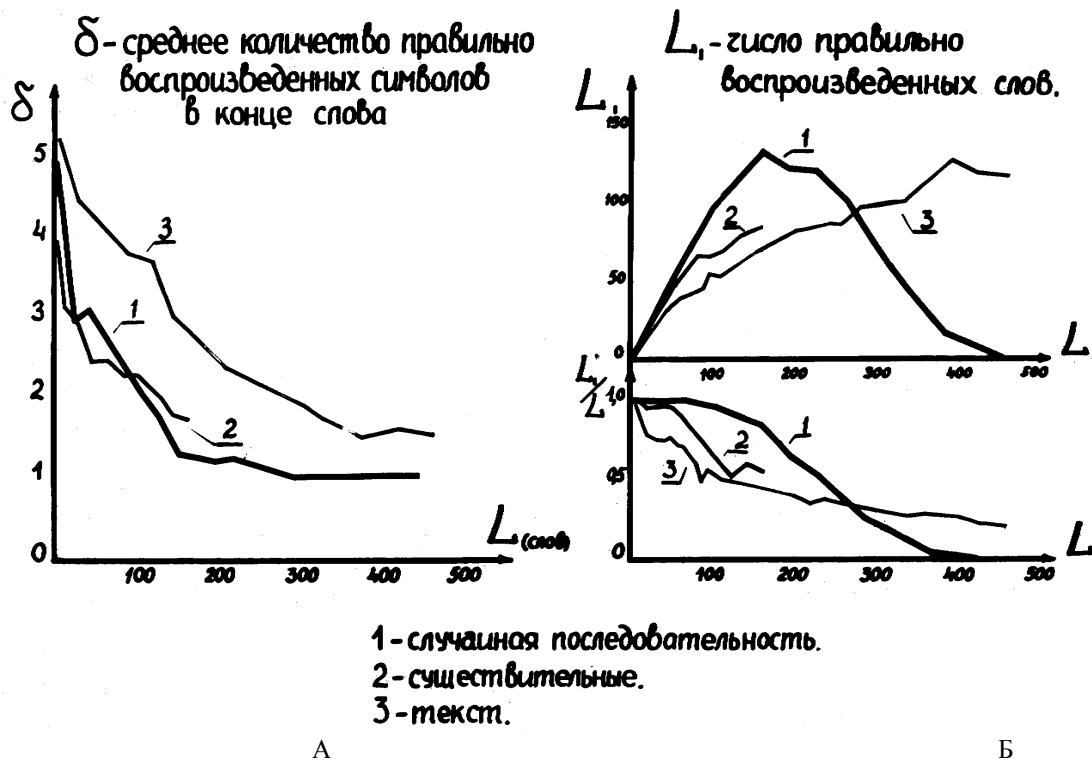


Рис.3.

В этой серии экспериментов случайные слова длиной 7 символов запоминались лучше при  $L$  до 270 слов. При дальнейшем увеличении  $L$  лучше запоминаются слова из текста.

Эксперименты с запоминанием и воспроизведением по отношению к опорному коду показали, что при ассоциативно-адресной схеме запоминания возможно получение относительно высокой вероятности правильного воспроизведения (порядка 0.97).

Эксперименты в целом показали, что в схеме с временными задержками возможно ассоциативное запоминание, определяемое смежностью во времени элементов запоминаемой информации. Вероятность правильного воспроизведения в ассоциативном автомате в значительной степени зависит от отношения объема запоминаемой информации к объему памяти автомата. Вероятность правильного воспроизведения в значительной степени зависит также от характера входной информации. Эта зависимость должна еще сильнее проявиться при вероятностной схеме запоминания.

### Вероятностная схема

Рассмотренная децирированная схема ассоциативного временного взаимодействия не учитывает закономерностей входной информации в том смысле, что часто повторяющиеся входные последовательности запоминаются так же как и однократные. Тем самым в системе не накапливается обобщающий опыт: случайные события фиксируются в памяти наравне с регулярными. В адаптивной, приспособляющейся к среде системе, должны запоминаться вероятностные связи между входными воздействиями.

Описанная детерминированная схема позволяет при воспроизведении предсказывать появление на входе событий (в наших экспериментах символов алфавита) с учетом динамической предыстории. Для того, чтобы та же схема обладала кроме того свойством предсказания с учетом частоты временных парных взаимодействий элементов  $b_i$  автомата необходимо изменить свойства связей и процесс обучения. Эта модификация была названа выше вероятностной схемой. Отличие заключается в том, что в вероятностной схеме веса связей могут принимать уже не только два значения 0 и 1, а изменяться от 0 до  $Q_{\max}$ .

Так же как и в детерминированной схеме вес связи  $q_{jk}$  увеличивается, если в момент ее возбуждения возбужден элемент  $b_j$ , на который идет связь. Если условия увеличения веса связи повторяются многократно, происходит увеличение с насыщением – вес связи стремится к пределу –  $Q_{\max}$ . С течением времени веса не возбуждающихся связей экспоненциально уменьшаются до величины  $Q_{\min}$ , чем обеспечивается “забывание”, необходимое для того, чтобы в памяти отражались не только суммарное количество временных парных взаимодействий элементов автомата, но и ‘свежесть’ опыта.

При компьютерной реализации вероятностного автомата для запоминания весов связей требуется значительно больший объем памяти, чем в случае детерминированной схемы, где вес связи – двоичная величина. Программно реализованный иллюстративный вероятностный автомат имел 33 элемента, т.е. был рассчитан на работу с 33 входными символами. Число связей на один элемент равнялась 6, т.е. длина учитываемой предыстории была 6 тактов (символов). Автомат запоминал и ассоциативно воспроизводил относительно короткие (200 знаков) грамматические тексты. Более подробное исследование свойств одноуровневого вероятностного ассоциативного автомата в работе не проводилось.

Вероятность правильных временных ассоциаций в одноуровневой схеме с временными задержками при передаче возбуждений существенно выше, чем в рассмотренном ранее случае, когда все связи имеют одинаковую задержку, равную одному такту. Однако, как мы видели, ошибки возникают и при схеме с задержками. В случае запоминания регулярных последовательностей вероятность правильного воспроизведения повышается при использовании вероятностной схемы запоминания.

Существенное повышение вероятности правильного воспроизведения временных последовательностей возможно, повидимому, только в многоуровневой иерархической системе памяти. В этой системе должны выработываться обобщения по смежности во времени и при воспроизведении должны учитываться не только взаимодействия между отдельными парами элементов, но и групповые взаимодействия внутри и между регулярными группами элементов.

### **Возможность реализации в активной нейронной модели.**

Нетрудно видеть, что описанная схема одноуровневых временных ассоциативных взаимодействий более или менее просто вписывается в логику работы рассмотренной в предыдущей лекции двухслойной активной нейронной сети.

В предыдущей лекции был описан упрощенный процесс формирования в модели одноуровневой двухслойной нейронной сети парных ассоциаций по смежности во времени поступления входных воздействий. Отличие этого процесса от процессов в описанном ассоциативном автомате состоит в наличии временных задержек при передаче сигналов от одного нейрона к другому.

Задержки и соответствующие задержкам дополнительные связи между элементами или нейронами вторичных нейронных сетей в активной нейронной сети не могут быть введены чисто формально. Образование связей и изменение величин их синаптических проводимостей, в соответствии с нашими общими постулатами, должно происходить в сети автоматически и должно быть направлено на оптимизацию функциональных состояний нейронов.

Изменения величин синаптических проводимостей связей с задержками, необходимые для реализации ассоциаций по времени, при качественном рассмотрении вытекают из принятых нами свойств нейронов. Для автоматического образования связей с задержками необходимо введение дополнительных постулатов. Можно, конечно, исходно закладывать в систему нужное количество “заготовок” - нейтральных связей с задержками, однако, это может потребовать слишком большой памяти.

Физиологичность представлений о наличии в нейронной сети задержек вытекает, во-первых, из того, что передающие сигналы аксоны имеют разную длину, разную толщину, разную скорость и, соответственно, разное время передачи импульсов. Во-вторых, можно предположить, что задержки формируются на самом нейроне. На один нейрон коры головного мозга может приходиться до 1000 связей. Связи оканчиваются синапсами на разных дендритах. Дендриты имеют разную длину, разную конфигурацию и разное расположение по отношению к телу клетки. Все это может определять задержку при передаче сигнала конкретным синапсом.

## **Лекция 14. Основные проблемы, связанные с моделированием нейронных механизмов мозга**

В этой лекции речь пойдет о принципиальных нерешенных проблемах, связанных с моделированием нейронных механизмов мозга. В последнее время в отношении перспектив нейронных моделей в научно-популярной и даже в научной литературе часто встречаются оптимистические прогнозы. Такие прогнозы не представляются обоснованными. Сразу необходимо отметить, что нерешенных и даже еще не решаемых принципиальных проблем много больше чем решенных. Более того, полностью и до конца решенных проблем попросту нет. В частности, в отношении описанных выше А-сетей есть лишь гипотетические представления об общих принципах, а также намечающиеся отдельные пути, причем не столько пути решения, сколько пути исследования.

### **Еще раз об общей оценке существующих нейронных моделей**

Общая оценка логических нейронных сетевых моделей Мак-Каллока, перцептронов и признаковых распознающих формальных нейронных сетей приводилась в предыдущих лекциях. Вывод состоит в том, что эти “пассивные нейронные сети” сколько нибудь существенными моделями механизмов мозга не являются. Еще раз

подчеркнем, что реализуемое в формальных распознающих нейронных сетях признаковое распознавание не обладает свойствами целостности, целенаправленности, синергичности и активности, т.е. не похоже на живое восприятие. Функция признакового восприятия безусловно существует в нервной системе, но занимает там частное вспомогательное место. Причем, вряд ли эта функция реализуется в нервной системе так же как в ПРНС. Идущие от модели Хебба и перцептрона растровые распознаватели несколько более физиологичны..

В 11-й лекции были сформулированы необходимые свойства и функции более или менее полных нейронных моделей механизмов мозга. Для удобства чтения приведем эти свойства и функции еще раз.

1. Для решения задач восприятия и поведения мозг должен при обучении строить модель проблемной среды. Модель проблемной среды субъективна, то есть, она строится с позиций субъекта восприятия и поведения и включает информацию, необходимую для решения задач восприятия и поведения.
2. Модель среды должна быть активна и моделировать изменения в среде, как зависящие, так и не зависящие от субъекта поведения.
3. В модели должна содержаться и использоваться информация, необходимая для реализации поведения, оптимального по критерию максТ и основывающегося на предвидении и эмоциональной оценке вариантов поведения.
4. В модели должны быть целостные структурные обобщенные отображения классов объектов, определяющие, с одной стороны, свойства целого, и с другой стороны, состав и свойства частей целого и отношения между ними. Объект в модели это не обязательно реальный физический объект среды. Объектами модели могут быть также ситуации в среде и действия, изменяющие ситуации.
5. Модель должна быть иерархической, то есть должна быть возможность построения объектов разного уровня укрупнения. Частями (элементами) объектов каждого уровня укрупнения должны являться объекты более низкого уровня. Основная цель построения уровней укрупнения – это сведение многопереборных многоэкстремальных задач поведения к малопереборным одноэкстремальным за счет одношаговых переходов между ситуациями на уровнях укрупнения
6. Должно осуществляться моделирование переходов как по вертикали – между уровнями укрупнения, так и по горизонтали – между ситуациями одного уровня.
7. В модели должны реализовываться ассоциации по сходству и по смежности во времени и на этой основе предвидение и преактивация.
8. Среда не является статической, поэтому модель среды также не должна быть статической. Модель должна быть активной и “жить” по законам среды.
9. В нейронной модели среды должно строиться отображение себя и эмоционально окрашенное отображение взаимодействия себя со средой.
10. В нейронной модели в пределе должно реализовываться абстрактное мышление.
11. Для моделирования мышления человека необходимо использование вербального уровня, то есть необходимо построение словестно-логической модели.

Как уже отмечалось, приведенными свойствами логические нейронные сети Мак-Каллока и Питтса, перцептроны и современные признаковые распознающие нейронные сети не обладают.

Глядя на этот список необходимых функций и свойств, можно придти к выводу, что не являются сколько-нибудь полными рабочими моделями механизмов мозга и рассмотренные нами активные нейронные сети (А-сети). Эти модели, повидимому, более физиологичны, чем модели, названные выше пассивными. Отличаются они, в первую очередь тем, что в них возможны ассоциации по сходству, возможно запоминание и экстраполяция временных последовательностей и на этой основе предвидение и преактивация нейронных ансамблей.

Но самое главное принципиальное отличие А-сети это синергичность и активность - неоднократно упоминавшиеся выше необходимые свойства моделей нейронных механизмов мозга. Свойства, которые могут сделать естественными и неформальными процессы обучения и работы сети. Свойства, на основе которых можно строить предположения и гипотезы о том, что делает нейронную сеть “живой”, определяя возникновение потребностей в получении информации и отображении среды в модели. На основе этих свойств можно строить и предположения о нейронных механизмах мышления, в том числе и творческого. Эти же свойства, повидимому, необходимы для моделирования сознания.

И тем не менее, на вопрос достаточно ли экспериментально получаемых или гипотетически предполагаемых свойств А-сетей для объяснения основных рабочих функций мозга можно еще раз ответить, что, безусловно, напрямую нет. Программные и умозрительные эксперименты с А-сетями имеют, в основном, иллюстративный характер. Остается еще очень много нерешенных проблем. Это, относится и к преимущественно умозрительным экспериментам Л.Б.Емельянова-Ярославского [9], вопреки его утверждениям о построении полной логики модели механизмов мозга. В связи с этим в дополнение к общим проблемам представляется полезным выделить и определить нерешенные конкретные проблемы.

Рассмотрим вопросы, связанные с получением в любой нейронной модели, в том числе и в А-сети, основных важных свойств, необходимых в первую очередь для восприятия и поведения, более сложного, чем рефлексы и инстинкты. Все эксперименты с активными сетями проводились на “плоской” модели, относящейся к одному уровню гипотетической полной иерархической модели. О том, как должна работать многоуровневая активная нейронная модель в очень общем плане говорилось в предыдущей лекции и будет еще говориться в дальнейшем. Однако, о том, как может или должна работать такая модель, можно говорить только предположительно.

Что заставляет работать мозг? Предполагается, что всякая работа “активного” мозга направлена на оптимизацию функциональных состояний отдельных неустойчивых нейронов и сети в целом. На это же должна быть направлена и работа А-сети. Поэтому, приводимые ниже основные проблемы, которые с неизбежностью возникнут при попытках более или менее полного моделирования нейронных механизмов мозга, нужно стараться решать не формально, а, по видимому, на основе свойств неустойчивости нейронов, синергичности и активности нейронной сети в целом.

## Сознание

Может быть не первая, но одна из самых принципиальных и наиболее важных проблем это моделирование сознания, включающая проблему концентрации и переключения внимания. Психология определяет сознание как субъективный образ части объективного мира, как непрерывно меняющуюся совокупность чувственных и умственных образов, непосредственно предстающих перед субъектом в его внутреннем опыте, включающем представление о своем “Я” (словарь практического психолога, 1998 г.).

Выше уже говорилось о том, что в нейронной модели среды должно строиться эмоционально окрашенное отображение фрагментов среды и взаимодействий себя со средой. Такое отображение наряду со свойством активности, лежащем в основе выбора актуального фрагмента, переключения и концентрации внимания, и должно формировать модельное сознание.

Применительно к гипотетической А-сети сначала определим сознание как эмоционально окрашенное реальное или умозрительное восприятие, выражающееся в направленном избирательном возбуждении нейронных ансамблей, соответствующих смысловым элементам какой-то части сформированной в мозге модели проблемной среды.

В А-сети любая, отображаемая в модели проблемной среды информационная единица, отождествляется с нейронным ансамблем. При работе А-сети должно происходить последовательное возбуждение нейронных ансамблей. При работе мозга в сознании последовательно возникают ситуации, слова, объекты. Если поглядеть в окно, то можно, например, увидеть улицу, дома, людей, машины. При этом в сознании возникает одна ситуация, которая может быть названа – вид на улицу из моего окна. Отдельный входящий в ситуацию объект находится на полу-подсознательном уровне и возникает в сознании только при концентрации на нем внимания, оставляя при этом на полу-подсознательном уровне все остальные, входящие в ситуацию объекты и вытесняя на полу-подсознательный уровень ситуацию в целом.

Очевидно, что при этом действует какой-то механизм управления сознанием, обеспечивая порядок в работе мозга. Например, в сознании не могут появиться два, а тем более три, слова, произнесенные одновременно разными людьми. Если одновременно говорят два человека, то в любой момент вы слушаете и слышите только одного из них. Нельзя одновременно делать и два дела, например, писать письмо и говорить по телефону. Способности типа приписываемых Цезарю если и могут существовать, то только как исключение.

Возможно, что сознание и, в частности, концентрация внимания связаны с доминантностью мышления в соответствии с принципом доминанты А.А. Ухтомского и с формированием конуса или “дерева” ассоциативно\_смысловой преактивации в нейронной модели среды. Такого же рода работа должна происходить в многоуровневой А-сети. Последовательное возбуждение нейронных ансамблей А-сети в целом обеспечивается свойствами нейронов, но в эксперименте чаще всего наблюдается частично одновременная работа последовательно возбуждавшихся нейронных ансамблей. Проблема концентрации внимания, имела в виду при экспериментах с одноуровневыми А-сетями, но детально не прорабатывалась.

Эту проблему Н.М.Амосов с сотрудниками решали при разработке своего сетевого автомата, введя внешнюю по отношению к сети формальную систему усиления-торможения (СУТ). СУТ подает дополнительное возбуждение на самый активный в данный момент информации элемент сети и торможение на все остальные. Нейронный уровень в ‘этих работах не рассматривался.

Таким образом, одна из важных проблем, решение которых необходимо при построении нейронных моделей механизмов мозга и, в частности, А-сети, это проблема моделирования составной части сознания, а именно направленной концентрации внимания. Использование идеи СУТ заманчиво, однако нужно, чтобы аналогичная функция выполнялась не внешним алгоритмом, а нейронным механизмом. Интересно было бы также получить в модели и промежуточные уровни активации нейронных ансамблей, с которыми может связываться определенная логика работы, интерпретируемая, например, в терминах стабильные очаги возбуждения, подсознание, полу-подсознание и тому подобное. При этом нейронный механизм управления сознанием, как и все остальные, должен формироваться автоматически на основе свойств нейронов и это формирование должно подчиняться единому общему критерию оптимизации функциональных состояний нейронов и активной сети в целом.



Как уже отмечалось концентрацией и переключением внимания не исчерпывается функция сознания. Не менее важная, а может быть и основная функция сознания это осознание себя во взаимодействии с проблемной средой. Причем не только осознание себя как физического, наделенного восприятием и возможностями действий объекта среды, но и как абстрактного умозрительного оператора в том числе и в абстрактной среде формальных, например математических задач. Из этого следует, что в нейронной модели проблемной среды на всех ее уровнях должен быть представлен и особый объект – “Я”.

Формирование модели “Я” во взаимодействии с другими объектами на всех уровнях укрупнения и обобщения модели проблемной среды происходит всю жизнь обладающего мозгом живого организма. В психологии “Я” (эго) определяется как “результат выделения человеком себя самого из среды, позволяющий ему ощутить себя субъектом своих физических и психических состояний, действий и процессов, переживать свою целостность и тождественность с самисм собой в отношении прошлого, настоящего и будущего.”

Таким образом, сознание можно определить как эмоционально окрашенную целенаправленную активацию фрагментов модели проблемной среды, включающих модель “Я” во взаимодействии с реальными или абстрактными объектами среды. При этом важно подчеркнуть, что в сознании возникает фрагмент модели среды, выходящий во времени (как в будущее, так и в прошлое) и в пространстве за пределы непосредственного восприятия текущей ситуации органами чувств.

Однако, и это определение не является полным. В сознании при восприятии среды могут практически без временной задержки появляться объекты, отсутствующие в модели проблемной среды. Кроме того, при умозрительном конструировании в сознании могут появляться и сохраняться в течение какого-то времени любые неизвестные, отсутствующие в реальной среде в том числе и физически невозможные объекты. В этих случаях сводить сознание к активации существующих в сети нейронных ансамблей, соответствующих фрагментам модели среды недостаточно. Нужны еще какие-то гипотезы о механизмах мозга, например гипотезы, включающие не только представления о долговременной памяти в виде нейронных ансамблей, но и представления о нейронной либо какой-то другой реализации моментальной и кратковременной памяти, а также гипотезы о механизмах воображения.

С учетом всего сказанного можно сказать, что полное нейронное моделирование сознания безусловно представляет собой сложную и еще очень далекую от решения проблему. Точно также можно сказать, что далекой от решения проблемой является и моделирование воображения.

Проблема моделирования сознания многими ставится на первое место при оценках возможности моделирования мышления. Однако, работа мозга в процессе мышления не состоит в выполнении отдельных функций. При целостном подходе может оказаться, что сознание это лишь одна из многих неразрывно связанных и одинаково важных компонент.

Часто считается важным соотносить сознание с какой-то конкретной областью головного мозга. Я полагаю, что локализация сознания не имеет принципиального значения. Для понимания проблемы привязка сознания к гипоталамусу, ретикулярной формации или какому либо другому нейронному субстрату полностью равноправны. Гораздо важнее понять его функциональную сторону в целостной картине работы мыслящего мозга и реализацию на уровне нейронных, либо каких-то других механизмов нервной системы.

## **Умозрительное моделирование и свобода воли**

Такой же важной как проблема сознания и трудно отделимой от него является проблема “умозрительного моделирования” и “свободы воли”. Очевидно, что “умозрительное” моделирование в нейронной сети не должно сводиться к простому предвидению, основывающемуся на запоминании при обучении и воспроизведении временных последовательностей событий. Как ранее так и в дальнейшем мы говорим о логической схеме сведения многопереборных многоэкстремальных задач к малопереборным одноэкстремальным, происходящего в мозгу человека или животного при решении задачи управления поведением в многоэкстремальной среде. При этом приблизительно определяется структура используемой многоуровневой модели среды. Приблизительно определяется и общая логика происходящих в этой модели процессов.

Отчасти несколько забегая вперед, кратко можно сказать, что эти процессы выражаются в моделировании, эмоциональной оценке, сравнении и выборе среди разных вариантов переходов между текущей и целевой ситуациями разных уровней укрупнения в многоуровневой иерархической модели среды. Сначала оцениваются наиболее укрупненные и общие переходы, а затем по мере необходимости все более и более конкретизируемые. Кроме того, к умозрительному моделированию относятся и упоминавшиеся выше очень важные для понимания мышления процессы перцептивного и абстрактного конструирования в воображении.

Что заставляет мозг непрерывно в процессе поведения или абстрактного мышления упорядоченно и направленно выполнять всю эту работу и как она может реализоваться реальными и модельными нейронными механизмами? Также как и при управлении сознанием нейронный механизм управления “умозрительным” моделированием и сравнением вариантов должен формироваться автоматически на основе свойств нейронов и это формирование должно подчиняться единому общему критерию оптимизации функциональных состояний нейронов и активной сети в целом. К этому критерию могут априорно (генетически) добавляться только какие-

либо структурные характеристики сети, а также специфические (для конкретного случая) и неспецифические возбуждающие и тормозящие воздействия на сеть.

Это очень общее положение должно работать при построении любых конкретных нейронных механизмов. Для этого нужно придумать и экспериментально проверить необходимые свойства нейронов и сети.

Очень грубо и приближенно можно говорить о том, что при решении любой задачи в нейронной сети, в том числе и задачи абстрактного мышления, должны возникать очаги возбуждения – активированные нейронные ансамбли, соответствующие, начальной и целевой ситуациям. Постоянная принудительная работа фиксированных ансамблей должна нарушать регулирование оптимальных состояний нейронов сети и стимулировать активацию цепочек ансамблей и переходов, связывающих эти ситуации. При моделировании необходимы привязанные к конкретным нейронным механизмам представления о формировании, поддержке и гашении очагов возбуждения.

В самом общем плане можно полагать, что для гашения очагов возбуждения и восстановления общей регуляции в сети должно быть необходимо физическое достижение целевых ситуаций, а при абстрактном мышлении – возбуждение очагов возбуждения одновременно с активацией нейронных ансамблей, соответствующих переходу между текущей и целевой ситуациями. При этом нужно отметить отдельную проблему – реализацию на нейронном уровне эмоциональных оценок переходов между ситуациями и сознательный направленный выбор на основе этих оценок наилучших вариантов переходов.

Осуществляемые в мозге в процессе мышления сознательный перебор, сравнение на основе эмоциональных оценок и фиксация лучшего варианта приводит к необходимости понимания не только на психологическом уровне, но и на уровне нейронных реальных и модельных механизмов еще одного феномена, который может быть назван “свобода воли”. Я думаю, что проблема понимания и моделирования свободы воли не менее существенна и не менее сложна, чем проблема понимания и моделирования сознания.

Выше говорилось о проблемах моделирования воображения и управления целенаправленным сравнением воображаемых вариантов. Воображение может строиться на основе возбуждения существующих в модели проблемной среды нейронных ансамблей. Но этого недостаточно. При творческом мышлении используются перцептивное и когнитивное конструирование, т.е. воображаемое построение в сознании объектов, отсутствующих в модели среды, а может быть, отсутствующих и в самой среде. Для полного моделирования мышления необходимо конструктивное решение и этой проблемы.

### **Возникновение и гашение очагов возбуждения**

Выше описывалось как возникает очаг возбуждения в одноуровневой А-сети при подаче на нее стабильного информационного воздействия. Описывалось как при этом активируется или формируется вторичная нейронная сеть, т.е. ансамбль нейронов, являющийся элементом памяти. При конкретной разработке многоуровневой нейронной модели необходимо будет определить логику и механизмы возникновения и гашения очагов возбуждения на всех уровнях нейронной сети.

Очаг возбуждения это постоянно, либо часто работающий (возбужденный) отдельный нейронный ансамбль или группа нейронных ансамблей. На возникновении и гашении очагов возбуждения в гипотетической полной нейронной А-сети строятся все представления о ее работе при восприятии, поведении и мышлении. Возможно, что подобным же образом работают и реальные нейронные механизмы.

В лекции 12 упоминались эксперименты, когда в А-сети создавался очаг возбуждения путем подачи на часть нейронов внешних возбуждающих воздействий. Очаг возбуждения нарушал работу сети, что приводило к образованию вторичной нейронной сети – элемента памяти. В более или менее общем виде можно представить и промоделировать появление в нейронной сети очагов возбуждения при возникновении первичных генетически определенных физиологических потребностей.

Первичным потребностям соответствуют первичные доминанты и отображающие их нейронные ансамбли. Как уже говорилось, доминанта по А.Ухтомскому это стабильно сохраняющийся очаг возбуждения. При наличии потребностей происходит возбуждение соответствующих нейронных ансамблей с интенсивностями, определяемыми величинами потребностей. В соответствии с принципом доминанты А.А.Ухтомского на основе сравнения потребностей и с учетом текущей внешней ситуации формируется и фиксируется главная поведенческая доминанта и тормозятся остальные. Одновременно передается возбуждение на те нейронные структуры, которые должны участвовать в решении поведенческой задачи и снятии доминанты, т.е. в гашении очага возбуждения.

В соответствии с рассмотренными выше гипотезами доминанта возникает при оптимизационном решении поведенческой задачи по критерию  $\max T$ . От главной доминанты, т.е. очагов возбуждения самого верхнего уровня обобщения и укрупнения модели проблемной среды на более низкие уровни распространяется конус или “дерево” возбуждений. Одновременно передается торможение на нейронные ансамбли, соответствующие альтернативным переходам. На каждом уровне иерархической модели среды выбирается и фиксируется своя временная производная доминанта, соответствующая выбранному на этом уровне действию (переходу).

Доминанта (очаг возбуждения) гасится при реальном выполнении соответствующего перехода. Наибольшую трудность при нейронном моделировании таких процессов будет, повидимому, представлять реализация перебора вариантов поведения, т.е. переходов на всех уровнях нейронной модели проблемной среды, их сравнение по эмоциональной оценке и фиксация в виде длительных и временных очагов возбуждения.

Моделированием процесса возникновения и гашения очагов возбуждения при формировании доминант, соответствующих генетически заданным первичным потребностям проблема далеко не исчерпывается.

В соответствии с рассмотренной логикой очаги возбуждения должны возникать и гаситься в нейронной сети также и при возникновении и удовлетворении вторичных потребностей, при постановке и решении абстрактных в том числе творческих задач, при достройке и коррекции модели проблемной среды.

Четких представлений о том, как все это происходит в реальных или моделируемых нейронных сетях, нет. Нет также конструктивных гипотез, т.е. представлений на уровне нейронных механизмов, о том, как возникают поддерживаются и изменяются постоянные вторичные потребности, т.е. потребности не заданные генетически. То же самое относится и к временным вторичным потребностям, связанным с постановкой и решением абстрактных в том числе и творческих задач мышления.

Любые потребности сводятся в конечном счете к потребности в гашении очагов возбуждения. В наиболее общем плане мышление основывается на формировании и гашении очагов возбуждения. Очаги возбуждения наряду со свойством накопления неустойчивости нейронных элементов являются причиной и источником активности мозга. Поэтому, можно сказать, что мыслящая система должна не только уметь, но и хотеть мыслить.

Можно сказать, что в нейронной сети, состоящей из активных неустойчивых элементов, постоянно возникают внутренние потребности в оптимизации своего функционального состояния. Никаких других потребностей в нейронной сети нет. Возникновение потребностей по оптимизации состояния связаны с формированием в сети стабильных очагов возбуждения, приводящих к неравномерному возбуждению нейронов сети, что ведет к ухудшению функционального состояния отдельных нейронов и сети в целом. Удовлетворение таких потребностей достигается за счет гашения очагов возбуждения.

Эти свойства нейронов и нейронной сети должны быть заданы генетически. Кроме того, генетически заданы внешние по отношению к нейронной сети первичные физиологические потребности и их отображение в нейронной сети в виде стабильно поддерживаемых первичных очагов возбуждения. Гасятся такие очаги возбуждения только при удовлетворении первичных потребностей. Как возникают вторичные очаги возбуждения, связанные с вторичными потребностями остается вопросом, как на физиологическом уровне, так и на уровне моделирования. Точно также вопросом остается и то, как работают реальные и модельные нейронные механизмы при возникновении и гашении вторичных очагов возбуждения в случае построения, коррекции или достройки модели среды.

Иногда говорят о том, что одной из генетически заданных первичных физиологических потребностей является получение информации. Это не верно. В описанном выше эксперименте Ротенберга и Аршавского в мозгу крысы при постоянном наблюдении приоткрытой двери в незнакомое темное помещение строился незаконченный фрагмент модели проблемной среды. и формировался вторичный очаг возбуждения, погасить который можно только исследовав это помещение и достроив фрагмент модели.

Таким образом, получение информации само по себе (любопытство) потребностью не является. Потребность же состоит в необходимости оптимизации функционального состояния нейронной сети при его ухудшении за счет появления очагов возбуждения. Вторичные очаги возбуждения гасятся, в частности, при активном и целенаправленном получении информации, необходимой для достройки целостной модели среды. Целенаправленность состоит в том, что по информации, имеющейся в существующей модели среды, строятся и затем проверяются гипотезы о свойствах исследуемого неизвестного объекта или исследуемой новой ситуации. Причем не о свойствах вообще, а о свойствах, имеющих смысл в отношении поведенческих задач, определяемых как первичными, так и вторичными потребностями.

### **Получение и запоминание признаков, отношений и метрических характеристик**

Эта казалась бы проблема значительно более низкого уровня может быть сопряжена при моделировании с очень серьезными затруднениями. Выше отмечалось, что входное поле любого, в том числе и первого самого низкого уровня А-сети, не отождествляется с рецепторным, например с сетчаткой глаза, хотя в частном случае при растровом распознавании такое отождествление возможно. В общем случае, “вертикальные” входы какого то уровня иерархии возбуждаются при активации информационных элементов другого уровня иерархии, например, на уровень слов должны подаваться входные воздействия с уровня букв или с уровня фонем. Части объединяются в целое не просто по принципу “И” Должны также формироваться “горизонтальные” связи между смысловыми элементами одного уровня. Такими связями, в частности, должны задаваться отношения и метрические характеристики.

Формирование элементарных смысловых единиц типа признак, отношение или метрическая характеристика не только в А-сети, но и ни в какой другой нейронной модели не рассматривается. Так в ПРНС на вход сразу

поступают признаки, полученные где-то и как-то вне нейронной модели. Таким образом, уже на самом нижнем уровне при входе в нейронную модель возникает проблема формирования в нейронной сети признаков. Проблема формирования отношений и метрических характеристик будет возникать на всех уровнях нейронной модели.

В зрительном анализаторе животных имеются рецептивные нейронные поля, выделяющие простые первичные признаки, такие как контуры, линии, направления линий, движения, грубая оценка размера объекта (большой, маленький). Такие простейшие преобразования входной информации в нейронной сети можно моделировать, но их явно недостаточно. В анализаторах входной информации животных и человека, повидимому, происходит последовательная, усложняющаяся, многоуровневая обработка, направленная на формирование все более сложных характеристик воспринимаемых объектов и ситуаций.

Как строить на нейронах функции формирования, в общем случае, сложных произвольных, признаков, отношений и метрических характеристик неизвестно.

Неизвестно так же, как на нейронах или нейронных ансамблях выделять, передавать, запоминать и сравнивать количественные характеристики. В частности это относится к количественной эмоциональной оценке, необходимой для сравнения вариантов поведения. В экспериментах, описанных в лекции 5, собака, выбирала вариант поведения в соответствии с формулой выбора:

$$\max M_j V_j$$

где  $M_j$  – мотивация,  $V_j$  – скорость удовлетворения соответствующей потребности.

Собака, конечно, делала приблизительные оценки, но, тем не менее, в мозгу собаки каким-то образом на нейронном уровне при обучении запоминались ожидаемые величины скоростей удовлетворения потребностей, оценивались величины соответствующих мотиваций и в процессе сравнения вариантов определялись какие-то величины, коррелирующие с произведением  $M_j V_j$ . (Вряд ли собака производила точное умножение).

Таким образом, еще раз подчеркнем, что в мозгу определяются и запоминаются величины (количества), а также производятся операции с этими величинами. Простейшая из необходимых операций с количествами это операция сравнения (больше-меньше). При построении моделей необходимо понять как это делается нейронными механизмами мозга.

### **Управление процессами в модели с понятийного уровня.**

Важной проблемой является построение нейронных механизмов управления выделением первичных и, вообще, любых информационных характеристик среды с понятийного уровня модели. При построении функции любого уровня, в том числе и функций получения первичных признаков и других характеристик нужно помнить о том, что недостаточно строить просто дешифратор вход-выход. Необходимо чтобы в модели реализовывалось управление с верхних понятийных уровней, обеспечивающее целенаправленность и преднастройку.

Рассмотрим следующий пример. Пусть дана фраза: в правом верхнем углу картинка должна быть красная прямая линия, отклоняющаяся на угол 30 градусов от горизонтали. Прочтя эту фразу, человек понимает о каких объектах и характеристиках идет речь, что искать – прямую красную линию, где искать – в правом верхнем углу. Понимает он также и то, что такое правый, что такое верхний, что такое горизонталь, что такое угол с горизонталью и что такое угол картинка. После этого направленно в нужном месте выделяется линия и проверяются ее характеристики – прямолинейность, цвет и направление. Ничего лишнего не делается. В приведенном примере речь идет о простых признаках и отношениях. И признаки и отношения могут быть гораздо более сложными. Информация, управляющая работой нижнего уровня, в общем случае, не поступает извне, а содержится в модели проблемной среды. Именно так, но не на нейронных механизмах, а чисто формально, организован описанный ранее процесс целенаправленного распознавания “с пониманием”.

Никаких конкретных конструктивных предположений о нейронных механизмах реализации подобных процессов ни в нейрофизиологии, ни на уровне моделирования нет.

### **Построение целостных многоуровневых иерархических структурно-метрических описаний**

Требующей решения является проблема образования на нейронном уровне иерархических целостных структурно-метрических описаний, обеспечивающих двунаправленное взаимодействие между частями и целым, а также между частным и общим, т.е. построение обобщений и укрупнений.

В одной из предыдущих лекций было показано как целостному объекту может быть поставлено в соответствие сложное структурно-метрическое описание, включающее структурные части объекта, их метрические характеристики и отношения между частями. Такое описание необходимо для реализации целостного целенаправленного восприятия. Подобные описания должны формироваться в нейронной модели и для решения других задач. Можно ли реализовать какой-то эквивалент таких сложных описаний в нейронной модели объекта?

Проблема выделения, запоминания и передачи количественных характеристик уже отмечалась. Кроме этой проблемы при представлении описания целостного объекта на нейронном уровне возникает еще три. Во-первых, нужно, чтобы в рамках строящегося описания отражались обобщенные метрические характеристики частей

объекта и взаимозависимости между входящими в описание элементами. Во-вторых, в пространстве структурных элементов, отношений и метрических характеристик должна каким-то образом выделяться область существования описываемого объекта или класса объектов. В третьих, должны моделироваться отношения между частями и целым.

Те же самые вопросы возникают при попытке представить себе как на нейронном уровне отражаются не только описания целостных объектов, но также и описания как обобщенных так и конкретных текущих, динамически изменяющихся ситуаций.

Мы уже говорили о том, что иерархическая многоуровневая модель среды должна иметь несколько уровней укрупнения. Упрощенно, взаимодействие между уровнями выглядит следующим образом. Выходы с элементов нижнего или промежуточного уровня укрупнения формируют (активируют) элементы более высокого уровня укрупнения. Выходы с элементов верхнего или промежуточного уровня укрупнения активируют входящие в это укрупнение элементы предыдущего, более низкого уровня. Упрощение состоит в том, что в таком представлении не учитывается главное – целое формируется не просто совокупностью частей, а частями, находящимися в каких-то взаимодействиях между собой. Эти взаимодействия могут задаваться соответствующими отношениями между частями целого.

Часто описания иерархических или как иногда говорят пирамидальных модельных представлений иллюстрируются схемой типа дерева. Например, подобной схемой иногда иллюстрируют иерархию и взаимодействие между уровнем букв и уровнем слов. Недостаточность такой схемы для описания связей между частями и целым очевидна, поскольку в такой схеме объединение элементов предшествующего уровня в элемент более укрупненного уровня происходит только за счет подразумеваемой логической функции И. В общем случае для задания связи между частями и целым могут требоваться другие отношения.

Например, одни и те же буквы в зависимости от их порядка формируют разные слова. То есть для описания целого (слова) необходимы отношения, определяющие порядок букв в слове. Для произносимого слова такими отношениями должны быть задаваемые между фонемами отношения следования во времени.

Реализовать такие отношения в модели нейронной сети можно с использованием схемы, подобной схеме рассмотренного в лекции 14 автомата для воспроизведения временных последовательностей. Напечатанное или написанное слово тоже может быть задано отношениями следования во времени между входящими в слово буквами. Однако, для напечатанного или написанного слова более естественны отношения типа “слева”, “справа” или “следует за”. Как реализовать в нейронном автомате такие отношения без сведения их к отношениям следования во времени неизвестно.

В более сложных случаях описания целого в терминах частей и отношений между ними могут оказаться значительно более сложными. Характеристики некоторых подобных описаний приведены в лекции 7. Реализация таких описаний в нейронном автомате представляет важнейшую нерешенную проблему.

## **Проблема моделирования памяти**

В нейрофизиологии и психологии имеются подкрепляемые многочисленными экспериментами теории о существовании двух видов памяти – кратковременной и долговременной, соответствующих запоминанию информации на короткий срок и надолго. Соответственно предполагается и существование двух разных механизмов запоминания. Точных конструктивных представлений о том, как устроены кратковременная и долговременная память в физиологии нет.

Гипотезы об устройстве долговременной памяти сводятся в основном к предположениям о запоминании информации на уровне нейронных ансамблей и синаптических проводимостей связей между нейронами, либо запоминание связывается с изменением структуры белка рибонуклеиновых кислот. Обоснованных и убедительных гипотез об устройстве кратковременной памяти нет. Чаще всего говорится, например, о каких-то абсолютно непонятно как устроенных “биоэлектрических контурах колебаний в нервной системе”, или о циркуляции информации в замкнутых цепочках возбуждения.

Функциональная роль этих двух видов памяти изучена и показана достаточно подробно. Очевидно, что важно понять и промоделировать механизмы кратковременной и долговременной памяти и режимы передачи информации из кратковременной памяти в долговременную.

Предполагается, что в А-сети долговременная память, т.е. модель проблемной среды, должна строиться на нейронных ансамблях разных уровней обобщения и укрупнения, а также на возбуждающих и тормозящих связях между ними. Обучение в А-сетях это довольно продолжительный процесс. При обучении изменяется топология и проводимости связей. В этих изменениях отражаются и запоминаются только повторяющиеся события – объекты, ситуации, последовательности ситуаций. Единичное и случайное при обучении не учитывается.

Мозг человека отражает не только типичное и повторяющееся. Отражаются также и мимолетные динамически изменяющиеся ситуации.

Обычно “моментальные” ситуации в памяти не хранятся и забываются так же быстро, как и возникают. Однако, некоторые “моментальные” ситуации, действия или слова иногда запоминаются на очень длительное время и всплывают в памяти через много лет. Это особенно характерно для эмоционально значимых ситуаций, хотя бывает, что мгновенно и надолго запоминаются и малозначимые события.

Механизм моментального восприятия и запоминания нового в нейронных моделях отсутствует. Не только в А-сетях, но также в перцептроне и ПРНС для восприятия и запоминания нового требуется продолжительный процесс обучения.

Таким образом, проблема, которую нельзя проигнорировать при разработке моделей механизмов мозга, это построение физиологичной модели памяти, включающей как моментальное запоминание отдельных объектов и событий, так и статистически корректируемое продолжительное запоминание на уровнях обобщения и укрупнения.

При моделировании, повидимому, нужно учитывать предположение о том, что информация передается из кратковременной памяти в долговременную, а также предположение, что эта передача происходит во сне, во время которого осуществляется многократное возбуждение соответствующих нейронных информационных элементов долговременной памяти. Эти предположения опираются на сведения о некоторых амнезиях, при которых забывается только та информация, которая была получена и запомнена после последнего сна.

### **Переход от нейрона к нейронному ансамблю**

Представляется, что отдельный ненадежный нейрон не может быть логической единицей мозга и носителем какой-либо значимой содержательной информации. Как уже отмечалось, из многих физиологических экспериментов следует, что логической единицей мозга является нейронный ансамбль - группа нейронов, объединенных взаимовозбуждающими связями. Нейронный ансамбль является логической единицей и в А-сети, если считать, что реализующие возвратное возбуждение элементы второго слоя это тоже ансамбли нейронов со своими отдельными функциями и связями с нейронами первого слоя и с другими ансамблями второго слоя.

Реальное взаимодействие – передача возбуждений и торможений в нейронной сети происходит по связям между нейронами. При описании логики работы нейронной сети нужно переходить к рассмотрению взаимодействий между ансамблями нейронов. Если описывается работа мозга или строится нейронная модель, то нужно определить что такое входные интегральные возбуждающие и тормозящие воздействия на ансамбль, как передаются возбуждения и торможения между ансамблями, как происходит обучение в ансамблях.

Более ясно характер возникающих в связи с этим проблем виден из следующего примера. В ПРНС разделяющие гиперплоскости строятся на отдельных нейронах – пороговых элементах, считающих взвешенную сумму входов. Предположим, мы хотим в целях повышения надежности заменить такой нейрон ансамблем нейронов. Возникает проблема как сделать ансамбль нейронов единым пороговым элементом и не только заставить считать взвешенную сумму входов, но и обучаться, то есть менять при обучении веса входов. Сходные проблемы возникнут в А-сетях, если при моделировании заменить тормозный элемент вторичной сети ансамблем нейронов.

Конечно, проблему нейрон – ансамбль нейронов, рассматриваемую как проблему надежности, можно решить простым дублированием нейронов и связей. Это обеспечит надежность, но потребует слишком большого числа нейронов и, особенно, связей.

### **Проблема моделирования языка и абстрактного мышления**

Язык и абстрактное мышление это то, что отличает мышление человека от мышления животного. Выше говорилось о двух строящихся в мозгу человека моделях: словестно-логической и образной. Понятно, что моделирование словестно-логической модели невозможно без моделирования языка, что представляет большую и очень сложную проблему. Говорилось также о взаимодействии между словестно-логической и образной моделями в процессе мышления. Как реализуется это взаимодействие, то есть передается информация и осуществляется управление процессом мышления на уровне взаимодействующих нейронных механизмов двух полушарий головного мозга неизвестно.

Далеко не все ясно и при рассмотрении процесса мышления отдельно только в словестно-логической модели или отдельно только в образной модели. Можно лишь, как уже говорилось, предположить, что любое мышление сходно с процессом поиска лучшего в каком-то смысле пути между двумя ситуациями. Путь отображается в мозге стабильными очагами возбуждения и последовательно активируемыми нейронными ансамблями.

Приблизительно можно сказать, что в словестно-логической модели это должен быть каким-то образом реализованный поиск непротиворечивого пути (последовательности возбуждений ансамблей нейронов) между двумя возбужденными ансамблями нейронов, определяющими задачу. Этот поиск должен основываться на построении одношаговых переходов в виде предположений или гипотез на уровнях обобщения и укрупнения, а также в проверке реализуемости этих переходов на более конкретных уровнях. Выбор между несколькими непротиворечивыми вариантами переходов должен, повидимому, управляться эмоциональным центром. Реализация этих процессов в модели является одной из наиболее трудных проблем.

### **Проблемы второго плана.**

Кроме выделенных основных проблем существуют и проблемы, которые условно можно отнести к проблемам второго плана. Относительно некоторых из них возможны приблизительные интерпретации, но безусловно, необходимы и тщательные, в том числе и экспериментальные проработки. Перечисляемые в этом

разделе дополнительные проблемы при их подробной проработке могут оказаться не менее важными и не менее сложными, чем проблемы, условно отнесенные выше к группе основных.

К проблемам второго плана можно отнести:

- отображение ситуаций, процессов и действий;
- отображение себя в среде;
- реализация переходов между уровнями модели снизу вверх и проекции сверху вниз;
- переходы по горизонтали – связь между причиной и следствием;
- экстраполяция временных последовательностей;
- предвидение результатов действий и процессов;
- связь между потребностью в оптимизации функционального состояния нейронной сети с необходимостью получения информации о среде и построения модели проблемной среды.

Кроме того, общие свойства активных нейронов и А-сетей позволяют строить гипотетические интерпретации таких особенностей работы мозга как сон и неспецифические эмоциональные реакции. Важно уточнить и промоделировать эти интерпретации.

Определенные качественные интерпретации неспецифических эмоциональных реакций уже приводились. На достаточно общем уровне возможны интерпретация и моделирование сна. При работе А-сети последовательно возбуждаются вторичные нейронные сети (нейронные ансамбли), соответствующие каким-то информационным элементам. Какие вторичные сети возбуждаются и в какой последовательности определяется решаемой А-сетью задачей. В любом случае при переработке какой-то информации возбуждается только часть вторичных сетей.

Параметр Q нейрона при каждом срабатывании уменьшается на дискретную величину. У несрабатывающих нейронов параметр Q увеличивается. Если какая-то часть вторичных сетей регулярно возбуждается, а другая часть не возбуждается, то происходит увеличение средних отклонений значений параметров Q от оптимального значения. У нейронов неработающих вторичных сетей параметры Q становятся больше значения  $Q_{opt}$ , у нейронов работающих вторичных сетей эти параметры могут стать меньше чем  $Q_{opt}$ . Это приводит к ухудшению общего функционального состояния сети. Нормализация функционального состояния сети возможна только при “отдыхе”.

Как уже говорилось в предыдущей лекции отдых сети при бодрствовании и отсутствии актуальных внутренних задач основан на равномерном возбуждении нейронов за счет внутреннего случайного возбуждения (перемешивания). Отдыху при бодрствовании способствует также подача не нейронную сеть внешних равномерных случайных возбуждающих воздействий.

Если реальный нейрон живого мозга также как и нейрон А-сети, имеет параметр, аналогичный параметру Q, и среднее отклонение этих параметров нейронов мозга от оптимального значения является эмоциональной переменной, то оптимизация при бодрствовании этой переменной будет восприниматься как приятное ощущение.

С этим, повидимому, связаны неспецифические эмоциональные реакции, например, такие, как реакции на спокойную музыку, свет костра или морской прибой. Эти положительные реакции могут возникать, когда случайный или ритмический характер внешних воздействий способствует равномерной работе нейронов и оптимизации функционального состояния нейронной сети. Известно, что для быстрого повышения работоспособности интенсивно работающих диспетчеров используется спокойная тихая музыка или специально подобранные случайные зрительные или, чаще, звуковые воздействия.

Для полного отдыха нейронной сети необходим специальный тип работы, когда прекращается обработка информации, блокируются внешние воздействия и включается описанный выше экспериментально полученный в А-сети режим саморегуляции. Этот режим полного отдыха интерпретируется как сон. Среднее отклонение параметров Q нейронов интерпретируется как “эмоциональная переменная”. Увеличение этой эмоциональной переменной требует смены режима работы сети от кратковременного отдыха до длительного, то есть сна. Приблизительно в том же аспекте могут интерпретироваться разные, зависящие от режимов работы мозга ритмы электроэнцефалограммы.

Вполне возможно, что со сном связаны не только нормализация функционального состояния нейронной сети, но и некоторые функции, направленные на “приведение в порядок” запомненной информации. Например, существуют гипотезы о том, что во сне может осуществляться коррекция горизонтальных и вертикальных связей в модели проблемной среды, или уже упоминавшаяся функция переноса информации из кратковременной памяти в долговременную.

На основе представлений о неустойчивости нейронов, синергичности и активности нейронной сети можно с разной степенью приближения интерпретировать и такие свойства мозга, как доминанта, навязчивый мотив, интуиция, инсайт (озарение) и творчество. О части таких интерпретаций мы поговорим в следующей лекции.

## **Выводы**

Рассмотрение проблем, связанных с моделированием нейронных механизмов мозга, приводит к выводу о том, что нерешенных проблем много и они сложны. В отношении многих проблем нет не только никаких конструктивных гипотез решения на уровне нейронного моделирования, но отсутствует и конструктивная

нейрофизиологическая информация. Эти проблемы будут представлять, повидимому, наибольшие и очень серьезные затруднения при моделировании более или менее полных нейронных механизмов мышления.

Появляющиеся иногда оптимистические оценки ситуации в науке и прогнозы о перспективах моделирования нейронных механизмов мозга малообоснованы. Внимательное рассмотрение перечисленных проблем приводит к значительному снижению уровня оптимизма в отношении существующих и предполагаемых ближайших результатов моделирования мышления. И, тем не менее, если рассмотренные функции мозга и реализующие их нейронные механизмы имеют реальную материалистическую природу, то они в принципе решаемы, а следовательно решать их можно и нужно.

Представляется, что для решения отмеченных проблем чрезвычайно полезным может быть продолжение и развитие компьютерного моделирования активных нейронных сетей, т.е., в конечном счете, изобретение и экспериментальная проверка вариантов мышления. Безусловно в плане понимания и конкретизации перечисленных вопросов значительную пользу могут и должны принести результаты целенаправленных нейрофизиологических исследований, выполняемых по возможности с позиций целостных представлений или гипотез о работе мозга.

## **Лекция 15. Виды мышления и гипотезы о некоторых механизмах мышления и творчества**

### **Общие замечания**

Сначала немного о терминологии. Часто используемое выше слово модель перегружено. Во-первых, говорится о строящейся в мозге модели проблемной среды и, во-вторых, о компьютерной модели нейронных механизмов мозга. Фактически во втором случае часто говорится о модели модели. Термин модель проблемной среды возник не случайно. Он выражает суть того, к чему он применяется. Информационное отображение среды в мозге не должно быть просто пассивным описанием. Тем не менее, для упрощения создавшейся ситуации в дальнейшем там, где это не слишком нарушает смысл изложения, вместо слов модель проблемной среды будет говориться – отображение проблемной среды. При этом нужно не забывать, что строящееся в мозге отображение среды должно быть активно и обладать свойствами действующей модели.

Мы исследуем возможность моделирования мышления. Поэтому представления о работе гипотетической компьютерной нейронной модели должны совпадать с представлениями о работе мозга. В начале книги говорилось о том, что к пониманию мышления полезно, а может быть и необходимо, идти от анализа задач восприятия и поведения, также как и от понимания работы активных нейронных механизмов мозга. Излагаемое в книге основано на этом, хотя и носит достаточно общий в основном гипотетический характер. Конкретная реализация полной многоуровневой нейронной модели относится пока еще к очень отдаленной перспективе и может быть связана с непростыми проблемами. Возможно, что при построении каких-то вариантов компьютерной модели полезны будут максимально допустимые упрощения.

В мозге строится иерархическое многоуровневое отображение проблемной среды. Это отображение используется для управления поведением, для восприятия среды и для решения абстрактных задач мышления. Описание этого отображения, отвечающее на вопросы о том, из чего оно состоит, как структурно устроено, как формируется и как работает, должно фактически определять функциональную сторону полной компьютерной нейронной модели механизмов мозга. Важные аспекты такого описания уже приводились в предыдущих лекциях. Этим же вопросам будет уделено основное внимание в последующем изложении.

Мышлением можно называть все происходящие в мозге процессы, связанные с осознанной переработкой информации. Это могут быть разные процессы, которые можно по-разному классифицировать. Естественно различать процессы, связанные с восприятием и познанием среды, управлением поведением, решением формальных задач и творчеством. Общим для любых процессов мышления является, во-первых, отражение их результатов в сознании и, во-вторых, сознательное целенаправленное управление процессом на основе использования отображения проблемной среды и умозрительного моделирования.

Далее мы рассмотрим процессы перцептивного, когнитивного (познавательного) и практического (поведенческого) мышления, а также гипотетические представления о механизмах мышления и творчества человека. Мы обсудим достаточно общие предположения о том, как может интерпретироваться мышление вообще и в частности на уровне модели нейронных механизмов мозга. Естественно, эти интерпретации, ориентированные на главное, упрощены и не затрагивают многих вариантов, особенностей и тонкостей реальных процессов.



## Необходимые свойства формируемой в мозге модели проблемной среды

Еще раз нужно отметить, что как в предыдущем изложении, так и в дальнейшем часто используется термин модель, а не отображение среды в мозге или описание среды, потому, что формируемое в мозге отображение среды и себя в среде должно отражать динамику происходящих в среде процессов, быть активным и “жить” по законам среды, иначе говоря, быть “действующей моделью”. Представляется, что именно свойство активности, рассматриваемое на уровне нейронных механизмов, может позволить сделать определенные гипотетические интерпретации феномена мышления.

Для переноса какого-то отражения реальной действительности в мозг человека или животного с целью формирования в мозгу модели среды должен быть выбран информационный срез. Общая модель среды может включать много разных информационных срезов. В дальнейшем мы ограничимся, в основном, рассмотрением взаимосвязанных информационных срезов, определяемых задачей управления поведением, задачей восприятия и задачей мышления, с надеждой, что в рамках таких срезов можно будет выделить основные и достаточно общие характеристики и свойства строящихся в мозгу моделей проблемной среды. В то же время, в тех же целях остается полезным рассмотрение и более простых информационных срезов, относящихся к рефлекторному и инстинктивному поведению.

О большей части общих характеристик и свойств модели проблемной среды уже упоминалось при рассмотрении понятия информация и при описании особенностей задач поведения и восприятия. Это уже дает возможность выделить наиболее важные свойства и функции модели проблемной среды, используемой для зрительного восприятия и управления поведением. Перечислим основные необходимые свойства и функции.

1. Отображение целостных объектов.
2. Отображение ситуаций, действий и процессов в среде.
3. Многоуровневость.
4. Сведение многоэкстремальных задач поведения к одноэкстремальным путем формирования иерархических уровней обобщения и укрупнения.
5. Построение иерархического отображения как реальной, так и абстрактной среды в словесно-логической модели, т.е. создание и использование вербального уровня (языка).
6. Умозрительное моделирование или предвидение как реально происходящих, так и возможных ситуаций.
7. Эмоциональная оценка ситуаций и вариантов поведения.
8. Активность.
9. Отражение в модели себя и себя в среде.

Все перечисленные свойства и функции модели проблемной среды представляются необходимыми для реализации основных механизмов мышления. Подчеркнем лишь, что три последних свойства – эмоциональная оценка ситуаций и вариантов поведения, активность и отражение в модели себя и себя в среде являются необходимыми для моделирования сознания и воли, без которых, в первую очередь, невозможно сложное поведение и абстрактное мышление.

Представляется, что модель проблемной среды должна быть организована приблизительно следующим образом. Модель должна состоять из иерархически организованных смысловых структур, объединяющих взаимосвязанные смысловые элементы. Смысловые элементы это объекты, ситуации, действия, процессы и абстрактные понятия. Между смысловыми элементами модели имеются “вертикальные” и “горизонтальные” двунаправленные активизирующие связи. На вертикальных связях строятся иерархии часть-целое и частное общее и соответственно реализуются процессы анализа-синтеза и индукции- дедукции. Горизонтальные связи между смысловыми элементами во-первых, задают отношения между смысловыми элементами и, в первую очередь, между частями в рамках целого. Во-вторых, горизонтальные связи задают причинно-следственные зависимости типа если – то, необходимые для предвидения или умозрительного моделирования. И, наконец, на горизонтальных связях строятся ассоциации по сходству и ассоциации по смежности во времени.

Приведенные необходимые свойства модели проблемной среды будут в определенной степени уточняться и дополняться при рассмотрении некоторых гипотетических процессов, связанных с практическим (поведенческим) и творческим мышлением.

Относительно нейронных механизмов гипотетической модели полных конструктивных гипотез нет. Можно говорить о том, что смысловым элементам модели должны соответствовать нейронные ансамбли. Можно также говорить о более или менее общих свойствах, не выходящих, в принципе, за рамки представлений, сформулированных при описании одноуровневой А-сети. Кроме того, можно конечно говорить и о проблемах, решение которых необходимо для создания полной нейронной модели механизмов мозга. Эти проблемы приведены и обсуждаются в предыдущей лекции.

Полная нейронная модель механизмов мозга должна быть такой же активной системой, как и рассмотренная выше А-сеть. Единство всех процессов, связанных с построением и работой полной нейронной модели, должно основываться на принципе активности, направленном на оптимизацию состояния сети, построенной из

неустойчивых элементов Общая схема по сути представляется такой же, как и рассмотренная выше схема формирования первичной сети и вторичных нейронных сетей – нейронных ансамблей. в одноуровневой А-сети. Суть схемы всегда, как при обучении, так и при работе системы, сводится к формированию стабильных очагов возбуждения нейронов, мешающих процессу оптимизации функционального состояния сети, и к гашению очагов возбуждения. Нужно еще раз подчеркнуть, что при построении конкретных нейронных механизмов модели можно широко варьировать свойства нейронов, сохраняя основные – синергичность (взаимосодействие) и активность, основывающиеся на свойстве внутренней неустойчивости и общей направленности всех свойств на оптимизацию функционального состояния сети.

### **Перцептивное мышление.**

Основная информация о среде поступает в мозг посредством зрения. Восприятие, обеспечивающееся другими органами чувств – слух, осязание, обоняние имеют свою специфику, которая в этой книге рассматриваться не будет. В дальнейшем под словом восприятие будет пониматься, в первую очередь, зрительное восприятие, хотя основные принципы с определенной коррекцией распространяются и на другие сенсорные модальности. Задача зрительного восприятия уже довольно подробно рассматривалась в 6-й и 7-й лекциях. Поэтому ниже кратко затрагиваются только основные моменты, относящиеся к организации процесса и необходимой структуре модели среды.

Восприятие окружающей среды на основе модели проблемной среды служит для организации зачастую очень сложного поведения в среде. Акт восприятия может рассматриваться как актуализация какого-то фрагмента модели среды. Подобная актуализация может быть и умозрительной (воображение) или актуализацией во сне. Восприятие является целостным. Ситуация не разбивается на отдельные объекты, а воспринимается целиком, со всеми входящими в ситуацию объектами, находящимися в каких-то пространственных и других отношениях. Выделение в восприятии отдельных объектов происходит при поведенческой необходимости за счет направленной концентрации внимания, управляемой пока еще не вполне определенным уровнем воли или сознания. Принцип целостности восприятия распространяется не только на описание в модели и узнавание отдельных объектов, но и на восприятие ситуаций любого уровня укрупнения или детализации.

Таким образом, информативная модель проблемной среды должна включать иерархию целостных представлений. При распознавании объектов восприятия любого уровня укрупнения принцип целостности предполагает интерпретацию любых распознаваемых объектов в составе более крупных целостных образований - конструкций, отображающих те взаимосвязи из внешней (поведенческой) задачи, в которых участвуют распознаваемые объекты. Эти взаимосвязи образуют внешний контекст интерпретации. Использование внешнего контекста интерпретации позволяет не только правильно выделять и воспринимать те части изображения, которые допускают при их отдельном восприятии неоднозначную интерпретацию, но и повысить надежность восприятия всех объектов, задействованных в той или иной семантической конструкции на всех уровнях укрупнения информационной модели, за счет их целенаправленной и совместной интерпретации.

В общем плане, целостность информационной модели дает возможность не только правильно интерпретировать части в составе целого и во взаимодействии между собой, но и осуществлять активный целенаправленный процесс поиска и выделения информации, необходимой как для интерпретации объектов восприятия всех уровней укрупнения, так и для использования во внешней по отношению к восприятию задаче поведения. Еще одной особенностью задачи восприятия является необходимое предвидение ситуаций по модели среды и формирование на этой основе гипотез восприятия. Как уже говорилось, для реализации предвидения модель должна быть динамичной и “жить” по законам среды.

Из сказанного уже очевидно, что восприятие и классическое распознавание образов это далеко не одно и то же. Как уже отмечалось в 6-й лекции при обычном распознавании образов на вход распознающей системы поступает не изображение целостной ситуации, а каким-то образом выделенное изображение отдельного объекта или его какие-то признаки. Распознающая система должна по признакам объекта определить к какому из известных классов относится данное изображение. Такое распознавание это не целостное восприятие среды, и даже не целостное восприятие отдельных объектов, а классификация по признакам объектов, рассматриваемых по отдельности. Все недостатки как обычного признакового распознавания так и аналогичных последовательных “горизонтальных” подходов уже достаточно подробно обсуждались.

Во многих работах говоря о моделировании или о каком-то приблизительном описании мышления зачастую ограничиваются ссылками на распознавание образов, полагая, что это почти одно и то же. Это не так. Классическое распознавание образов это не только не мышление, но и не восприятие среды. В то же время следует отметить, что как признаковое, так и растровое распознавание может быть определенным важным компонентом в общей схеме восприятия. В свою очередь, восприятие среды это и не мышление в каком-то более или менее полном объеме, а лишь специфическая очень сложная и очень интересная его часть, проливающая, тем не менее, свет на структуру модели проблемной среды и некоторые важнейшие принципы организации работы механизмов мозга. Строящаяся на основе восприятия модель среды должна обеспечивать умозрительное моделирование (представление в воображении) реальных объектов и ситуаций. Соответствующие процессы можно назвать перцептивным мышлением.

Таким образом, из особенностей задачи восприятия следует необходимость специфической структурно-функциональной организации модели проблемной среды, обеспечивающей реализацию таких функций как:

- построение и использование целостных обобщенных структурных представлений объектов и ситуаций в модели среды;
- принципиальная отображаемость полных структурных описаний объектов восприятия;
- предвидение по модели среды и формирование на этой основе гипотез восприятия;
- целенаправленное взаимодействие со средой или ее изображением в процессе восприятия;
- многоуровневая иерархичность модели среды;
- использование в восприятии контекста, в том числе и внешнего контекста, задаваемого семантическими конструкциями уровней обобщения и укрупнения модели;
- управление восприятием с верхних понятийных уровней модели среды;
- активность восприятия.

Итак, перцептивное мышление и распознавание образов вещи совершенно разные. Процесс перцептивного мышления связан с восприятием, осознанием и пониманием реальной ситуации в среде. Перцептивное мышление это актуализация в мозге и вывод на уровень сознания части модели среды, включающей реально воспринимаемый органами чувств фрагмент среды. Возникающая при этом в сознании часть среды может выходить за пределы непосредственно воспринимаемого фрагмента как в пространстве, так и во времени. Кроме того, к перцептивному мышлению будем относить процессы умозрительного перцептивного конструирования воображаемых объектов и ситуаций. Поясним это на следующих примерах.

Я гляжу на экран и вижу появляющиеся на нем символы, слова и фразы. Я вижу монитор, клавиатуру, иногда мышку и поверхность стола. Периферическим зрением менее четко я вижу настольную лампу телефон и другие находящиеся на столе предметы. При желании я могу переключить внимание и вывести на уровень сознания с любой доступной мне степенью детализации любой из этих предметов. Не оборачиваясь назад я представляю, что находится у меня за спиной, но это представление является уже обобщенным и не таким точным и детальным. Я слышу направляющиеся ко мне шаги и представляю кто идет и даже догадываюсь зачем. Я могу представить себе не только ситуацию, которая будет через какое-то время, но и ситуацию, которая была в прошлом.

Описанные примеры относятся к актуализации фрагментов перцептивной части модели среды. Эти фрагменты могут включать как конкретные объекты и ситуации, так и представления разной степени обобщения и укрупнения. Кроме того, я могу себе представить не только то, что в явном виде отображено в моей модели проблемной среды. В моем умозрительном представлении могут возникать объекты и ситуации недоступные прямому восприятию. Я могу себе представить и то, что я никогда не видел и никогда не увижу, например, избушку на курьих ножках, паровоз, стоящий на рельсах вверх колесами, либо другие физически невозможные объекты и ситуации. Представления такого рода можно назвать умозрительным перцептивным конструированием.

Приведенные примеры умозрительного моделирования и умозрительного конструирования объединяет с процессами непосредственного реального восприятия то, что все они являются принципиально отображаемыми перцептивными представлениями хотя бы только в воображении.

Таким образом, перцептивное мышление включает как узнавание отдельных объектов так и целостное восприятие окружающей среды. но не ограничено только этим. Классическое распознавание образов, в том числе и реализованное на формальных нейронах, прямого отношения к перцептивному мышлению не имеет. Описанное выше “восприятие с пониманием” можно считать упрощенной, но в то же время важной частью перцептивного мышления. Восприятие и умозрительное моделирование среды в процессе перцептивного мышления должно основываться на тех же принципах целостности, целенаправленности и активности. Должны разворачиваться и взаимодействовать процессы “снизу-вверх” и “сверху-вниз” в иерархически организованной модели проблемной среды, отражающей двусторонние связи между частным и общим, а также между частями и целым. Восприятие должно умозрительно разворачиваться во времени и в пространстве за пределы отражающегося в мозге с помощью органов чувств фрагмента среды. Должно происходить управление процессом восприятия с понятийного уровня. Восприятие с пониманием должно основываться на предвидении – “акцептор восприятия”. Должен использоваться контекст и максимально полная семантическая модель проблемной среды.

Полная компьютерная реализация всего этого, тем более реализация на основе моделей нейронных механизмов, представляет собой проблему, очень далекую от решения. Еще большую проблему представляет собой нейронная компьютерная имитация второй составляющей перцептивного мышления, а именно имитация умозрительного перцептивного конструирования.

## **Когнитивное (познавательное) мышление**

Когнитивное мышление это формирование в мозге модели среды. Исторически, как в эволюционном плане, так и в онтогенезе, построение в нервной системе отображения проблемной среды начинается с формирования в ней простых конкретных чувственных образов объектов, ситуаций и действий. В нейронной модели смысловым элементом, т.е. конкретным объектам, ситуациям или действиям, должны соответствовать нейронные ансамбли

нижних уровней иерархии. Далее в модели строятся причинно-следственные связи между образами на основе непосредственных парных сочетаний во времени - временных ассоциаций. Эти связи могут строиться примерно так же, как и рассматриваемые в лекции 13 связи между последовательно возбуждающимися нейронными ансамблями А-сети. Там же рассмотрена и более сложная и более логически емкая схема формирования временных ассоциаций с использованием задержек при передаче сигналов между нейронами.

Конкретные чувственные образы и простые связи между ними формируются при отображении проблемной среды в нервной системе животных. У человека образное отображение среды формируется в правом полушарии головного мозга. Понятийное словесно-логическое отображение формируется в левом полушарии. Образная и понятийная модели функционально и структурно разделены, но работают как единое целое, взаимно дополняя друг друга. С этим связаны структурно-логические особенности модели проблемной среды.

Как уже отмечалось выше, отображение среды в мозге человека, должно быть многоуровневым и иерархическим. Построение иерархической многоуровневой системы представления объектов, ситуаций, действий и процессов происходит одновременно как по линии формирования укрупнений (часть-целое), так и по линии формирования обобщений (частное-общее). Входами на нейроны каждого уровня иерархии являются выходы с информационных объектов – нейронных ансамблей более конкретного (более низкого) уровня.

В такой схеме естественно реализуется связь между целым и входящими в него частями на основе функции объединения (И). Эти конструкции легко реализуют упрощенные ассоциации по сходству. Возможны также парные ассоциации по смежности во времени. Реализация в нейронной сети более сложных отношений между входящими в целое частями представляет важную принципиальную проблему. Формирование в модели отображения среды, включающего понятия высокого уровня укрупнения, обобщения и абстракции невозможно без вербализации, т.е. языка.

В гипотетической многоуровневой нейронной модели должна происходить активация смысловых элементов (нейронных ансамблей) по вертикали – снизу вверх, т.е. от частного к общему или индукция, и от частей к целому или синтез. Должны происходить и встречные возбуждения сверху вниз, т.е. от общего к частному или дедукция, и от целого к частям или анализ.

Построение в мозге модели проблемной среды происходит при взаимодействии процессов, называемых перцептивным мышлением, состоящем в восприятии и воображении, и когнитивным (познавательным) мышлением, состоящем в исследовании общих и частных свойств воспринимаемых объектов разных уровней обобщения и укрупнения.

Когнитивное мышление во многом опирается на наглядно-образное конструирование (воображение), а также и на наглядно-действенное конструирование. При наглядно-образном конструировании в воображении представляются объекты, в том числе могут устанавливаться непривычные или невероятные сочетания предметов и их свойств. Эти представления играют роль гипотез, требующих проверки. При наглядно-действенном конструировании происходит реальное физическое преобразование ситуации и опробывание свойств объектов. Все сказанное может относиться не только к реальной воспринимаемой или отображаемой среде, но и к формальным и абстрактным, например, математическим объектам.

Для того, чтобы модель “жила” по законам среды кроме общего свойства активности необходимо представление в модели действий и процессов, определяющих варианты возможных переходов от одной ситуации к другой. Иерархия действий и процессов необходима для того, чтобы на любом уровне укрупнения ситуаций между ними были представлены одношаговые переходы, т.е. укрупненные действия, а не цепочки, состоящие из последовательности действий и ситуаций. Обобщенные действия при их конкретизации порождают варианты, сравниваемые в процессе решения задач поведения и мышления.

Как уже говорилось, основная часть модели среды строится на основе личного опыта индивида. Что заставляет животное или человека получать из среды информацию, т.е. проводить эксперимент, выявлять причинно-следственные связи и структурировать эту информацию, т.е. обобщать, укрупнять и привязывать к уже существующей? Последнее в каком-то смысле эквивалентно построению теории.

Выше уже говорилось о том, что у живого организма имеется потребность не в получении информации вообще, а в построении модели проблемной среды, при этом строятся информационные срезы, определяемые его первичными и вторичными потребностями. Говорилось, что живой организм направленно и активно входит во взаимодействие со средой для удовлетворения всех своих потребностей, в том числе и для удовлетворения потребности в построении и дополнении модели среды.

Итак, познание среды активно и целенаправленно. Состоит оно в определении причинно-следственных связей, формировании гипотез о свойствах исследуемых объектов и в получении информации, необходимой для их проверки. Как уже говорилось, познание активно еще и потому, что

оно непосредственно направлено на гашение очагов возбуждения и оптимизацию функционального состояния нейронной сети. В конечном счете этот процесс должен приводить к формированию в мозге целостной модели среды.

Интересно в связи с этим привести цитату из словаря практического психолога (1998г.): “непрерывное генерирование взаимосвязанной системы познавательных гипотез, идущих навстречу внешним стимулам есть выражение активной природы образа мира – в противоположность традиционным представлениям о познавательных образах как возникающих в результате рефлекторных процессов – реактивных, развертывающихся в ответ на внешние воздействия”. Это очень созвучно излагаемому в книге и, в определенной степени, подкрепляет приводящиеся представления. Можно добавить еще что активный целенаправленный процесс познания имеет много общих черт с описанным в лекции 7 процессом активного целенаправленного восприятия.

## Практическое (поведенческое) мышление

Особенности задачи поведения

Мозг возник и развивался для решения задач поведения и восприятия. В начале курса высказывалось мнение, что для того, чтобы понять суть интеллекта ( временно отвлекаясь от механизмов реализации), его нужно рассматривать как средство, необходимое для решения особых “человеческих” задач и что в качестве таких задач в первую очередь полезно рассматривать задачу целенаправленного поведения и задачу восприятия.

Рассматривая задачу поведения можно выделить ее особенности. Главная особенность состоит в накапливающейся внутренней неустойчивости организма, т.е. в непрерывном росте потребностей, определяемом в конечном счете принципом устойчивого неравновесия Э.С.Бауэра. Из этого следует принцип управления внешним поведением, формулируемый как принцип оптимальности  $\max T$ .

Вторая важная особенность – это постоянная необходимость взаимодействия со средой в определенных целевых ситуациях удовлетворения потребностей. Из этого следует необходимость постоянной активности и целенаправленности.

Третья особенность это динамичность, изменчивость среды. Из этого следует необходимость выявления и фиксации причинно-следственных связей типа “если-то”.

Четвертая принципиальная особенность состоит в многоэкстремальности большинства задач поведения. Из этого следует необходимость сведения многопереборных многоэкстремальных задач к малопереборным одноэкстремальным, дающая возможность пользоваться локальным градиентным правилом принятия решения в ситуации выбора.

Мозг решает многоэкстремальные задачи поведения за счет построения многоуровневой иерархической модели проблемной среды. С использованием этой модели в процессе мышления ищутся оцениваются и сравниваются укрупненные варианты одношаговых переходов между текущей и целевой ситуациями. Для этого необходимы умозрительное моделирование и предвидение. (Эти вопросы еще будет рассматриваться).

Пятая особенность это многообразие и практически неповторяемость среды на конкретном уровне. Из этого следует необходимость формирования обобщенных знаний о среде в виде иерархии, выстраиваемой по линии частное-общее.

Шестая особенность задачи поведения это необходимость выбора и фиксации цели поведения (доминанты), а также необходимость оценки и сравнения вариантов достижения цели.

Ранее для случая упрощенной одноэкстремальной задачи было получено локальное правило выбора варианта поведения,

$$\min \Delta Z_j = \min \left( \sum_i \frac{\partial F(X)}{\partial x_i} U_i - \frac{\partial F(X)}{\partial x_j} V_j \right)$$

j

а также сводящиеся при определенных допущениях к этому же правилу формальные критерии “эмоциональной” оценки вариантов выбора, определяемые формулами:

$$Z = F(X) - F(0) \quad (1.2), \quad R = \sum_i \frac{U_i}{V_i} - 1 \quad (1.9) \text{ и}$$

$$E_j = - \Delta Z_j = m_j V_j - \sum_i m_i U_i \quad (1.10)$$

Из этих формул видно, что механизм принятия решения в ситуации выбора работает на переборе и сравнительной оценке вариантов даже в случае одноэкстремальной задачи.

Отмеченные особенности задачи поведения в первую очередь определяют структуру, функции и работу строящейся в мозге модели проблемной среды.

Нужно отметить, что модель проблемной среды должна обеспечивать и более простое поведение. В некоторых случаях эмоциональный механизм сравнения и выбора вариантов поведения может отключаться. Во-первых, это случаи экстремальных эмоциональных состояний, таких как: страх, горе, гнев, ярость и т.п. Как уже

отмечалось при ситуациях, соответствующих таким эмоциональным состояниям, обычный эмоциональный механизм принятия решения чаще всего бесполезен и должен быть заблокирован. Экстремальные эмоциональные состояния выполняют функцию прерывания. Совершающиеся при этом действия как правило просты, рефлекторны и безальтернативны, например, замереть, бежать, кричать, плакать и т.п.

Во-вторых, в часто повторяющихся стандартных ситуациях также можно действовать без перебора и сравнения вариантов - по шаблону. Если ситуации повторяются не точно появляется риск ошибиться. Риск оправдывается выигрышем времени. Элементарное стандартное поведение, т.е. поведение по схеме стимул – реакция, это в первую очередь безусловные и условные рефлексы, а также динамические стереотипы.

В схему стимул – реакция вписывается и более сложное чем рефлексы и простые динамические стереотипы стандартное поведение, например, иногда очень сложное инстинктивное поведение птиц, рыб и особенно насекомых. Как рефлекторное так и инстинктивное стандартное поведение часто с необходимостью включает не только восприятие простых сигнальных стимулов, но и восприятие сложных ситуаций, задаваемых (описываемых) на уровнях обобщения и укрупнения модели проблемной среды.

Однако, как уже отмечалось выше, используемая во многих работах по описанию или моделированию поведения схема стимул-реакция не пригодна для решения задачи сложного поведения. Поведение не ограничивается ответными реакциями на внешние стимулы и не преследует традиционно рассматриваемой цели “уравновешивания со средой”. Главная и постоянная задача поведения – это удовлетворение первичных и вторичных потребностей, для которого необходимо достижение целевых ситуаций в среде. Эта задача становится особенно сложной если среда динамична и многоэкстремальна.

В свете сказанного рассмотрим гипотетический процесс практического (поведенческого) мышления на типичном примере планирования и управления поведением в многоэкстремальной среде.

В психологии термин практическое мышление чаще всего относят к процессу постановки целей поведения и выработки планов действий. Для определенной привязки к гипотетическим механизмам мышления будем считать практическим (поведенческим) мышлением поиск и выбор лучшего по эмоциональной оценке пути, соединяющего в модели проблемной среды смысловые элементы, соответствующие текущей и целевой ситуациям. Эта же схема подходит для описания, постановки и умозрительного решения формальной задачи, не связанной с реальным поведением в среде. Для конкретности в дальнейшем будем рассматривать задачу управления поведением. Ниже рассматривается приближенная гипотетическая схема решения многоэкстремальных задач поведения с использованием многоуровневого иерархического отображения (модели) проблемной среды.

Итак, ввиду многоэкстремальности задач поведения и сходных задач управления в активных динамических системах их глобальное решение может быть практически невозможно. Однако, когда задача одноэкстремальна, она может быть решена с применением локального алгоритма, при котором не нужен просчет вариантов до конечной целевой ситуации, а ограниченным перебором определяется лишь наилучший одношаговый переход в каждом узле дерева траекторий, то есть движение осуществляется по градиенту целевой функции.

Каким же образом применить локальный алгоритм к многоэкстремальным задачам? Естественно, прямое применение локального алгоритма в этом случае не дает решения, то есть может приводить к ошибке. Так человек, начинающий учиться играть в шахматы, рассматривая возможности очередного хода, выбирает тот ход, который непосредственно приводит к локальной максимизации функции ценности, что очевидно, не может привести к успеху ввиду многоэкстремальности шахматной задачи.

Применять локальный алгоритм можно только после сведения многоэкстремальной задачи к одноэкстремальной. Многоэкстремальность может проявиться только тогда, когда система двигаясь от текущей ситуации к целевой, последовательно проходит промежуточные ситуации, в которых оценивается целевая функция. При переходе от текущей ситуации к целевой в один шаг многоэкстремальности возникнуть не может.

Таким образом, свести многоэкстремальную задачу к одноэкстремальной можно перейдя на новый информационный уровень модели проблемной среды. Этот новый информационный уровень всегда должен быть уровнем обобщения и укрупнения. Однако, просто перейти на более высокий уровень укрупнения недостаточно. Нужно перейти на такой уровень, где возможны *одношаговые* переходы между текущей и целевой ситуациями.

Рассмотрим пример. Пусть я нахожусь в Москве и должен по каким-то причинам быть завтра в Петербурге. Спланировать и оценить перемещение из Москвы в Петербург на уровне элементарных действий невозможно. Поэтому я сначала по имеющейся у меня модели среды выделяю и сравниваю наиболее укрупненные одношаговые переходы. При этом происходит взаимодействие словесно-логической и образной части модели.

Приближенно и грубо это можно представить себе следующим образом. В словесно-логической модели отражены одношаговые переходы, например, лететь самолетом, ехать на поезде, ехать на автомашине, идти пешком. При выборе между ними идет обращение к образной модели и в сознании возникают образы, соответствующие сравниваемым вариантам, и их эмоциональная оценка.

Вариант ехать на поезде связан с возникающей из памяти одной последовательностью образов, вариант лететь на самолете – с другой. Эти образы и их эмоциональная оценка зависят от конкретного опыта человека и у разных людей могут быть разными. Например, могут возникнуть такие конкретные образы: ночной поезд, сервис в вагоне, соседи в купе, световые мелькания за окном, плохой сон и т.д. Все это в совокупности дает общую

эмоциональную оценку. Таким образом, варианты одношаговых переходов порождает словесно-логическая модель, а их сравнение основывается на эмоциональных оценках образной модели.

Выбрав вариант одношагового перехода – ехать в Петербург на поезде, я намечаю достижимые в один шаг последовательные подцели укрупненного, но более низкого уровня: собрать необходимые вещи, ехать на вокзал, взять билет, сесть в поезд, ехать в поезде. Эта последовательность подцелей определяются выбранным одношаговым переходом верхнего уровня и сами являются одношаговыми переходами.

На каждом из этих одношаговых переходов я могу сравнивать и выбирать варианты, выделять подцели и одношаговые переходы более конкретного уровня вплоть до конкретных действий, например, “встать со стула”. В рамках каждого одношагового перехода возможна цепочка действий и ситуаций более конкретного уровня, на которых происходит изменение в том числе может происходить и ухудшение эмоциональной оценки рассматриваемой ситуации. Однако это уже не имеет решающего значения – за все “отвечает” более высокий и в первую очередь самый высокий уровень укрупнения модели среды.

На каждом уровне модели выбирается просто лучший из возможных одношаговых переходов, даже если все они имеют отрицательную эмоциональную оценку. На самом высоком уровне – уровне удовлетворения потребностей (в рассматриваемом случае потребность это быть завтра в Петербурге) выбираемый лучший одношаговый переход должен иметь положительную общую эмоциональную оценку, являющейся совокупной оценкой пути и цели. В противном случае человек (животное, система) в данной ситуации не справляется или плохо справляется с задачей удержания регулируемых переменных внутри области допустимых значений.

Итак, целевая функция максимизируется на наиболее высоком одноэкстремальном уровне модели проблемной среды. Таким образом реализуется локальный алгоритм принятия решения. При выборе решений с необходимостью делаются лишь предварительные приблизительные оценки изменения целевой функции, основанные на умозрительном моделировании и предвидении с использованием многоуровневой модели проблемной среды.

Описанная схема планирования и реализации поведения в многоэкстремальной среде, использует одношаговые переходы в многоуровневой иерархической модели проблемной среды. На основе сравнения одношаговых переходов и выбора лучшего из них по эмоциональной оценке в соответствии с принципом доминанты А.А.Ухтомского формируется и фиксируется главная поведенческая доминанта и тормозятся остальные. Доминанта по А.Ухтомскому это стабильно сохраняющийся очаг возбуждения. От главной доминанты, т.е. очага возбуждения самого верхнего уровня обобщения и укрупнения модели проблемной среды на более низкие уровни распространяется конус или “дерево” возбуждений и формируется ассоциативно-смысловая зона преактивации. Одновременно передается торможение на нейронные ансамбли, соответствующие альтернативным одношаговым переходам. На каждом уровне иерархической модели среды выбирается и фиксируется своя временная доминанта, соответствующая выбранному на этом уровне одношаговому переходу. Доминанта (очаг возбуждения) гасится при реальном выполнении соответствующего перехода.

Таким образом за счет оцениваемых одношаговых переходов достигается сведение многоэкстремальных задач поведения к одноэкстремальным, что лежит в основе процесса поведенческого мышления. К этой же схеме может сводиться и процесс мышления при решении многих умозрительных конкретных и абстрактных задач.

## **Структурно-логическая особенность модели проблемной среды человека**

Развитым мозгом и мышлением обладают животные и человек. В то же время мышление человека во многом отличается от мышления животных. В значительной степени эти различия определяются тем, что строящаяся в мозге модель проблемной среды человека принципиально отличается от модели проблемной среды животных. В мозгу животного строится так называемая образно-двигательная модель. У человека модель проблемной среды состоит из двух частей – словесно-логической и образной. Эти части имеют не только функциональное различие, но и четкое структурное и пространственное разделение. Словесно логическая модель расположена в левом полушарии коры головного мозга, образная модель – в правом. Соответственно в психологии различаются словесно-логическое и образное мышление.

Мне кажется, что можно и полезно вместо термина словесно-логическая модель применять применять термин понятийная модель.

Чем отличаются образная и понятийная модели и связанные с ними типы мышления? В понятийной модели входящие в нее объекты, ситуации, действия и процессы именуется. Это позволяет создавать иерархические многоуровневые обобщения и оперировать не только с конкретными единичным сущностями, но и с их классами. В понятийной модели возможно формирование понятий разной степени абстракции. Например, возможны такие понятия, как – “здоровье”, “ошибка” или “плохое обслуживание”. Часто считают, что словесно-логическое, а следовательно и понятийное мышление является последовательным. Это не совсем точно. Информация, содержащаяся в модели проблемной среды, не является последовательным описанием, а представляется в виде параллельной иерархической структуры. На этой структуре разворачиваются во многом параллельные процессы поиска путей между очагами возбуждения. Результатом такого поиска (мышления) становятся непротиворечивые в пределах имеющегося опыта причинно-следственные цепочки ситуаций и действий разных уровней обобщения, укрупнения и абстракции. Выбор пути между очагами возбуждения в модели также повидимому в основном

параллельный процесс. Последовательным может быть процесс проверки реализуемости или непротиворечивости цепочки переходов.

В образной модели отражаются конкретные объекты, ситуации и действия. Возможность построения обобщений и укрупнений минимально достаточна для процесса восприятия, в первую очередь зрительного. В нашей терминологии это только два или три уровня иерархии. Помимо зрительных образов могут восприниматься, запоминаться и воспроизводиться в воображении звуки, музыка и простая информация других сенсорных модальностей. Возможность восприятия и понимания более или менее сложной информации, выражаемой словами минимальна.

Образная модель работает параллельно. В образной модели происходит одномоментное многозначное целостное восприятие объектов и ситуаций. Образное мышление ограничивается преимущественно ассоциациями по сходству и по смежности во времени. И что очень важно, в образной модели происходят основные эмоциональные оценки, в том числе и оценки, на основании которых принимаются поведенческие решения в ситуации выбора. С этими особенностями образной модели связано то, что правое полушарие в основном отвечает за воображение и творчество. В нормальном здоровом мозге человека понятийная и образная модели работают совместно.

Очень приблизительно это взаимодействие может происходить следующим образом. Вспомним Маяковского: “Взвоят и замрут сирены над Гудзоном, будто бы решают выть или не выть. Лучше бы не выли, пассажирам сонным нужно просыпаться, думать, есть, любить”. Поэт написал слова, взятые из его понятийной модели. Слова эти появились на основании конкретных, эмоционально оцениваемых образов из его образной модели и их обобщений, сформированных в его понятийной модели. При этом поэту удалось создать типичные узнаваемые образы так, что у читателя эти слова вызывают из его образной модели его частные образы, их обобщение и эмоциональную оценку, близкие, в общем случае, образам и их эмоциональной оценке поэта.

Существует представление, состоящее в том, что для науки, в отличие от искусства, главным, а может быть и единственным, является сухое строгое словесно-логическое или, что то же самое, символично-логическое мышление. Это не верно. Для любой творческой деятельности и, может быть, в первую очередь для науки, необходимы фантазия, воображение, образное мышление, целостное, часто зрительное, представление проблемы, эмоциональная оценка. Многие ученые (Адамар, Пуанкаре, Пенроуз) пишут, что их творческое мышление - момент озарения (инсайт) часто не имеет словесного выражения. Это можно объяснить не только тем, что в творческом процессе на первом месте стоят образные представления, но и тем, что информация в мозге человека представлена в виде иерархической параллельной структурной модели. Это относится не только к образно-двигательной, но и к понятийной модели. Формулирование результатов творческого процесса в виде линейной символической (словесной) последовательности происходит на втором этапе мышления.

## **Механизмы простого репродуктивного практического мышления**

В психологии репродуктивным называется мышление, основанное на применении готовых знаний и умений. Практическое (поведенческое) репродуктивное мышление может быть как простым так и творческим. Текущая и целевая ситуация представляются в нейронной модели проблемной среды возбужденными нейронными ансамблями – очагами возбуждения. Путь в модели проблемной среды это последовательность переходов ситуация – действие - ситуация, выражающаяся в последовательном возбуждении нейронных ансамблей.

На качественном уровне процесс поиска и сравнения вариантов пути в многоуровневой модели среды рассмотрен выше. Представления о возможной реализации такого процесса в нейронной модели реальных механизмов мышления достаточно приблизительно. Очаги возбуждения это стабильная активация соответствующих нейронных ансамблей. Формироваться и фиксироваться главные поведенческие доминанты и соответствующие им очаги возбуждения должны при выборе между постоянными первичными или вторичными потребностями в процессе решения задачи оптимизации поведения по критерию maxT. Гаситься очаги возбуждения должны при удовлетворении потребностей, для которого необходимо либо достижение реальных целевых ситуаций в среде, либо, при постановке абстрактных задач, достижение ‘умозрительных’ целевых ситуаций в модели. Достижение целевых ситуаций в нейронной модели проблемной среды это последовательное возбуждение нейронных ансамблей, соответствующих переходам от текущей к целевой ситуации.

При поиске пути в модели проблемной среды сначала происходит выделение области поиска, а затем порождение и сравнение вариантов. Будем называть область поиска ассоциативно-смысловой зоной преактивации нейронных ансамблей. Выделение ассоциативно-смысловой области поиска не является абсолютно жестким. Осуществляется оно фоновой возбуждающей преактивацией нейронных ансамблей, связанных с нейронными ансамблями, непосредственно возбуждаемыми поставленной задачей, т.е. с главной доминантой и соответствующим очагом возбуждения, а также с подчиненными доминантами на более низких уровнях иерархической модели среды.

“Связанные” это в данном случае означает входящие в конус или “дерево” возбуждений преактивации и связанные в модели проблемной среды по “вертикали”, т.е. по линиям часть – целое и общее частное, либо связанные по “горизонтали”, т.е. отношениями типа причина-следствие или ассоциациями по сходству или смежности во времени. Таким образом, активируются не только ‘вертикальные’ связи между уровнями иерархии



внутри зоны доминантной преактивации, но и ассоциативно-смысловые “горизонтальные” связи, идущие как от главной доминанты, так и от производных доминант на всех уровнях модели проблемной среды.

Выделение области поиска направленно на то, чтобы рассматривать только осмысленные варианты и снизить вероятность рассмотрения вариантов, не имеющих отношения к решаемой задаче. Так при планировании поездки из Москвы в Петербург врядли нужно рассматривать варианты, включающие подводную лодку или ракетодром.

Порождение вариантов преимущественно в выделенной ассоциативно-смысловой зоне преактивации происходит на основе случайного перебора и взаимного торможения нейронных ансамблей, соответствующих альтернативным вариантам переходов между моделируемыми ситуациями. На все нейроны поступает дополнительное случайное возбуждающее воздействие, которое складывается с преактивацией нейронных ансамблей в области поиска. Будем называть это дополнительное случайное воздействие внутренним шумом. Можно предположить, что в мозге такие случайные внутренние возбуждающие воздействия, поступают на нейроны коры из каких-то других отделов головного мозга, например из ретикулярной формации.

Дополнительное случайное возбуждающее воздействие поступает на все нейронные ансамбли и не должно быть слишком большим, иначе оно “забьет” и нивелирует ассоциативно-смысловую преактивацию. Именно поэтому мышлению как правилу мешает сильный внешний шум.

. В свете всего сказанного, будем называть простым репродуктивным практическим (поведенческим) мышлением поиск по модели проблемной среды и выбор лучшего пути между текущей и целевой ситуациями, т.е. выбор одношаговых переходов и последовательных переходов между ситуациями на всех уровнях иерархии модели, при условии, что искомый путь существует в модели и лежит в зоне ассоциативно-смысловой преактивации. Найденный и все промежуточные пути реализуются в возбуждении соответствующих смысловых элементов модели проблемной среды, т.е. в возбуждении представляющих эти элементы модели нейронных ансамблей. Все сказанное относится как к задачам управления поведением в реальной среде, так и к решению абстрактных задач.

Таким образом, простое репродуктивное поведенческое мышление характеризуется тем, что искомый путь между очагами возбуждения существует в модели среды и находится в зоне ассоциативно-смысловой преактивации. Творческий вариант репродуктивного поведенческого мышления отличается тем, что искомый путь между очагами возбуждения существует в модели среды, но находится вне зоны ассоциативно-смысловой преактивации. Это обычно существенно затрудняет его поиск. Некоторые предположения об общей схеме творческого мышления рассматриваются ниже.

## **Репродуктивное творческое мышление**

Рассмотрим некоторые в принципе моделируемые и возможно основные аспекты творческого мышления. К репродуктивному творческому практическому (поведенческому) мышлению будем относить случаи, когда искомый путь между очагами возбуждения в модели проблемной среды существует, но лежит вне зоны ассоциативно-смысловой преактивации. Этот случай можно назвать творческим поиском. Типичный простейший пример необходимости расширения зоны ассоциативно-смысловой преактивации это задача: построить четыре равносторонних треугольника из шести спичек, требующая для своего решения перехода от напрашивающихся представлений на плоскости к представлениям в трехмерном пространстве.

Если искомый путь в модели проблемной среды существует, но находится вне зоны прямой ассоциативно-смысловой преактивации, то выход на него, т.е. возбуждение соответствующих не преактивируемых нейронных ансамблей, затруднен и является случайным событием, вероятность которого меньше, чем вероятность возбуждения преактивируемых нейронных ансамблей. Кроме того, в соответствии с принципом доминанты А.А.Ухтомского и по некоторым модельным гипотезам, от доминантных очагов возбуждения не только передаются активизирующие воздействия на все ансамбли, которые имеют прямые ассоциативные и смысловые связи с возбужденными, но также передается и торможение на все другие ансамбли. В результате возбуждение от внутреннего шума оказывается недостаточным для активации нужных нейронных ансамблей.

Такая ситуация может породить “зацикливание”, выражающееся в последовательном повторном возбуждении одних и тех же нейронных ансамблей, относящихся к зоне ассоциативно-смысловой преактивации. “Зацикливанию” может способствовать и соответствующее гипотезе об А-сетях снижение порогов возбудимости у часто возбуждающихся нейронов. Это часто и происходит у человека при решении трудных задач. Выход из зацикливания и активация искомого нейронного ансамбля соответствует состоянию, которое в психологии называется инсайт или озарение.

Факторов, способствующих возникновению инсайта несколько. Во-первых, это может быть повышение порога возбудимости у нейронов, входящих в ансамбли зоны ассоциативно-смысловой преактивации при отдыхе. При продолжительном отдыхе и особенно во сне могут нормализоваться и выравниваться пороги возбуждения нейронов по всей сети. (Вспомните – утро вечера мудреннее). Во-вторых, это, часто случайное, получение дополнительных возбуждений на нейроны искомого ансамбля через не прямые, иногда достаточно отдаленные ассоциации, зависящие от накопленной в модели среды информации. Эти дополнительные возбуждения могут создаваться слабо возбужденными и не выходящими на уровень сознания нейронными ансамблями. Последнее можно интерпретировать как влияние подсознания или интуицию. Интуиция делает поиск более направленным –

вспомните “мне кажется, или интуиция мне подсказывает, что в этом что-то есть”. Правда, интуиция, т.е. связь между нейронными ансамблями через ансамбль, слабо возбужденный и не выходящий на уровень сознания, может оказаться и ложной. В третьих, таким фактором может быть случайное внешнее возбуждение.

Таким образом, инсайт это случайное событие, вероятность которого повышается при совпадении во времени перечисленных факторов со случайным возбуждающим воздействием, поступающем от внутреннего “шума”. Результат выхода за пределы ассоциативно-смысловой зоны поиска, приводящий к решению творческой задачи, может запоминаться в модели проблемной среды путем образования новых связей между нейронными ансамблями. На этой основе новые сходные задачи могут решаться проще с использованием более широкой области ассоциативно-смысловых ассоциаций.

Творческому мышлению часто помогает и искусство. Искусство создает конкретные и в то же время типичные эмоционально оцениваемые образы. Одновременно могут предлагаться (зрителю, слушателю) варианты обобщения этих образов. При восприятии произведения искусства происходит процесс сотворчества, приводящий к расширению стандартных ассоциативно-смысловых связей и к построению новых обобщений. Это может происходить как в образной (в первую очередь) и словесно-логической частях модели среды, так и в связях между ними.

### **Креативное творческое мышление**

Второй вариант творческого мышления относится к случаю, когда искомый путь между очагами возбуждения в нейронной модели среды отсутствует, а в самой среде существует или может возникнуть в результате направленного преобразования среды. Конкретный простой пример это ставящаяся перед обезьяной задача: достать высоко подвешенный банан, для решения которой нужно представить, а затем построить сооружение из находящихся в помещении ящиков. (Животные, в первую очередь приматы и дельфины, могут решать не очень сложные творческие задачи). Отсутствие пути в модели проблемной среды, если в самой среде искомые ситуации и/или переходы между ситуациями существуют, либо могут быть созданы, означает, что модель не полна и ее нужно достраивать. Когнитивное (познавательное) и креативное (созидательное) мышление трудно делимы. Как, впрочем, и все другие условно выделяемые типы мышления. Будем считать, что когнитивное мышление направлено на отражение и обобщение причинно-следственных связей типа если-то, а креативное мышление направлено на отображение более сложных зависимостей, в том числе и поиск ответов на вопросы как и почему.

Достройка модели в обычном случае это отображение в ней конкретной среды в процессе ее изучения либо направленного изменения, т.е. отображение непосредственно воспринимаемых органами чувств новых объектов, ситуаций и действий. Достройка модели может происходить и при умозрительном перцептивном или абстрактном конструировании. В этом случае воображение поставляет требующие проверки гипотезы. При творческом мышлении достройка модели может состоять не только в отображении конкретных или абстрактных новых объектов, ситуаций и действий, но и в построении отсутствующих связей между объектами модели, а также в построении обобщений и укрупнений, т.е. в переходах между наблюдаемыми частными явлениями и их обобщениями (индукция, дедукция) и между частями и целым (синтез, анализ).

Примером такого творческого мышления является построение научной теории. Процесс мышления всегда основывается на умозрительной работе модели среды, т.е. на воображении. Творческое мышление первого типа (когнитивное), так же как и простое репродуктивное мышление, использует процесс, который может быть назван умозрительным моделированием. Творческое мышление второго типа (креативное) использует кроме того процесс, который может быть назван умозрительным конструированием. В творческом мышлении первого типа важное значение имеет случайное возбуждение необходимых нейронных ансамблей. Точно так же случайность и инсайт участвуют и в креативном мышлении. Однако вероятности возбуждений нужных нейронных ансамблей всегда зависят от существующих знаний (существующей модели проблемной среды) и от прорабатываемой направленной умственной работы.

Построение обобщений в общем творческом случае не должно сводиться к перечислению прецедентов, т.е. к построению объединения частных на основе логической функции ИЛИ, а должно фактически состоять в создании теории. Построение укрупнений в общем творческом случае не должно сводиться к перечислению частей целого, т.е. к построению объединения элементов на основе логической функции И, а должно состоять в отображении взаимосвязей частей между собой, в предсказании качеств целого и в отображении взаимосвязей между частями и качествами целого. Реализация и того и другого в нейронной или какой-нибудь иной модели весьма проблематична.

У человека построение обобщений и укрупнений происходит как в словесно-логической (понятийной), так и в чувственно-образной частях общей модели проблемной среды. В словесно-логической части (в левом полушарии) строится иерархия понятий, в чувственно-образной части (правое полушарие) строятся конкретные и обобщаемые образы. Обе части модели работают вместе и согласованно. Если нет патологии, то параллельной независимой работы, которая могла бы приводить к раздвоению личности не бывает.

Образные представления помогают формированию понятийных представлений и, как правило, опережают их. С образно-чувственным представлением объектов, ситуаций, действий и процессов в первую очередь

связывается их эмоциональная оценка. С образными представлениями чаще всего связано построение гипотез, с понятийными представлениями - проверка гипотез. Таким образом, понятийные и образные представления работают совместно и дополняют друг друга.

Второй вариант творческого мышления чаще всего реализуется в науке и искусстве. При этом, если наука, ориентируется, в первую очередь, на словесно-логическую понятийную модель, формируя в ней свои результаты и используя образную модель как вспомогательную, то искусство ориентируется, в первую очередь, на чувственно-образную модель, формируя в ней свои результаты и используя словесно-логическую модель как вспомогательную.

Особенность искусства состоит в том, что это не просто процесс творчества создателя художественного произведения, но и средство приобщения к процессу сотворчества других людей (зрителей, слушателей, читателей), помощь в формировании в их моделях мира эмоционально окрашенных образов и связей между ними в образной модели, помощь в обобщении образов, помощь в образовании и обобщении понятий и помощь в образовании связей между образами и понятиями.

Наиболее простые естественные примеры сотворчества дает театр. Для этого вида искусства характерно то, что процесс творчества актера и сотворчества (зрителя, слушателя) часто, хотя и не всегда, совпадают во времени. В других видах искусства когда создание и восприятие произведения разнесены во времени, если произведение понимается, то также происходит процесс сотворчества.

В творчестве много общего с простым мышлением. Так же как и при простом мышлении процесс запускается от очагов возбуждения. Так же ищутся или строятся в модели проблемной среды пути между очагами возбуждения. Так же как и при простом мышлении ведущую роль играет эмоциональная оценка, связывающаяся, в первую очередь, с гашением очагов возбуждения. Именно поэтому как художественное и научное творчество, так и просто изобретательство сопровождаются сначала мучениями, вызываемыми возникновением стабильных интенсивных очагов возбуждения, а затем получением положительных эмоций при их гашении. Точно так же возникновение очагов возбуждения и их гашение, сопровождаемое получением положительных эмоций, происходит в процессе сотворчества зрителя, слушателя, читателя.

Для первого варианта творческого мышления, соответствующего случаю когда искомым путем между очагами возбуждения в модели проблемной среды существует, но лежит за пределами зоны ассоциативно-смысловой преактивации, определенные конструктивные предположения о соответствующих модельных нейронных механизмах построить можно. Некоторые предварительные, безусловно требующие уточнения предположения приведены выше. В то же время, нужно еще раз подчеркнуть, что пока еще нет конструктивных (потенциально моделируемых) представлений о том, как на уровне активных нейронных механизмов происходят процессы творческого мышления во втором описанном случае, т.е. процессы связанные с умозрительным конструированием, с умозрительной достройкой иерархической модели проблемной среды, включающей построение новых объектов и связей.

## **Работа мозга и некоторые функции творчества**

Третий вид творчества ни понятийным, ни образным мышлением не является. Этот вид творчества реализуется в любом виде искусства, часто называется формой и может быть направлен на создание дополнительной эмоциональной окраски смысловой содержательной составляющей. Этот же вид творчества может быть и не связан со смысловой составляющей искусства и состоять только в эмоциональной окраске так называемого "свободного потока сознания". Наиболее четко это проявляется в музыке и абстрактной живописи.

По приведенным выше модельным представлениям эмоциональные оценки в первую очередь связываются с возникновением и удовлетворением потребностей. В свою очередь потребности связываются с возникновением стабильных очагов возбуждения и их гашением. Можно предположить, что в формальной составляющей музыки эмоциональное воздействие также связывается с созданием, кратковременным поддержанием и гашением очагов возбуждения. Временные очаги возбуждения могут возникать на основе предвидения и гасится при совпадении предвидения с приходящим внешним воздействием. Подобные процессы на уровне восприятия формы независимо от содержания, т.е. еще без смысловой нагрузки, или параллельно с ней, происходят, по видимому, и в других видах искусства.

Например в стихах есть ритм (закономерное чередование ударных и безударных слогов) и рифма. На основе и того и другого возникает предвидение. Ритм (как и в музыке) создает временные ассоциации (экстраполяции), рифма – ассоциации по сходству. Кроме того, рифма может являться знаком конца ритмической фразы. Процесс временных экстраполяций можно пытаться интерпретировать с использованием логики работы автомата, описанного в лекции 14. Предвидение выражается в преактивации определенных нейронных ансамблей. Совпадение предвидения с реальным событием создает положительное эмоциональное ощущение. Несовпадение, например, диссонанс, нарушение ритма или отсутствие рифмы сопровождается отрицательным эмоциональным ощущением.

Таким образом, к творчеству в искусстве относится не только создание у воспринимающего новых образной и понятийной моделей мира или активация их фрагментов, но и создание внешнего воздействия, прямо направленного на положительное эмоциональное восприятие в процессе оптимизации работы нейронной сети.

Кроме того, и музыка, и абстрактная живопись могут способствовать равномерному возбуждению нейронной сети, приводящему к оптимизации функциональных состояний нейронов. Эта составляющая внешнего возбуждения, может сопровождать восприятие произведения искусства и действовать непосредственно на нейронную сеть, запуская и поддерживая режимы оптимизации функциональных состояний нейронов и сети в целом. ... Приведенные упрощенные гипотетические представления о некоторых процессах творчества имеют очень общий характер. Нейрофизиологические интерпретации процессов творчества и особенно творчества в искусстве наверняка должны учитывать много профессиональных нюансов. С учетом этого было бы очень интересно определить количественные характеристики процессов возбуждения и изменения функциональных состояний нейронов мозга, например при слушании музыки, а также характеристики работы нейронной сети на разных уровнях иерархии отображения среды, и сравнить эти характеристики с количественными характеристиками музыкальной гармонии и диссонанса. Точно также, было бы интересно установить корреляции между физиологическими параметрами и количественными характеристиками вариантов формальной составляющей поэзии.

Воздействие искусства на мозг человека без конкретной смысловой нагрузки, например, воздействие музыки, далеко не исчерпывается приведенными упрощенными представлениями. Так одна и та же музыка может вызывать у разных людей разные конкретные или абстрактные образы. Однако эмоциональная окраска этих образов по шкале от печали до радости чаще всего очень близка. Механизмы этих воздействий неизвестны. Во всяком случае, вряд ли сейчас можно проинтерпретировать эти воздействия в таких терминах как предвидение, создание и гашение очагов возбуждения, неустойчивость и т.п.. Вот как описывает воздействие музыки Баха поэт М. Анчаров. “ И только музыкой не словами всколыхнулась земная твердь. Звуки поплыли над головами вкрадчивые как смерть. И звуки начали душу нежить и зов любви нарастал. И небыль, ненависть, нечисть, нежить бежали как от креста.”. Подобное восприятие музыки не только формализовать, но и попросту конкретизировать с использованием каких-то модельных представлений затруднительно.

### **Ненаправленная умственная работа**

И, наконец, несколько слов о ненаправленной умозрительной умственной работе. Точнее не о работе, а об отдыхе. Нейроны мозга работают, т.е. возбуждаются, всегда. Ненаправленная умозрительная работа происходит в состоянии бодрствования при временном отсутствии задач. (Постоянное отсутствие задач, то есть, потребностей, у активной поддерживающей неустойчивость динамической системы невозможно). Если очагов возбуждения в нейронной сети нет, то под действием внутреннего шума происходит ненаправленная активация случайных отрывочных цепочек смысловых элементов модели проблемной среды - последовательных возбуждений нейронных ансамблей, соответствующих образам, понятиям и конкретным объектам. Происходит так называемый “случайный поток сознания”, не заданный внешней или внутренней задачей и состоящий из случайных обрывков мыслей, ассоциаций и представлений.

Этот процесс, поддерживаемый внутренними случайными возбуждающими воздействиями, способствует равномерному возбуждению нейронов, т.е. оптимизации функционального состояния сети. Однако, для полного отдыха всей сети необходима равномерная нагрузка всех нейронов, что может происходить только при полном отключении от всех потребностей и от восприятия внешних ситуаций, объектов и событий, например во сне. Этому же может способствовать и подача на вход помогающего внутреннему внешнему равномерного шума, создающего дополнительные случайные возбуждения нейронов сети и переключающему на себя внешнее восприятие. Таким внешним шумом может быть, например, свет костра или морской прибой. Как уже говорилось, отчасти и при соблюдении некоторых условий эту роль могут играть музыка и составляющая формы в живописи.

## **Лекция 16. Заключение**

### **Моделирование мышления миф или реальность?**

Бурная дискуссия по этому вопросу, проходившая в середине прошлого века ни к каким убедительным результатам не привела просто потому, что не было определено, что такое мышление. В конечном счете, все свелось к тесту Тьюринга и стало понятно, правда не сразу, что этот тест машина, в принципе, пройти может, но это еще ничего не значит.

Для того, чтобы мыслить по-человечески машина, во-первых, должна научиться решать неалгоритмические творческие задачи. Некоторые интерпретации процесса решения творческих задач мы рассматривали в предыдущей лекции. Следует еще раз подчеркнуть, что для варианта творческого мышления, соответствующего случаю необходимости расширения области ассоциативно-смысловой преактивации модели среды, определенные конструктивные предположения о механизмах реализации сделать можно. В то же время, для второго варианта творчества, когда для решения задачи требуется умозрительная достройка модели среды и, в первую очередь, достройка модели на уровнях обобщения и укрупнения, конструктивные гипотезы о механизмах реализации отсутствуют. В первую очередь это касается представлений о необходимом в этом случае умозрительном

конструировании. Правда это не означает, что такие представления о работе нейронных механизмов принципиально не могут быть созданы.

Во-вторых, для того, чтобы мыслить по человечески машина должна быть активной, иметь собственные потребности, мотивации и эмоциональный механизм целеполагания и выбора пути к цели на основе активной модели среды, умозрительного моделирования и умозрительного конструирования. Это в рамках рввиваемого в направлении Искусственный интеллект алгоритмического подхода невозможно. Стало также понятно что компьютер и мозг устроены и работают совершенно по-разному. Правда, к этому вопросу до сих пор еще время от времени возвращаются.

Важнейшей функцией мозга, отличающей творческое человеческое мышление от мышления компьютера часто называется сознание. Под сознанием обычно понимается нечто большее, чем просто активация актуального фрагмента нейронной модели проблемной среды. Это нечто большее, как правило, считается интуитивно понятным и четко не определяется. В последнее время вслед за Р.Пенроузом [12] в поисках не определенного четко и невычислимого феномена сознания часто опускаются вглубь нейрона на молекулярный уровень или на уровень нанобиологии и добавляют к нерешенным проблемам мышления нерешенные проблемы квантовой механики. Попутно иногда вновь всплывает ничем не обоснованный и, казалось бы, давно уже похороненный со времен Мак-Каллока и Розенблатта вопрос о биологических вычислениях внутри клетки, но теперь уже о вычислениях, реализуемых молекулярными клеточными автоматами. (См., например, предисловие Г.Г.Малинецкого к книге Р.Пенроуза “Новый ум короля” [12])

Утрируя можно сказать, что поиск сознания на микроуровне подчиняется примерно такой логике. Что такое мышление непонятно. Для мышления необходимо сознание или осознание. Что это такое тоже непонятно, но это неалгоритмично. Одновременно предполагается, что как устроены и работают нейроны и связи между ними на макроуровне в целом понятно, а именно: нейрон это пассивный суммирующий пороговый элемент осуществляющий простую нелинейную передаточную функцию, т.е. формирующий на своем выходе сигнал в зависимости от входов и передающий этот сигнал другим нейронам, но эти вычисляемые и алгоритмически реализуемые функции непонятного и неалгоритмичного сознания не объясняют. Поэтому в поисках сознания нужно опускаться на молекулярный уровень – уровень квантовой механики, где не все однозначно и не все вычислимо. Например, в нейроне есть микроуровень, содержащий мельчайшие элементы – тубулины. Как они устроены, как связаны с макроуровнем и для чего нужны непонятно. А не являются ли эти непонятные микроэлементы-тубулины или какие-то другие объекты квантовой механики носителями непонятного невычислимого сознания и как следствие необходимыми элементами непонятного мышления?

Все это очень интересно написано, но никакой даже приблизительно намечающейся теории работы мозга за этим не стоит. Кроме того, полезно вспомнить Эйнштейна: теория должна быть проста, насколько это возможно. Правда у Эйнштейна есть и продолжение фразы: но не проще. Рискую слишком упростить проблему, можно попробовать все же не опускаться на микроуровень и не использовать аппарат квантовой механики пока это не будет действительно необходимо.

Надежда на то, что для объяснения и даже почти полного моделирования мышления будет достаточно уровня объединенных в сеть синергических взаимодействующих нейронов основывается на гипотезах о том, что нейрон не просто пороговый сумматор, а активный элемент, поддерживающий оптимальное изменяющееся неравновесное функциональное состояние, имеющий изменяющийся порог возбудимости и связи, изменяющиеся в зависимости от параметров и возбуждений связываемых нейронов. Все это, в принципе, для каждого отдельного нейрона вычислимо. Но интегральные функции мозга определяются не отдельными нейронами, а специфически организованной активной нервной сетью, подверженной к тому же случайным шумовым воздействиям. Движущим механизмом мышления должна быть активность, порождаемая необходимостью поддерживать неустойчивость. Основанные на таких представлениях определенные интерпретации процессов мышления, в том числе и некоторых вариантов творческого мышления, приведены выше.

Отрицая возможность компьютерной имитации мышления и ставя сознание на первое место по важности, часто считают достаточным уже упоминавшийся аргумент, состоящий в неалгоритмичности сознания. Вполне возможно, что алгоритм полной реализации функции сознания построить трудно или вообще невозможно. Более того, мозг и в целом в большинстве своих высших функций, связанных с мышлением, не алгоритмичен. Уже отмечалось, что мозг не имеет ничего общего с компьютером и то, что делает активный мозг в процессе мышления ни реализацией каких-то алгоритмов, ни вычислениями не является. Однако это не значит, что мозг и мышление нельзя моделировать на компьютере. Например, рекуррентно пошагово вычисляемые цифровые модели физических объектов и процессов в них можно моделировать на вычислительных машинах с любой задаваемой точностью.

Все свойства живой системы можно получить в искусственной системе если воспроизвести в ней не только внешние, но и все внутренние свойства живого, т.е. создать полную копию. Можно ли это физически сделать на неорганической основе? Не известно. Может быть и нельзя. Однако, любой физический объект можно смоделировать программно с любой задаваемой точностью, если, конечно, воссоздавать не внешние функции и проявления, а полностью моделировать

внутреннее устройство. Возможно ли это применительно к мозгу? Пока еще имеющих деталей знаний недостаточно. Будет ли это возможно в будущем? Может быть. В настоящее время можно рассчитывать только на какие-то шаги в последовательном приближении. Важным шагом может стать моделирование свойства активности.

Таким образом, моделирование мозга как физического объекта и реализуемого мозгом не алгоритмического в целом процесса мышления в принципе возможно. При этом для целей моделирования строить особый сверхсложный специальный компьютер не нужно. Достаточно иметь обычный универсальный компьютер с необходимыми вычислительными ресурсами и рекуррентно вычислимую физическую модель элементов мозга.

Отрицая возможность компьютерной имитации мышления, нужно логически обосновывать невозможность объяснения разума (мышления, сознания, свободы воли, творчества) без введения некоторого нематериального субстрата. Такой субстрат может, например, иметь название – душа. Логического обоснования необходимости души для процесса мышления пока еще нет. Возможна также еще и логика, строящаяся на следующих допущениях. В мире есть вполне материальные, но не объясняемые современной наукой сущности и поэтому представляющиеся нематериальными. Можно даже допустить, что материальная основа этих сущностей в принципе человеком непостижима. И, наконец, последнее допущение может состоять в том, что такая материальная, но неизвестная, а может быть и непостижимая сущность лежит в основе жизни и мышления и, в частности, сознания. Такая непостижимая материальная сущность будет в принципе немоделируема. Однако, при обсуждении вопроса о возможности моделирования мышления, приведенные допущения могут рассматриваться в качестве аргументов только в случае их хотя бы минимального фактического или логического обоснования.

Тем не менее, многие считающие себя материалистами психологи, не вводя в явном виде представление о нематериальной душе, в то же время говорят (без доказательства) о принципиальной невозможности свести уровень психологического анализа к уровню анализа физиологического, психологические законы к законам деятельности нейронных механизмов мозга. Я думаю, что это неверно. Если жизнь имеет материальную основу и эта основа познаваема, то рассмотрение и изучение процессов в активных неравновесных нейронных сетях может со временем позволить интерпретировать на физиологическом уровне, многие, а в конечном итоге и все, психологические факты, законы и явления. Не исключено, что все интерпретации психологического уровня окажутся возможными и на уровне компьютерного моделирования мозга. В том числе и интерпретации таких особенностей работы мозга, как интуиция, инсайт, творчество и, даже, юмор. Во всяком случае, теоретические запреты возможности построения активной творческой мыслящей машины не определены ни на формальном ни на качественном уровне. Правда, создание такой машины если и возможно, то в очень отдаленной перспективе.

## Заключение

Итак, что такое мышление? Можем ли мы теперь дать короткий, исчерпывающий, и строгий ответ на этот вопрос? Нет, не можем. Человеческое (“живое”) мышление это сложный многоплановый процесс, в котором важно не только то, что мозг делает на функциональном уровне, но и то, как он это делает. Поэтому ответ неоднозначен и может быть сформулирован лишь на разных информационных срезах и на некоторых уровнях приближения. Рассмотренные нами модели не являются полными определениями, а дают, как правило, лишь схему, в рамках которой можно обсуждать рассматриваемые явления по возможности в их существенных и общих аспектах. Таким образом, с учетом изложенного в книге можно дать много разных определений.

Можно, например, дать такое достаточно общее качественное определение: мышление это процесс решения задач поведения, восприятия или абстрактного в том числе и творческого мышления реализующийся в мозге (человека), обладающем сознанием, пониманием и активностью. Приведенное определение очень обще и требует интерпретации понятий сознание, понимание, активность и творчество. Можно, например, дать и такое достаточно общее качественное определение: мышление это процесс построения иерархической модели проблемной среды и использования этой модели для решения задач поведения или абстрактного мышления путем сведения многопереборных многоэкстремальных задач к малопереборным одноэкстремальным. Приведенное определение очень обще и требует многих интерпретаций, в том числе интерпретаций понятий многоэкстремальность и малоэкстремальность задач поведения. Выделяя главное, в самом обобщенном виде можно также сказать, что мышление характеризуется пониманием на основе использования иерархической семантической модели проблемной среды, умозрительным моделированием на основе построения и использования семантической модели проблемной среды и активностью как при построении модели среды, так и при решении задач на основе синергического взаимодействия неустойчивых неравновесных “живых” элементов нервной системы. Возможны и другие определения, например, ставящие на первое место процессы творчества, или специфику работы активных нейронных механизмов мозга в процессе мышления, например, процессы возникновения и гашения очагов возбуждения. Эти определения, как и приведенные выше, также требуют многих сложных интерпретаций.

Таким интерпретациям и посвящено основное содержание лекций. Я думаю, что рассмотренные нами модели и проблемы и сформулированные на этой основе гипотетические представления в определенной степени приближают нас к пониманию феномена мышления.

Мыслящим мозгом обладают только животные и человек. Поэтому, к объяснению мышления естественно идти от понимания различия между живым и не живым. Кстати, при обсуждении вопроса “может ли машина мыслить?” нужно определить не только термин “мыслить”, но и термин “машина”. В этом плане главным является представление о том, что живое строится из обладающих свободной энергией активных неустойчивых элементов и что в основе всего живого лежит сформулированный Э.С.Бауэром принцип устойчивого неравновесия. Это приводит к ряду выводов. На уровне целого организма выводы таковы.

Первый: живой организм не только на клеточном уровне, но и в целом – это система с непрерывно накапливающейся внутренней неустойчивостью и активно поддерживаемым устойчивым неравновесием.

Второй: поддержание устойчивого неравновесия на уровне организма происходит в результате активного поведения в среде, и подчиняется принципу  $\max T$  – максимизации времени пребывания системы внутри условной области существования, определенной как область допустимых значений регулируемых переменных – первичных и вторичных потребностей.

Третий: живой организм должен быть постоянно активен – поведение непрерывная задача. Поведение не ограничивается ответными реакциями на внешние стимулы и не преследует цели “уравновешивания со средой”. Главная и постоянная задача поведения – это удовлетворение потребностей, для которого необходимо достижение целевых ситуаций в среде.

Что из этого следует для понимания мышления? Во-первых, рассмотрение задачи поведения позволяет сформулировать гипотезу о специфике основной решаемой мозгом задачи.

Человек и животные в отличие от растений существуют в многоэкстремальной среде. Именно необходимость достижения целевых ситуаций в многоэкстремальной среде и определила, в первую очередь, образование и развитие такого управляющего органа как мозг. Поэтому, к пониманию мышления ведет рассмотрение задачи управления поведением и построение на этой основе предположения о том, что главная решаемая мозгом первичная проблема - это формирование и использование многоуровневой иерархической модели среды для решения задачи управления поведением путем сведения многопереборных многоэкстремальных задач к малопереборным одноэкстремальным.

В результате такого сведения возникает возможность приближенно решать многоэкстремальные задачи с использованием локальных правил принятия решения в ситуации выбора, путем сравнения *одношаговых* переходов на разных уровнях обобщения и укрупнения иерархически организованной модели проблемной среды. Такого же рода организация и использование знаний необходимы, повидимому, не только для решения задачи управления поведением в сложной многоэкстремальной среде, но и при решении других задач, в том числе и задач абстрактного мышления.

Во-вторых, к объяснению мышления ведет и рассмотрение задачи восприятия, в результате которого оказалось возможным сформулировать гипотетические представления о полных целостных структурно-метрических моделях зрительного восприятия, а также об организации процесса “восприятия с пониманием”, основанного на управлении процессом с понятийного уровня. Это управление включает формирование гипотез (предвидение) и целенаправленное подтверждение гипотез, т.е. восприятие базируется на активных взаимодействующих процессах “сверху-вниз” или от модели к среде и “снизу-вверх” или от среды к модели.

Переходя от особенностей и принципов решения задач целого организма к механизмам реализации этих принципов, иными словами, от общих способов решения задач поведения и восприятия к нейронным механизмам мышления, можно сказать следующее. Рассмотрение всех процессов в живом организме с единых позиций позволяет сделать гипотетическое предположение о том, что активная работа мозга как при построении модели среды, так и при решении задач восприятия, поведения и абстрактного мышления, определяется его элементами - нейронами, обладающими изменяющимся функциональным состоянием и поддерживаемой внутренней неустойчивостью. Пассивное состояние нейронов приводит к ухудшению функционального состояния нейронной сети. В результате в мозге имеется постоянная внутренняя задача – оптимизация функционального состояния и разряды в нейронах, а также передача возбуждений и торможений по связям между нейронами являются средством для решения этой задачи.

Из этого вытекает необходимость синергичности и постоянной активности нейронной сети. Это же предположение позволяет сформулировать ряд гипотез о том, как в нейронной сети происходит организация нейронных ансамблей, необходимая как для первичной оптимизации функциональных состояний отдельных нейронов и сети в целом, так и для последующей оптимизации нейронной сети в процессе получения, обработки и хранения информации.

Из этих гипотез также следует, что целенаправленное получение информации, необходимой для построения и коррекции целостной модели проблемной среды, является постоянной потребностью живого организма. Такой же потребностью является и решение мозгом задач, состоящее в нахождении наилучшего в каком-то отношении (по эмоциональной оценке) пути между очагами возбуждения. Очаги возбуждения и путь между ними отражаются

в мозге в виде возбужденных нейронных ансамблей, соответствующих текущей, целевой и промежуточным ситуациям.

Удовлетворение этих потребностей как при формировании модели проблемной среды, так и при решении мозгом задач приводит к нормализации функционального состояния нейронов, что, в свою очередь, ведет к нормализации функционального состояния всей нейронной сети.

Все затронутые проблемы рассматривались, по возможности, с общих позиций. Это обеспечивалось тем, что везде при объяснении тех или иных явлений, сущностей или вопросов во главу угла ставилось представление об изменяющемся функциональном состоянии и внутренней неустойчивости, а также принципы целостности, целенаправленности и активности, являющиеся, в конечном счете, производными от общего принципа оптимальности –  $\max T$ . Я надеюсь, что удалось показать роль и значение этих одинаково важных принципов. В то же время, хочется подчеркнуть, что главное необходимое свойство существования живых организмов, проявляющееся не только в восприятии и поведении, но и в любых процессах умственной деятельности это активность.

Я полагаю, что активность это главное необходимое свойство мышления. Означает ли это, что активностью и мышлением могут обладать только живые организмы? Нет, не означает. Увеличивающимися во времени и требующими удовлетворения потребностями, и соответственно, необходимой активностью могут обладать и искусственные системы. Могут обладать активностью и компьютерные модели нейронных механизмов мозга. Это в какой-то степени я пытался показать в предыдущих лекциях..

Представляется, что синергичность (взаимодействие) и активность как необходимые свойства в первую очередь должны лежать в основе моделирования механизмов мозга. Однако, следует еще раз подчеркнуть, что из этих необходимых свойств все функции мышления автоматически не вытекают. На уровне модельных механизмов мышления нужно еще очень многое осмысливать, придумывать и проверять.

Итак, подводя самый общий итог, можно еще раз сказать, что главное отличие работы мыслящего мозга от работы компьютера, запрограммированного на решение каких-то в том числе и так называемых интеллектуальных задач, это сознание, понимание и активность. Основное содержание лекций фактически состоит в интерпретации этих и сопряженных с ними понятий.

Понимание всех этих моментов не привело к построению законченной рабочей или хотя бы умозрительно работающей более или менее полной нейронной модели мозга, но позволило все же сформулировать некоторые гипотезы, а также выделить некоторые, возможно основные, нерешенные проблемы. Эти проблемы относятся, главным образом, к логической структуре модели проблемной среды, к некоторым процессам, реализующимся с использованием этой модели, а также к пониманию и моделированию устройства и работы активных нейронных механизмов.

Нерешенных принципиальных проблем, относящихся к уровню нейронного моделирования, много и они очень сложны. Рассмотрение этих проблем позволяет сделать вывод о том, что оптимистические оценки современного уровня нейрофизиологии, также как и современного уровня так называемого нейрокомпьютинга необоснованы.

Чем мышление человека отличаются от нынешних и возможных будущих компьютерных имитаций? На этот вопрос есть два разных ответа. Первый – принципиально ничем. Второй – отличаются принципиально, см., например, уже упоминавшуюся книгу Р. Пенроуза [12]. Сторонники второй точки зрения полагают, что живое и неживое вещи принципиально разные и, поэтому, моделировать живое в неживом невозможно. Как уже отмечалось, если эти принципиальные различия не носят нематериальный или непознаваемый характер и не состоят, например, в наличии у живого нематериальной души, а определяются все-таки чем-то познаваемым и материальным, то это материальное различие можно моделировать.

Более того, выше уже было определено, в чем, возможно, заключаются эти материальные и в принципе моделируемые различия. Я думаю, что нынешние компьютерные имитации работы мозга, а именно пассивные программные системы искусственного интеллекта и формальные распознающие нейронные сети, действительно принципиально отличаются от мышления человека и это отличие состоит, в первую очередь, в отсутствии устойчивого неравновесия и вытекающего из него как следствие активного динамического принципа перавновесной организации, в том числе в отсутствии неустойчивости, синергичности и активности на уровне решающих механизмов. В то же время, я думаю, что поддержание неравновесности и активность, могут быть реализованы в компьютерных решающих нейронных механизмах будущего. Эти свойства могут быть смоделированы и в современных универсальных компьютерах.

Часто говорят, что многие принципиально важные проявления человеческого мозга не алгоритмичны и по этой причине компьютерная имитация мозга невозможна. В первую очередь, в качестве таких проявлений называются сознание, творчество и свобода воли. Это, повидимому, действительно так, если имеется в виду алгоритмическая компьютерная имитация процесса мышления, сводящаяся к организации памяти компьютера и решению запрограммированных задач. Однако, уже отмечалось, что имитация может быть и совершенно иной, если строить, например, рекуррентно-вычислимую компьютерную модель мозга, рассматривая моделируемый мозг как физический объект. Такая модель мыслящего мозга может быть также как и живой мозг “эвристической машиной”, порождающей на основе иерархической модели реальной или абстрактной среды эвристики для



решения поведенческих или абстрактных задач. Эта модель должна эвристически решать все задачи, решаемые человеком, в том числе неалгоритмичные, невычислимые и творческие, правда, без гарантии решения.

В то же время возможно, что для полной имитации человеческого мышления потребуются в терминах неравновесность и активность более точно и конструктивно, чем это было нами сделано ранее, определить феномены сознания и свободы воли. Также необходимы, повидимому, будут конкретизация и моделирование эмоциональных механизмов.

Имитация работы и реальное построение 'живого' мыслящего мозга вещи разные. Интуитивно представляется, что смоделировать потребности, вычислить величину даже полностью соответствующую совокупности всех составляющих эмоциональной оценки ситуации и реальное ощущение эмоции это не одно и то же. Точно так же реальное или умозрительное эмоционально окрашенное осознание себя во взаимодействии с реальными или умозрительными объектами модели проблемной среды это не то же самое, что вычисление интенсивностей срабатывания модельных нейронов. Таким образом, очевидные отличия живого мозга от машины состоят в наличие чувств, ощущений и переживаний, сопровождающих эмоциональные оценки. Но это, скорее всего, специфические "живые" механизмы, необходимые для оценки, сравнения и выбора вариантов. Оценка, сравнение и выбор это процедуры, поддающиеся в принципе алгоритмизации. И, наконец, важнейшее отличие состоит в наличии у живого потребностей, мотиваций и эмоций, лежащих в основе целеполагания и поведенческой активности, подчиняющейся принципу  $\max T$ , т.е. направленной на выживание - сохранение своей качественной определенности.

Проявятся ли эти различия на функциональном уровне, то есть на уровне решения задач – это вопрос, остающийся пока еще открытым. Однако, можно еще раз отметить, что четкого логического обоснования необходимости какого-то нематериального (или непознанного) субстрата для объяснения работы мозга не существует. Не исключено, что строгое объяснение недостаточности (или достаточности) существующих физических знаний для полного понимания работы мозга возникнет при более глубоком исследовании мышления и, в том числе, феноменов сознания и свобода воли, а также нейронных механизмов умозрительного конструирования и активного выбора.

## Приложение

### **Актуальность проблемы - риск создания машинного интеллекта.**

Сейчас в прессе и научно-популярных статьях часто обсуждается проблема глобальных рисков, с которыми может столкнуться человечество. Одним из этих рисков называется искусственный интеллект. Предполагается, что "Большой искусственный интеллект" превзойдет по своему уровню интеллект естественный и сможет сам себя улучшать. Прогнозируются довольно близкие сроки. Например, 2030 год, но не дальше 2050 года. Все предположения о близких (постоянно отодвигающихся) сроках строятся со ссылкой на якобы имеющиеся успехи и производные развития алгоритмического (программного) искусственного интеллекта, распознавания образов и формальных нейронных сетей.

Обсуждается, каким будет искусственный интеллект по отношению к человеку – дружественным или не дружественным. В частности, предполагается, что машины с таким интеллектом могут вступить в конкуренцию и борьбу с человечеством. Предполагается, что результат борьбы "Большого искусственного интеллекта" с человеком может быть не в пользу человека. Существует международное общественное "Трансгуманитарное движение", одной из главных целей которого является подготовка к появлению Большого искусственного интеллекта и призыв к принятию мер, направленных на то, чтобы машинный интеллект был дружественным.

Эти проблемы многим представляются весьма актуальными. Особенно на фоне интенсивно развивающихся информационных и коммуникативных технологий, которые часто, особенно на бытовом уровне, путаются с технологиями искусственного интеллекта. Однако, полагаю, что можно всех успокоить. Обсуждаемая проблема не актуальна.

Искусственный интеллект сам по себе не будет ни дружественным, ни недружественным. Он будет пассивным. Таким же пассивным, как компьютерная шахматная программа, выигрывающая у чемпиона мира, но не осознающая этого и не стремящаяся к этому.

Кроме того, разрозненные результаты, получаемые в направлении работ, называемом "Искусственный интеллект", включая распознавание образов и формальные нейронные сети, активного мышления не образуют и к полноценному машинному мышлению не ведут. Никаких конструктивных наметок создания Большого искусственного интеллекта в настоящее время не существует. Производная развития теории искусственного мышления близка к нулю. Нет теории работы мозга в процессе мышления и в биологии. Из всего этого можно сделать вывод, что никакая "точка сингулярности" на кривой развития искусственного интеллекта в районе 2050 года не предвидится.

Таким образом, еще раз повторим, что проблема "риск создания машинного мышления" не актуальна. Станет ли эта проблема актуальной в будущем? Нет, не станет. Системы, которые могли бы соперничать и бороться с

людьми, должны были бы обладать потребностями, мотивациями, волей, активностью и эмоциональным механизмом управления поведением. Они должны были бы обладать и сознанием, т.е. осознанием себя как индивида или личности во взаимодействии со средой. Кроме того, эти системы должны обладать возможностью какого-то поведения, т.е. эффекторами. Кроме того, эти системы должны соперничать с людьми за какой-то ограниченный ресурс. То есть с органической жизнью на земле могла бы соперничать только какая-то другая жизнь.

Из этого следует, что в принципе можно обсуждать не надуманную проблему риска создания Большого искусственного интеллекта, а проблему риска искусственного создания альтернативной небелковой жизни. Правда, никаких конструктивных предпосылок для искусственного создания альтернативной жизни нет. При этом, возникает не только неразрешимый пока вопрос “как”, но также и вопрос “зачем”.

В то же время, следует отметить, что существует движение 2045. В рамках этого движения ставится цель создания альтернативной жизни - бессмертного “существа” с искусственным неорганическим телом и искусственным интеллектом. Это не обладающее физиологическими потребностями “существо” должно заменить смертного человека.

Планируемые конкретные работы не содержат ничего, что могло бы их принципиально отличить от давно ведущихся работ по созданию роботов, киборгов и моделированию мышления. Провозглашение метода обратного инжиниринга мозга без какой либо теории работы мозга в процессе мышления это попросту несерьезно. При этом, можно отметить, что ни разработка роботов, ни разработка так называемых киборгов, ни каких-либо других подобного рода систем на неорганической основе никакого отношения к созданию альтернативной жизни не имеет.

Неживые системы находятся в состоянии равновесия со средой, либо в движении к равновесию. (Искусственно поддерживаемое неравновесие в пассивных диссипативных самоорганизующихся системах это отдельный вопрос, который рассматривается в [23] и на котором мы здесь останавливаться не будем). Открытые, имеющие потребности живые активные самоорганизующиеся системы находятся в состоянии активно поддерживаемого неравновесия, т.е. в состоянии непрерывной борьбы со смертью. Равновесие это смерть. По-видимому, именно эта особенность биологической формы движения материи, т.е. особенность живых систем, является основой таких “виталистических” свойств живого, как активность, целенаправленность, сознание, воля, эмоции.

Неорганической искусственной системе, даже имеющей какой-то интеллект, для сохранения своего существования, в отличие от живого организма, не нужна постоянная активная самоорганизация, т.е. выполнение внутренней, и внешней работы по поддержанию своего неравновесного состояния [4]. Такое “существо” живым не будет. Оно не будет иметь не только потребностей, но повторим еще раз, не будет иметь и целей, мотиваций, желаний, активности, эмоций и ощущений. Не будет оно обладать и сознанием.

Таким образом, планируемые в рамках движения 2045 конкретные работы ничем не отличаются от уже давно ведущихся работ по созданию роботов, киборгов и искусственного интеллекта со всеми проблемами и недостатками этих работ. Точка технологической сингулярности, связываемая с созданием искусственного интеллекта, ни в обозримом, ни в отдаленном будущем не предвидится. В то же время, альтернативной жизни на основе планируемых работ построить нельзя. Интенсивно пропагандируемые цели движения 2045, никакого научного обоснования не имеют.

В добавление сделаем еще некоторые замечания по поводу положений, выдающихся идеологами движения 2045 за его научно-методическую основу.

Во-первых, часто ссылаются на продвижение в решении проблемы расшифровки мозговых кодов психических явлений. Можно ли говорить о задаче расшифровки мозговых кодов психических явлений? Думаю, что постановка такой задачи применительно к более или менее сложным психическим явлениям неправомерна. Кодов сложных психических явлений нет. Смысловым единицам иерархической нейронной модели мира, включающей модель собственного “Я”, соответствуют нейронные ансамбли – выделенные в сети группы нейронов, взаимодействующих через взаимовозбуждающие (не обязательно прямые) положительные обратные связи. Психические явления выражаются не кодами, а сложными взаимодействиями нейронных ансамблей в процессе их возбуждений и торможений. Эти взаимодействия рождают сознание и понимание. Как это делается нам не известно. Часто считается, что эти процессы неалгоритмичны и ставится под сомнение принципиальная возможность их формального выражения, а соответственно и возможность моделирования.

Кодироваться и декодироваться естественным образом могут описания или алгоритмы, зафиксированные на каком-то языке. Описания или передаваемые сообщения такого рода внутри мозга в процессе мышления не создаются. В плане расшифровки кодов психических явлений, по-видимому, можно решать лишь упрощенную задачу, а именно задачу установления причинно-следственных связей между возбуждениями каких-то групп нейронов и действиями в пределах простой схемы стимул-реакция. Но даже и в этом случае установление такой связи вряд ли можно назвать кодом психического явления.

Следующий опорный тезис - информация ивариантна по отношению к физическим свойствам своего носителя. Это так в случае если информация это зафиксированное или передаваемое описание. В случае целостной

нейронной модели мира информация как-бы “растворена” в своем физическом носителе и вне его существовать не может. При этом нейронная модель является одновременно и формой и содержанием информации. Кроме того, нейронная модель является инструментом для работы с информацией. Можно ли аналогичную модель с аналогичным представлением информации и аналогичными свойствами реализовать на другом субстрате или программно? Может быть и можно, хотя это очень далекая от решения проблема. Физическим субстратом, обеспечивающим аналогичные свойства, могла бы быть, например, альтернативная небелковая жизнь. Программная реализация скорее всего полностью аналогичной быть не может. Однако, все это требует отдельного обсуждения.

Третий опорный тезис это принцип изофункционализма – “ один и тот же комплекс функциональных отношений может быть воспроизведен на разных субстратах, т.е. системами, имеющими разные физические свойства”. Что такое комплекс функциональных отношений не определяется.

Говоря об этом принципе обычно ссылаются на Тьюринга. Я не нашел у Тьюринга такой формулировки. Известен тезис Тьюринга: каждая функция, вычислимая в обычном естественном смысле, может быть реализована машиной Тьюринга, или физический тезис Чёрча — Тьюринга: любая функция, которая может быть вычислена физическим устройством, может быть вычислена машиной Тьюринга. Эти тезисы нельзя ни строго доказать, ни опровергнуть. Принцип изофункционализма из этих тезисов не вытекает.

Кроме того, Тьюринг математик. Для него функция это связь между переменными или закон, ставящий в соответствие каждому элементу некоторого множества некоторый элемент другого множества.

Мышление не сводится к вычислению функций. Мышление это сложный многоплановый процесс. В лекциях рассматривались такие типы мышления как восприятие, обучение, управление поведением, решение формальных задач, творчество.

Все эти типы мышления (и тем более творческое мышление) не сводятся не только к формальным функциям, но, чаще всего, и к любому автоматному процессу, организованному по схеме вход-преобразование-выход. Что бы ни понималось под фразой “комплекс функциональных отношений” принцип изофункционализма либо не имеет отношения к тезисам Тьюринга и Черча, либо не имеет отношения к мозгу и процессу мышления.

Движение 2045 интенсивно пропагандируется (например, письма в правительство и ООН). Полностью проигнорировать это было бы неправильно. Не включая, естественно, эти вопросы в курс лекций, я вынес их в приложение. Часто говорят, что движение 2045 лженаучно. Это неверно. Никакой науки ни ложной, ни истинной там нет.

Все сказанное имеет отношение к общей проблеме моделирования мышления, но ни в коем случае не перечеркивает проблему. Моделировать активное человеческое мышление можно и нужно. Но это очень сложная и очень далекая от решения проблема. Это отдельная тема, для обсуждения которой нужно поднимать очень многие вопросы и, в частности, такие сложные вопросы, как жизнь и смерть, устойчивое неравновесие, активная антропийная самоорганизация, репродуктивное и творческое мышление, сознание, воля, активность, эмоции, ощущения, чувства. Эти вопросы в определенной мере затрагивались в лекциях.

## **Литература**

1. Анохин П.К. Узловые вопросы теории функциональных систем, М. Наука, 1980, 198 с.
2. Асмаян Н.В., Голицын Г.А., Особенности афферентного синтеза в условиях выбора. Вопросы радиоэлектроники, Сер. Бионика, 1967, вып.9, с.48–58.
3. Байков А.М., Кузин Е.С., Шамис А.Л.. Целостное целенаправленное распознавание изображений в ЭВМ. В сб. Вопросы кибернетики. Автоматизированные системы ввода-вывода графической информации. Москва, 1987, с.78–90.
4. Бауер Э.С., Теоретическая биология, М.-Л., ВИЭМ, 1935, 350с.
5. Бонгард М.М. Проблемы узнавания. М., Наука, 1967.
6. Н. Винер. Кибернетика. М. , 1968.

7. Вайнберг С. Мечты об окончательной теории. Москва, УРСС, 2004, 253с.
8. Голицын, В.М. Петров, Информация поведение творчество, Москва, Знание, 1991, с.222.
9. Емельянов-Ярославский Л.Б., Интеллектуальная квази-биологическая система Индуктивный автомат. М., Наука, 1990, 110 с.
10. Мак-Каллок У., Питтс В.. Логическое исчисление идей, относящихся к нервной активности. В сб. Нейронные сети: история развития теории. с.с. 5-22, М. Радиотехника, 2001
11. Психология машинного зрения, М,Мир,1978.
12. Минский М., Пайперт С.. Перцептроны М,Мир, 1971.
13. Пенроуз Р. Новый ум короля. Москва, УРСС, 2004,398с.
14. Розенблатт.Ф. Принципы нейродинамики. М.Мир.1965, 480 с.
15. Тьюринг А. Может ли машина мыслить?, М., физматгиз, 1960.
16. Шамис А.Л., Левит Б.Ю. Подход к построению формальной модели поведения, Сб. Механизмы и принципы целенаправленного поведения. Москва, Наука, 1972, с. 34-49.
17. Шамис А.Л. Пути моделирования мышления. М. Комкнига. 2006, 333с.
18. Шеннон К. Работы по теории информации и кибернетике, М.,1963.
19. Шредингер Э. Что такое жизнь с точки зрения физики. М. 1947.
20. Eccles J.C. The Neurophysiological Basis of Mind. Oxford University Press, 1953.
21. Hebb. D.O. The Organisation of Behavior. New York. 1949.
22. Hopfield J.J. Neural network and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci/ US. 1982. Vol 79/ P 2554-2558.
23. Шамис А.Л. Вектор эволюции - жизнь, эволюция, мышление с точки зрения программиста. М., УРСС, 2013.
24. Шамис А.Л. Модели поведения восприятия мышления. М., УРСС, 2010.
25. Николис Г., Пригожин И.. Самоорганизация в неравновесных системах: от диссипативных структур к упорядоченности через флуктуации. М.: Мир. 1979
26. Ухтомский А.А. Доминанта. СПб. Питер, 2002.