**On Intelligence**

**Jeff Hawkins**

with **Sandra Blakeslee**

Содержание

[Пролог 3](#__RefHeading___Toc142709371)

[1. Искусственный интеллект. 6](#__RefHeading___Toc142709372)

[2. Нейронные сети 11](#__RefHeading___Toc142709373)

[3. Человеческий Мозг 18](#__RefHeading___Toc142709374)

[4. Память 29](#__RefHeading___Toc142709375)

[5. Новая структура интеллекта 38](#__RefHeading___Toc142709376)

[6. Как работает кортекс 47](#__RefHeading___Toc142709377)

[6.1. Инвариантное представление 48](#__RefHeading___Toc142709378)

[6.2. Интеграция чувств 51](#__RefHeading___Toc142709379)

[6.3. Новая точка зрения на V1 53](#__RefHeading___Toc142709380)

[6.4. Модель мира 55](#__RefHeading___Toc142709381)

[6.5. Последовательности последовательностей 57](#__RefHeading___Toc142709382)

[6.6. На что похожа область кортекса 61](#__RefHeading___Toc142709383)

[6.7. Как работают области кортекса: детали 66](#__RefHeading___Toc142709384)

[6.8. Вверх и вниз 71](#__RefHeading___Toc142709385)

[6.9. Действительно ли обратные связи могут делать это? 72](#__RefHeading___Toc142709386)

[6.10. Как обучается кортекс 73](#__RefHeading___Toc142709387)

[6.11. Гиппокамп: вершина всего этого 75](#__RefHeading___Toc142709388)

[6.12. Альтернативный путь по иерархии 77](#__RefHeading___Toc142709389)

[6.13. Завершающие мысли 77](#__RefHeading___Toc142709390)

[7. Сознание и творчество 79](#__RefHeading___Toc142709391)

[7.1. Обладают ли животные интеллектом? 79](#__RefHeading___Toc142709392)

[7.2. Чем отличается интеллект человека? 80](#__RefHeading___Toc142709393)

[7.3. Что такое творчество? 81](#__RefHeading___Toc142709394)

[7.4. Являются ли одни люди более творческими, чем другие? 83](#__RefHeading___Toc142709395)

[7.5. Можем ли мы натренировать себя так, чтоб стать более творческими? 84](#__RefHeading___Toc142709396)

[7.6. Может ли творчество завести меня в тупик? Могу ли я обмануть сам себя? 85](#__RefHeading___Toc142709397)

[7.7. Что такое сознание? 85](#__RefHeading___Toc142709398)

[7.8. Что такое воображение? 88](#__RefHeading___Toc142709399)

[7.9. Что есть реальность? 89](#__RefHeading___Toc142709400)

[8. Будущее интеллекта 91](#__RefHeading___Toc142709401)

[8.1. Сможем ли мы построить интеллектуальные машины? 91](#__RefHeading___Toc142709402)

[8.2. Должны ли мы строить интеллектуальные машины? 94](#__RefHeading___Toc142709403)

[8.3. Зачем строить интеллектуальные машины? 96](#__RefHeading___Toc142709404)

[Эпилог 104](#__RefHeading___Toc142709405)

[Приложение: проверяемые предсказания 105](#__RefHeading___Toc142709406)

[Библиография 110](#__RefHeading___Toc142709407)

[Благодарности 114](#__RefHeading___Toc142709408)

[Об авторах 115](#__RefHeading___Toc142709409)

**Об Интеллекте**

# Пролог

**Эта** книга и моя жизнь наполнены двумя моими увлечениями.

В течение 25 лет я был увлечен мобильными компьютерами. В мире высоких технологий Силиконовой Долины я известен как зачинатель двух проектов - Palm Computing и Handspring, и как разработчик множества наладонных компьютеров и сотовых телефонов, таких как PalmPilot и Treo.

Но у меня есть и другое увлечение, которое предшествует моему увлечению компьютерами – и оно мне кажется более важным. Я увлечен изучением мозга. Я хочу понять, как работает мозг, не только с философской точки зрения, не только в общих чертах, а более детально. Мое желание – не только понять, что такое интеллект и как работает мозг, но и как построить машины, которые будут работать так же. Я хочу построить действительно интеллектуальную машину.

Вопрос об интеллекте – последний великий территориальный рубеж науки. Большинство важных научных вопросов затрагивают очень маленькие или очень большие масштабы, или события, происходившие миллиарды лет назад. Но мозг есть у всех. Вы – это ваш мозг. Если вы хотите понять, почему вы чувствуете, как вы осознаете мир, почему вы делаете ошибки, почему вы способны к творчеству, почему вас вдохновляет музыка и живопись, и вообще, что такое быть человеком – тогда вам нужно будет понять, что такое мозг. К тому же, успешная теория интеллекта и функций мозга будет иметь огромное общественное значение, и не только помогать лечить болезни, связанные с мозгом. Мы будем способны построить действительно интеллектуальные машины, хотя они совершенно не обязательно должны быть похожи на роботов и компьютеры из фантастических произведений. Наоборот, интеллектуальные машины будут базироваться на совершенно новых принципах о природе интеллекта. Как таковые, они помогут нам расширить наши знания о мире, помогут нам изучить вселенную, сделать мир лучше. А попутно будет создана огромная индустрия.

К счастью, мы живем в то время, когда проблема понимания интеллекта может быть решена. Наше поколение имеет доступ к горам информации о мозге, собранным за сотни лет, и скорость, с которой мы получаем новые данные, увеличивается. Только в США в области нейронаук работают тысячи ученых. У нас пока нет продуктивных теорий о том, что такое интеллект, или как работает мозг в целом. Большинство нейробиологов не задумывается о целостных теориях мозга, потому что они с головой ушли в эксперименты по сбору дополнительной информации о функционировании подсистем мозга. И хотя легионы программистов пытаются изготовить компьютерный интеллект, их попытки безуспешны. Я уверен, что их попытки останутся безуспешными до тех пор, пока они будут продолжать игнорировать разницу между компьютером и мозгом.

Чем же таким является интеллект, чего такого нет в компьютере, но есть в мозгу? Почему шестилетний ребенок может аккуратно перепрыгнуть с камня на камень через ручей, тогда как самые продвинутые роботы нашего времени – неуклюжие зомби? Почему трехлетний ребенок без проблем осваивает язык, тогда как компьютеры не могут, несмотря на полвека усилий лучших программистов? Почему вы можете отличить кошку от собаки за долю секунды, тогда как суперкомпьютер не может совсем? Это великие тайны, ожидающие ответов. У нас есть множество ключей к их разгадке; чего нам не хватает – это чуть более глубокого понимания.

Вы можете удивиться, почему разработчик компьютеров пишет книгу о мозге. Или с другой стороны, если я увлекаюсь изучением устройства мозга, почему я не сделал карьеру в нейронауке или в области ИИ? Ответ: я пытался несколько раз, но я отверг те способы изучения, которыми пользовались другие до меня. Я верю, что лучший способ решить проблему – использовать более детальную биологию мозга как ограничение и как путеводитель, но продолжая думать об интеллекте как о вычислительной проблеме – где-то между биологией и компьютерами. Многие биологи склонны отвергать или игнорировать идею рассуждать о функциях мозга в вычислительных терминах, а компьютерщики часто не верят, что им нужно знать что-то из биологии. К тому же, мир науки меньше приемлет риск, чем мир бизнеса. В технологическом бизнесе человек, преследующий новую идею с обоснованным подходом, может продвинуть свою карьеру, не смотря на то, что идея может не достигнуть успешного завершения. Многие предприниматели достигли успеха только после более ранних безуспешных попыток. Но в академических кругах несколько лет, потраченных на преследование бесплодной идеи, могут начисто разрушить вашу карьеру. Таким образом, я увлечен двумя идеями в моей жизни, веря, что успех в индустрии должен помочь мне достигнуть успеха в понимании работы мозга. Мне нужны финансы, чтоб заниматься той наукой, какой хочу, мне нужно изучать, как повлиять на мир, как продать новые идеи, до которых я дошел, работая в Силиконовой Долине.

В августе 2002 я основал исследовательский центр, Институт Нейронаук в Редвуде (RNI), занимающийся теорией мозга. В мире есть множество центров, занимающихся нейроисследованиями, но ни один из них не занимается поисками общего теоретического понимания работы неокортекса – части человеческого мозга, ответственной за интеллект. Именно это мы изучаем в RNI. По многим статьям RNI является начинающей компанией. Мы преследуем мечты, которые некоторым людям кажутся недостижимыми, но мы счастливы быть группой замечательных людей, усилия которых начинают приносить плоды.

\* \* \*

Намерения данной книги амбициозны. Она описывает всеобъемлющую теорию того, как работает мозг. В ней описывается, что такое интеллект и как он возникает в вашем мозгу. Представляемая мной теория не является полностью новой. Многие из отдельных идей, которые вы прочтете, уже существовали ранее в той или иной форме, но не согласованно друг с другом. Будьте готовы к этому. Говорят, что новые идеи - это чаще всего старые идеи, переупакованные и переинтерпретированные. Это определенно применимо к предложенной здесь теории, но упаковка и интерпретация могут создать существенное различие между кучей деталей и удовлетворительной теорией. Я надеюсь, что это произведет на вас впечатление, так же как и на других людей. Обычно я слышу такую реакцию: «Это впечатляет, я не мог и подумать об интеллекте таким образом, но после того, как вы мне это описали, я вижу, как это все укладывается вместе». С такими знаниями большинство людей начинают видеть самих себя немного по-другому. Вы начинаете наблюдать ваше собственное поведение, говоря: «я понимаю, что только что произошло в моей голове». Надеюсь, что когда вы прочтете эту книгу, у вас будет новое понимание того, почему вы думаете, что вы думаете и почему вы ведете себя именно так, а не иначе. Я также надеюсь, что некоторые читатели будут настроены на то, чтоб направить свою карьеру на построение интеллектуальных машин, основанных на принципах, изложенных на этих страницах.

Я часто ссылаюсь на эту теорию и на мой подход к изучению интеллекта как на «естественный интеллект», чтоб отличить его от «искусственного интеллекта». Специалисты в области ИИ пытаются запрограммировать компьютеры, чтоб они действовали как люди, не ответив сначала на вопрос, что такое интеллект и что означает понимание. Они оставляют в стороне наиболее важную часть в построении интеллектуальных машин – интеллект! «Естественный интеллект» указывает, что прежде чем построить интеллектуальные машины, мы должны сначала понять, как думает мозг, без чего-либо искусственного. Только потом мы можем спрашивать, как же нам построить интеллектуальные машины.

Книга начинается с рассмотрения того, почему предыдущие попытки в понимании интеллекта и построении интеллектуальных машин были безуспешными. Затем я введу и разовью коренную идею теории, которую я называю модель «память-предсказание». В главе 6 я детально покажу, как физический мозг воплощает модель «память-предсказание» – другими словами, как в действительности работает мозг. Затем мы обсудим социальные и другие следствия теории, которые для многих читателей может оказаться раздражающим разделом. Книга заканчивается обсуждением интеллектуальных машин – как мы можем их построить и на что будет похоже будущее. Я надеюсь, вы найдете это завораживающим. Вот несколько вопросов, которые мы затронем попутно:

***Могут ли компьютеры быть интеллектуальными?***

Десятилетия ученые в области ИИ заявляли, что компьютеры станут интеллектуальными, когда они станут достаточно мощными. Я так не думаю, и я объясню, почему. Мозг и компьютер занимаются совершенно разными вещами.

***Неужели предполагается, что нейронные сети приведут к интеллектуальным машинам?***

Конечно мозг – это нейронная сеть, но без понимания того, что делает мозг, простые нейронные сети будут не более успешными в создании интеллектуальных машин, чем компьютерные программы.

***Почему так сложно понять, как работает мозг?***

Многие ученые говорят, что мозг слишком сложен, и разобраться в нем займет у нас много времени. Я не согласен. Сложность – это всего лишь симптом неразберихи, но не причина. Я утверждаю, что у нас немного интуитивные, но неправильные предположения, которые ведут нас не туда. Вера в то, что интеллект определяется интеллектуальным поведением – это большая ошибка.

***Что же такое интеллект, если он не определяется поведением?***

Мозг использует громадное количество памяти, чтоб создать модель мира. Все, что вы знаете и изучили, хранится в этой модели. Мозг использует эту модель, основанную на памяти, чтоб осуществлять непрерывное предсказание будущих событий. Способность делать предсказания о будущем – основная проблема интеллекта. Я достаточно глубоко опишу способность мозга к предсказанию; это коренная идея данной книги.

***Как работает мозг?***

Место интеллекта – в неокортексе. Хотя у интеллекта огромное число возможностей и колоссальная гибкость, неокортекс имеет на удивление регулярную структуру. Различные части неокортекса, ответственные ли за зрение, слух, осязание или языковые способности – все работают по одним и тем же принципам. Ключ к пониманию неокортекса – понять эти общие принципы и, в особенности, их иерархическую структуру. Мы изучим неокортекс с достаточной детализацией, чтоб показать, как его структура вмещает структуру мира. Эта дискуссия будет наиболее технической частью книги, но заинтересованные неспециалисты должны ее понять.

***Каковы следствия этой теории?***

Эта теория мозга может помочь в объяснении многих вещей, например, почему мы способны к творчеству, почему мы осознаем наши ощущения, почему у нас бывают предубеждения, как мы обучаемся, и почему «старая собака» не способна выучить «новые трюки». Я обсужу несколько таких пунктов. В конце концов эта теория даст нам понимание того, кто мы такие и почему мы делаем то, что делаем.

***Можем ли мы построить интеллектуальные машины и что они будут делать?***

Да, мы можем и мы сделаем это. Ясно, что в течение нескольких десятилетий возможности таких машин стремительно возрастут и займут интересные направления. Некоторые люди боятся, что машины могут быть опасными для человечества, но я утверждаю обратное. Роботы не обгонят нас. Намного легче будет построить машины, которые превысят наши возможности в физике, математике, чем построить что-то наподобие ходячих говорящих роботов из фантастических произведений. Я исследую те невероятные направления, в которых эти технологии скорее всего пойдут.

Моя цель – объяснить эту теорию интеллекта и работы мозга таким способом, чтоб любой смог понять. Хорошая теория должна быть легкой для осмысления, не затуманенной жаргоном или скомканными рассуждениями. Я начну с базовых вещей и буду добавлять детали по ходу. Что-то будет основано только на логических рассуждениях; что-то – затронет некоторые аспекты устройства мозга. Некоторые из предложенных мною деталей определенно неверны, что нередко бывает в области науки. Уйдет много лет на то, чтоб разработать полноценную теорию, но это не уменьшает ценность коренной идеи.

\* \* \*

**К**огда много лет назад я начал интересоваться устройством мозга, я пошел в свою личную библиотеку в поисках хорошей книги, которая объяснила бы работу мозга. В юности я приучился доставать великолепные книги, объясняющие большинство интересующих меня вещей. Это были книги по теории относительности, про черные дыры, магию, книги по математике – все, что могло меня интересовать в данный момент. Однако, мои поиски удовлетворительной теории по работе мозга оказались безрезультатными. Я пришел к выводу, что ни у кого нет идей, как действительно работает мозг. Не было даже плохой или недоказанной теории; не было вообще никакой. Это было странно. Например, никто точно не знал, почему вымерли динозавры, но было изобилие теорий, о которых можно было прочитать. Но относительно работы мозга не было ничего похожего. Поначалу я не мог в это поверить. Меня донимало, что мы до сих пор не знаем, как работает этот важный орган. По мере изучения того, что уже было известно, я начал верить, что существует непосредственное объяснение. Мозг не был чем-то волшебным, и мне даже не казалось, что ответы могут быть сложными. Математик Пол Эрдос верил, что простейшие математические доказательства уже существуют в некой «божественной книге», и работа математика заключается в том, чтоб «прочитать эту книгу». Я был убежден, что объяснение интеллекта «где то там». Я готов был попробовать. Я хотел «прочесть книгу».

В течение 25 лет я видел образ той маленькой, простой «книги» о мозге. Это было похоже на приманку, которая манила меня все эти годы. Этот образ оформился в книгу, которую вы держите сейчас в своих руках. Мне никогда не нравилась сложность, ни в науке, ни в технике. Вы можете увидеть это в изделиях, спроектированных мной - часто отмечают простоту их использования. Все гениальное – просто. Таким образом, эта книга предлагает простую и непосредственную теорию интеллекта. Я надеюсь, она вам понравится.

# 1. Искусственный интеллект.

**Когда** я окончил Корнельский университет в 1979 году в качестве радиоинженера, у меня не было каких-либо грандиозных планов на жизнь. Я начал работать инженером в новом кампусе компании Intel в Портлэнде, штат Орегон. Микрокомпьютерная индустрия только что начиналась, и Intel был ее сердцем. Моя работа состояла в том, чтоб анализировать и устранять проблемы, найденные другими инженерами, работающими с нашей главной продукцией – одноплатными компьютерами (расположение целого компьютера на одной печатной плате только что стало возможным, благодаря изобретению микропроцессора компанией Intel). Я публиковал объявления, должен был путешествовать по командировкам и встречаться с потребителями. Я был молод, и для меня это были веселые деньки, хотя я скучал по своей возлюбленной из колледжа, которая получила работу в Цинцинатти.

Несколько месяцев спустя я встретил нечто изменившее направление моей жизни. Это «нечто» было только что опубликованным сентябрьским выпуском «*Scientific American*», посвященным полностью мозгу. Это вновь разожгло интересы моей юности относительно мозга. Это было великолепно. Оттуда я узнал об организации, развитии и химических процессах мозга, нейронных механизмах зрения, движения и других функций, а также о биологической основе психических расстройств. Это было одним из лучших выпусков «*Scientific American*» за все время. Несколько нейрофизиологов, с которыми я общался, сказали мне, что это сыграло значительную роль в выборе их карьеры, так же как и в моей.

Последняя глава, «Раздумья о Мозге», была написана Френсисом Криком, одним из соавторов открытия структуры ДНК, который затем обратил свой талант в направлении изучения мозга. Крик заявил, что назло устойчивому приросту знаний о мозге, до сих пор остается глубокой тайной, как же работает мозг. Ученые обычно не пишут о том, что им неизвестно, но Крика это не беспокоило. Он как мальчик, который сказал «а король то голый!». Согласно Крику, у нейроученых много данных, но нет теории. Его точные слова «чего очевидно не достает – так это идейного каркаса». Для меня это по-джентельменски сказанное «Мы понятия не имеем, как это работает». Это было верно и тогда, и остается верным сейчас.

Слова Крика придали мне силы. Желание моей жизни понять работу мозга и построить интеллектуальную машину было возрождено к жизни. Хотя я только что окончил колледж, я решил поменять карьеру. Я собрался изучить мозг, не только понять, как он работает, но и использовать эти знания как фундамент новых технологий, чтоб построить интеллектуальную машину. Ушло некоторое время, пока эти планы воплотились в действие.

Весной 1980 я переехал в Бостонский офис компании Intel, чтоб воссоединиться со своей будущей женой, которая поступила в ВУЗ. Я занимался обучением пользователей и служащих тому, как проектировать микропроцессорные системы. Но мое рвение было направлено к другой цели: я пытался понять, как же построить теорию мозга. Инженер во мне осознавал, что как только мы поймем, как работает мозг, что мы сможем построить искусственный мозг, из кремния. Я работал на компанию, разработавшую кремниевые чипы памяти и микропроцессор; таким образом, возможно, я мог бы заинтересовать Intel позволить потратить часть моего времени на размышления об интеллекте и о том, как спроектировать мозгоподобные чипы памяти. Я написал письмо председателю Intel, Гордону Муру. Содержимое письма было примерно таким:

Дорогой Доктор Мур,

Я предлагаю организовать исследовательскую группу, занимающуюся изучением работы мозга. Ее можно начать всего с одного участника – с меня – и так далее. Я уверен, что мы можем понять, как он работает. Когда-нибудь это станет большим бизнесом.

— Джеф Хокинс

Мур направил меня к главному ученому компании Intel, Теду Хоффу. Я вылетел в Калифорнию на встречу с ним и выложил свои планы по изучению мозга. Хофф был известен по двум вещам. Первая, о которой я был осведомлен – это разработка первого микропроцессора. Вторая, о которой я не знал до этого времени – его работа над теорией нейронных сетей. У Хоффа был опыт в искусственных нейронных сетях и в том, что с ними можно сделать. Я не был готов к этому. Выслушав мои предложения, он сказал, что не верит в то, что в обозримом будущем понять работу мозга, и что для Intel нет смысла поддерживать меня. Хофф был прав, потому что только сейчас, 25 лет спустя, мы только начинаем продвигаться в понимании мозга. В бизнесе время – это все. Однако, тогда я был слегка разочарован.

Я склонился к тому, чтоб с наименьшими потерями достигнуть своей цели. Работа над теорией мозга в Intel могла бы быть наилучшим выходом. Когда эта возможность была отвергнута, я стал искать другую. Я решил обратить внимание на Массачусетский Технологический Институт (MIT), который был известен своими исследованиями в области искусственного интеллекта и был удобно расположен по дороге. Это казалось великолепным совпадением. У меня большой опыт в вычислительной технике – «подходит». У меня желание построить интеллектуальную машину, «подходит». Я хочу сначала изучить мозг, чтоб понять, как он работает… «хмм.., с этим проблемы». Эта последняя цель, желание понять работу мозга, было ненужным в глазах ученых из лаборатории искусственного интеллекта MIT.

Это было все равно что ломиться сквозь стену. MIT был родиной искусственного интеллекта. Когда я подал заявление в MIT, он был домом для множества интересных людей, порабощенных идеей запрограммировать компьютер так, чтоб он демонстрировал интеллектуальное поведение. Для этих ученых зрение, язык, роботы и математика были всего лишь вычислительными проблемами. Компьютеры могли бы делать все, что мог бы мозг, и даже больше, зачем же ограничивать мышление биологическими недостатками природных компьютеров? Изучение мозга ограничило бы наше мышление. Они были уверены, что лучше изучать предельные ограничения вычислений, как наиболее выраженные в цифровых вычислительных машинах. Их «Святым Граалем» было желание написать компьютерные программы, которые сначала сравнялись бы, а затем и обогнали человеческие способности. Они выбрали подход «цель оправдывает средства»; их не интересовало, как в работает реальный мозг. Они гордились игнорированием нейробиологии.

Меня поразило, как совершенно неверным способом берутся за проблему. Интуитивно я чувствовал, что ИИ-подход не только будет безуспешным в создании программ, могущих то же, что и человек - он ни за что не объяснит нам, что же такое интеллект. Компьютер и мозг построены на совершенно различных принципах. Первый программируется, второй – самообучается. Первый должен точно и четко работать с любыми данными, второй обладает естественной гибкостью и толерантностью к сбоям. У первого есть центральный процессор, у второго – нет централизованного управления. Список различий можно продолжать и продолжать. Основная причина, по которой я думал, что компьютер не может быть интеллектуальным – это то, что я понимал, как работает компьютер, вплоть до уровня физических процессов в транзисторе, и эти знания давали мне сильное интуитивное ощущение, что компьютер и мозг фундаментально различны. Я не мог этого доказать, но я знал это настолько точно, насколько человек может что-либо интуитивно знать. В конечном счете, я был убежден, что ИИ может привести к полезным изделиям, но он не приведет к построению действительно интеллектуальных машин.

В отличие от этого, я хотел понять реальный интеллект и восприятие, изучить физиологию и анатомию мозга, принять вызов Френсиса Крика и представить миру четко определенную систему взглядов на то, как работает мозг. Я обратил свой взор в особенности на неокортекс – наиболее молодую часть мозга млекопитающих и место локализации интеллекта. После понимания того, как работает неокортекс, мы смогли бы продвинуться в построении интеллектуальных машин, но не раньше.

К несчастью, преподаватели и студенты, которых я встретил в MIT, не разделяли моих интересов. Они не верили, что необходимо изучать реальный мозг, чтоб понять интеллект и построить интеллектуальные машины. Так они мне и сказали. В 1981 году университет отклонил мое заявление.

\* \* \*

**Б**ольшинство людей сегодня верит, что ИИ-подход жив и здоров, и всего лишь ожидает достаточных компьютерных мощностей, чтоб оправдать свои многочисленные обещания. Когда компьютеры будут обладать достаточным объемом памяти и производительностью, продолжается мысль, программисты ИИ смогут сделать интеллектуальные машины. Я не согласен. ИИ-подход страдает от такого фундаментального недостатка, что он не может адекватно указать, что такое интеллект или что обозначает понимание чего-либо. Краткий взгляд на историю ИИ и на догма, на которых он построен, объяснят, почему это направление сбилось с курса.

ИИ-подход родился с появлением цифровых вычислительных машин. Ключевой фигурой в ИИ-движении был английский математик Алан Тьюринг, один из соавторов идеи компьютера общего назначения. Его великолепной работой стала формальная демонстрация концепции универсальных вычислений: то есть, все компьютеры фундаментально эквивалентны, несмотря на то, как они построены. Как часть своего доказательства, он придумал воображаемую машину из трех основных частей: процессорного блока, бумажной ленты и устройства, которое считывало и записывало метки на ленту, двигая ее взад и вперед. Лента предназначалась для хранения информации, наподобие компьютерных 1 и 0 (это было до изобретения чипов памяти и дисковых накопителей, так что Тьюринг вообразил бумажную ленту для хранения). Блок, который теперь мы называем центральным процессором (CPU), следовал фиксированному набору правил для чтения и изменения информации на ленте. Тьюринг математически доказал, что если вы выберете верный набор правил для процессорного блока и дадите ему бесконечно длинную ленту, он сможет выполнить любые определяемые множества операций во вселенной. Такая одна из многих эквивалентных машин называется Универсальной Машиной Тьюринга. Является ли задачей извлечение квадратного корня, вычисление баллистической траектории, компьютерная игра, рисование изображений или согласование банковской транзакции – в основе нее лежат единицы и нули, и любая Машина Тьюринга может быть запрограммирована, чтоб выполнять ее. Преобразование информации это преобразование информации. Все цифровые компьютеры эквивалентны

Вывод Тьюринга был бесспорно истинным и феноменально плодотворным. Вся компьютерная революция и все ее продукты базируются на нем. Позже Тьюринг обратился к вопросу как построить интеллектуальную машину. Он чувствовал, что компьютеры могут быть интеллектуальными, но не хотел вдаваться в аргументацию того, возможно это или нет. Он не только не задумывался, сможет ли он формально определить интеллект, он даже не пытался этого сделать. Вместо этого он предложил доказательство существования интеллекта, известный Тест Тьюринга: если сможет обмануть человека-экзаменатора, заставив его думать, что он общается также с человеком, то по определению компьютер интеллектуален. И таким образом, с Тестом Тьюринга в качестве мерила и с Машиной Тьюринга в качестве средства, Тьюринг помог стартовать направлению ИИ. Его центральная догма: мозг всего лишь другой тип компьютера. Не важно, как именно вы проектируете систему искусственного интеллекта, главное сымитировать поведение, подобное человеческому.

Сторонники ИИ увидели параллель между вычислениями и мышлением. Он говорят: - «Смотрите, большинство впечатляющих проявлений человеческого интеллекта несомненно затрагивают манипуляции абстрактными символами – и это именно то, что могут делать также и компьютеры. Что мы делаем, когда говорим или слушаем? Мы манипулируем ментальными символами, называемыми словами, используя хорошо определенные правила грамматики. Что мы делаем, когда играем в шахматы? Мы используем ментальные символы, которые представляют свойства и позиции различных шахматных фигур. Что мы делаем, когда мы смотрим? Мы используем ментальные символы для представления объектов, их положения, их названий и других свойств. Конечно, люди делают все это с помощью мозга, а не компьютера, но Тьюринг показал, что не важно, как вы реализуете манипуляцию символами. Вы можете делать это с помощью системы зубчиков и шестеренок, системы электронных ключей, или с помощью нейронной сети мозга – чем угодно, лишь бы ваше средство могло реализовать функциональный эквивалент Универсальной Машины Тьюринга».

Это предположение было поддержано важным научным документом, опубликованным в 1943 году нейрофизиологом Уорреном Мак-Каллоком и математиком Уолтером Питтсом. Они описали, как нейроны могли бы выполнять цифровые функции, то есть, как нервные клетки предположительно могли бы воспроизводить формальную логику компьютера. Идея заключалась в том, что нейроны могли бы выступать в качестве того, что инженеры называют логическими вентилями. Логические вентили реализуют простейшие логические операции, такие как И, НЕ, ИЛИ. Компьютерные чипы собраны из миллионов логических вентилей, соединенных в определенные сложные контуры. Процессор – это всего лишь набор логических вентилей.

Мак-Каллок и Питтс указали, что нейроны также могли бы быть соединены определенным образом, чтоб выполнять логические функции. Следовательно, нейроны собирают входные сигналы друг с друга и обрабатывают эти сигналы, чтоб решить, следует ли активировать выход, таким образом, предположительно нейроны могли бы быть живыми логическими вентилями. Таким образом, предположили они, мозг мог бы рассматриваться состоящим из И-вентилей, ИЛИ-вентилей и других логических элементов, построенных исключительно из нейронов, в прямой аналогии с соединением цифровых электронных контуров. Не ясно, действительно ли МакКаллок и Питтс верили, что мозг работает именно так, они всего лишь сказали, что так могло бы быть. И, логически рассуждая, такой взгляд на нейроны возможен. Нейроны могут теоретически реализовать цифровые функции. Однако никто не удосужился спросить, как же действительно соединены нейроны в мозгу. Они взяли в качестве доказательства, что несмотря на недостаточность биологических подтверждений, что мозг – всего лишь другой тип компьютера.

Нет ничего хуже, чем ИИ-философия, подкрепленная доминирующей точкой зрения психологии первой половины двадцатого века, называемой бихевиоризмом. Бихевиористы верили, что невозможно узнать, что творится внутри мозга, который они называли неприступным черным ящиком. Но возможно наблюдать и измерять окружение и поведение животного - что оно ощущает и что оно делает, его входы и выходы. Они признавали, что мозг содержит механизмы рефлексов, которые могли бы обуславливать адаптацию поведения животного через поощрения и наказания. Но что либо кроме этого они не считали необходимым изучать в мозге, особенно неопределенные субъективные переживания такие как голод, страх или то, что они значат для понимания чего либо. Излишне говорить, что такая исследовательская философия поблекла во второй половине двадцатого века, но ИИ продолжал гулять возле этого гораздо дольше.

Когда после Второй Мировой Войны электронные цифровые вычислительные машины стали доступны для широкого применения, пионеры ИИ засучили рукава и взялись за программирование. Перевод с языка на язык? Легко! Это всего лишь способ дешифрации. Нам просто надо отобразить каждый символ из Системы А в Систему Б. Зрение? Это тоже кажется легко. Мы уже знаем геометрические теоремы, которые оперируют с поворотами, масштабированием и смещением, и мы легко можем закодировать из в виде компьютерного алгоритма – и полдела сделано! Ученые мужи сделали делали заявления насчет того, как быстро компьютерный интеллект догонит и перегонит интеллект человеческий.

По иронии судьбы, наиболее всего приблизилась к прохождению Теста Тьюринга программа, которая называлась Элиза, прикидывающаяся психоаналитиком и перефразирующая ваши же фразы обратно вам. Например, если кто-то писал «Мой парень и я больше не общаемся», Элиза могла сказать «Расскажи мне о твоем парне» или «Почему ты думаешь, что ты и твой парень больше не общаетесь?». Разработанная в качестве шутки, эта программа действительно обманывала некоторых людей, хотя она была тупой и тривиальной. Более серьезные усилия были приложены к такой программе, как Блочный Мир, эмулирующей комнату с блоками различного цвета и формы. Вы могли задавать Блочному Миру вопросы типа «Есть ли зеленая пирамида на большом красном кубе?» или «Переместить синий куб на маленький красный куб». Программа должна была отвечать на ваши вопросы или пытаться выполнить то, что вы попросили. Она все это эмулировала и она работала. Но она была ограничена своим очень искусственным миром блоков. Программисты не могли обобщить ее, чтоб она делала что-то полезное.

Публика, меж тем, была впечатлена продолжительным потоком кажущихся успехов и новостей об ИИ-технологии. Одной из программ, вызвавшей возбуждение публики, была программа решения математических теорем. Даже начиная с Платона, многошаговые дедуктивные умозаключения виделись вершиной человеческого интеллекта, так что наперво казалось, что ИИ сорвал куш. Но, подобно Блочному Миру, программа оказалась ограниченной. Она могла найти только очень простые теоремы, которые уже были известны. Затем были большие телодвижения насчет «экспертных систем», БД фактов, которые могли отвечать на вопросы, заданные человеком-пользователем. Например, медицинские экспертные системы могли диагностировать болезнь пациента по заданному списку симптомов. Но снова оказалось, что она имеет ограниченное применение и не проявляет чего-либо близкого к обобщенному интеллекту. Компьютеры могли играть в шашки на уровне эксперта и в конечном счете IBM-овский Deep Blue превосходно обыграл Гари Каспарова, мирового чемпиона по шахматам, в его собственной игре. Но эти успехи ушли впустую. Deep Blue выиграла не за счет ума; он выиграл за счет того, что был в миллионы раз быстрее, чем человек. У Deep Blue нет интуиции. Опытный игрок смотрит на позицию на доске и сразу видит, какие варианты игры наиболее выгодны или опасны, тогда как компьютер не имеет врожденного чувства того, что важно, и должен исследовать гораздо больше вариантов. У Deep Blue также нет ощущения истории игры, и он не знал ничего о своем оппоненте. Он играл в шахматы так и не поняв, что такое шахматы, аналогично этому калькулятор выполняет арифметические операции, но понятия не имеет о математике.

Во всех случаях ИИ-программы были хороши только в одной определенной области, для которой они были разработаны. Они не обобщали и не выказывали гибкости, и даже их создатели признавались, что их программы не мыслят подобно человеку. Некоторые ИИ-проблемы, которые изначально казались легкими, не добились прогресса. Даже сегодня ни один компьютер не понимает язык так же хорошо, как может трехлетний ребенок, и не видит даже так, как может мышь.

По истечение многих лет усилий, неосуществленных обещаний и несостоявшихся успехов, ИИ начал блекнуть. Ученые из этой области ушли в другие области исследований. Компании, завязанные на ИИ, оказались неудачными. Вложения стали скудными. Стало казаться невозможным запрограммировать компьютер, чтоб он выполнял даже наиболее базовые задачи восприятия, языка и поведения. Сегодня немногое изменилось. Как я уже сказал ранее, до сих пор есть люди, которые верят, что ИИ-проблемы могут быть решены более быстрыми компьютерами, но большинство ученых думают, что в целом такие попытки ущербны.

Мы не должны порицать пионеров ИИ за их неудачи. Алан Тьюринг был блестящим человеком. Все могли бы сказать, что Машина Тьюринга должна изменить мир – и она сделала это, правда не путем ИИ.

\* \* \*

**М**ой скептицизм насчет притязаний ИИ был обострен примерно в то самое время, когда я подал заявление в MIT. Джон Серл, влиятельный профессор философии в Калифорнийском Университете в Беркли, в то время говорил, что компьютеры не были и не смогут быть интеллектуальными. Чтоб доказать это, в 1980 году он предложил мысленный эксперимент, называемый Китайской Комнатой. Это было примерно следующее:

Предположим, у вас есть комната с прорезью в одной из стен, и внутри находится англоговорящий человек, сидящий за столом. У него есть большая книга с инструкциями и все карандаши и бумага для черновиков, которые ему как-нибудь могли бы понадобиться. Перелистывая книгу, он видит, что инструкции, написанные на английском языке, указывают ему способы манипулирования, сортировки и сравнения китайских символов. Представим себе, что указания ничего не говорят о значении китайских символов; они только задают, как символы должны быть скопированы, стерты, переупорядочены, перекодированы, и т.д.

Кто-то снаружи просовывает кусочек бумаги через прорезь. На ней написан рассказ и вопросы по этому рассказу, все на китайском. Человек внутри не говорит и не читает ни слова по-китайски, но он берет бумажку и идет работать с книгой. Он трудится и трудится, следуя инструкциям в книге. В некоторых ситуациях инструкции говорят ему записывать символы на клочке бумаги, в других – перемещать и стирать символы. Применяя правило за правилом, записывая и стирая символы, человек работает до тех пор, пока инструкции из книги не скажут ему, что все готово. По окончании, наконец у него есть новая написанная им страница символов, которая без его ведома стала ответами на вопросы. Книга говорит ему передать эту бумагу через прорезь. Он делает это и удивляется, для чего было это утомительное упражнение.

Снаружи человек, говорящий на китайском, читает страницу. Ответы правильные, он замечает – даже проницательные. Если его спросят, исходят ли эти ответы от интеллектуального разума, он определенно скажет «да». Но прав ли он? Кто понял историю? Определенно, это не человек внутри, несведущий в китайском и понятия не имеющий, о чем история. Это не книга, которая просто хорошая книга, спокойно лежащая на письменном столе среди кипы бумаг. Так где же возникло понимание? Сеел говорит, что никакого понимания не возникало, это всего лишь связка бессмысленного листания страниц и черкания карандашом. А теперь чудесное превращение: Китайская Комната – это в точности цифровая вычислительная машина. Человек – это процессор, бездумно выполняющий инструкции, книга – это программа, дающая инструкции процессору, черновики – это память. Таким образом, не важно, как хитро спроектирован компьютер для эмуляции интеллекта путем имитации человекоподобного поведения, у него нет понимания и нет интеллекта. (Серл дает понять, что он не знает, что такое интеллект; он только говорит, чем бы он ни был, компьютер не может иметь этого).

Этот аргумент создал огромную пропасть между философами и апологетами ИИ. Это расплодило сотни статей, а в дополнение – еще больше сарказма и вражды. Защитники ИИ выступили с массой контраргументов Серлу, таких как заявление, что хотя ни один из компонентов комнаты не понимает китайский, комната в целом понимает, или что человек в комнате действительно понимает китайский, но просто не знает этого. Что относительно меня, я думаю, что Серл был прав. Когда я размышлял над аргументом Китайской Комнаты и о том, как работают компьютеры, я не видел понимания ни там ни там. Я был убежден, что нам необходимо понять, что такое «понимание», найти способ формально определить его, что могло бы прояснить является система интеллектуальной или нет, когда она понимает китайский, и когда – нет. Одно лишь ее поведение не сможет нам этого сказать.

Человеку не требуется «делать» что-то, чтоб понять рассказ. Я могу спокойно прочитать рассказ, и хотя нет явного поведения, мое понимание и осмысление наглядны, по крайней мере для меня. С другой стороны, исходя из моего спокойного поведения вы не сможете понять, понял я рассказ или нет, даже если я знаю язык, на котором написан рассказ. Вы могли бы позже спросить меня, чтоб узнать, понял ли я, но мое понимание возникло, когда я читал рассказ, а не когда я отвечал на ваши вопросы. Тезис этой книги в том, что понимание не может быть измерено внешним поведением; как мы увидим в следующих главах, это внутренняя метрика того, как мозг помнит что-либо и использует эти воспоминания для того, чтоб делать предсказания. Китайская Комната, Deep Blue и большинство компьютерных программ не делают ничего похожего на это. Они не понимают, что они делают. Единственный же способ, по которому мы можем судить, является ли компьютер интеллектуальным – это по его выходным данным, или поведению.

Последним аргументом в защиту ИИ является то, что компьютер теоретически мог бы эмулировать целый мозг. Компьютер мог бы смоделировать все нейроны и их соединения, и, если так, то не останется различий между интеллектом мозга и интеллектом, эмулированным на компьютере. Хотя, это может быть практически и невозможно, я с этим согласен. Но исследователи ИИ не эмулируют мозг, и их программы не интеллектуальны. Вы не можете эмулировать мозг без того, чтоб сначала понять, что он делает.

\* \* \*

**П**осле того, как я получил отказ и в Intel и в MIT, я не знал, что делать. Когда вы не знаете, как продолжать, часто наилучшей стратегией является ничего не делать, пока ваше мнение не прояснится. Таким образом, я продолжил работу в компьютерной отрасли. Меня удовлетворяло остаться в Бостоне, но в 1982 году моя жена захотела переехать в Калифорнию, так что мы переехали (снова путем наименьшего сопротивления). Я добился назначения в Силиконовой Долине, в проекте, названном Grid Systems. Grid занимался лэптопами, прекрасными машинами, которые стали первыми компьютерами в коллекции Музея Современного Искусства в Нью-Йорке. Работая сначала в маркетинге, а потом инженером, я создал в конце концов высокоуровневый язык программирования GridTask. Он и я становились все более и более важными для успеха Grid; моя карьера пошла в гору.

Я все еще не мог выкинуть из головы мое любопытство насчет мозга и интеллектуальных машин. Я был поглощен желанием изучать мозг. Так я прошел заочные курсы по психологии человека и обучался самостоятельно (никто еще не получал отказа от заочной школы!). После изучения порядочного объема биологии, я решил подать заявление в аспирантуру по биологии и изучать интеллект в биологических науках. Если компьютерные науки не нуждаются в мозговых теоретиках, то может быть мир биологии откроет двери для компьютерщика. Тогда еще не было такой вещи, как теоретическая биология, и в особенности теоретическая нейронаука, так что биофизика показалась наиболее подходящим полем для моих интересов. Я усердно учился, сдал требуемые вступительные экзамены, подготовил резюме, рекомендательные письма, и Вуаля!, я был принят на дневное отделение биофизики в Калифорнийский Университет в Беркли.

Я был взволнован. Наконец то я смог приступить к серьезной работе над теорией мозга, или мне так казалось. Я завершил свою работу в Grid без намерений работать снова в компьютерной индустрии. Конечно, это обозначало неопределенный отказ от моей зарплаты. Моя жена уже подумывала что «пора покупать дом и заводить семью», и я к счастью перестал быть кормильцем. Это определенно не было путем наименьшего сопротивления. Но это было наилучшим вариантом, который у меня был, и она поддержала мое решение.

Джон Элленби, исследователь из Grid, затащил меня в свой офис почти перед моим уходом и сказал: - «я знаю, ты не собираешься когда-либо возвращаться в Grid или компьютерную индустрию, но ты никогда не сможешь сказать заранее, что произойдет. Вместо того, чтоб полностью отказываться, почему бы тебе не взять отпуск? В таком случае, если через год или два ты надумаешь вернуться, ты сможешь сохранить зарплату, положение и возможность распоряжаться пакетом акций». Это было прекрасным предложением. Я принял его, но я чувствовал, что оставляю компьютерную индустрию навсегда.

# 2. Нейронные сети

**Когда** я поступил в Калифорнийский Университет в Беркли в январе 1986, первым, что я сделал, было собрание истории теорий интеллекта и функций мозга. Я прочел сотни статей анатомов, физиологов, философов, лингвистов, компьютерщиков и психологов. Множество людей из множества областей широко описали мышление и интеллект. В каждой область науки был свой набор журналов и в каждой использовалась своя терминология. Я нашел их описания неподходящими и неполными. Лингвисты говорили об интеллекте в терминах «синтаксис» и «семантика». Для них мозг и интеллект существовали только как язык. Специалисты по зрению ссылались на 2-мерные, 2.5-мерные, 3-хмерные модели. Для них мозг и интеллект существовали только как визуальное распознавание паттернов. Компьютерщики говорили про схемы и фреймы, новые термины, которые они придумали для представления данных. Никто из этих людей не говорил про структуру мозга и про то, как бы в нем укладывалась любая из их теорий. С другой стороны, анатомы и нейрофизиологи широко описали о структуре мозга и о том, как ведут себя нейроны, но они в большинстве своем избегали любых попыток построить крупномасштабную теорию. Было сложно и тщетно пытаться ухватить смысл этих разнообразных подходов и гор экспериментальных данных, сопровождающих их.

Примерно в это же время на сцену вышел новый обещающий подход к интеллекту. Нейронные сети существовали еще с конца 60-х в той или иной форме, но нейронные сети и движение ИИ были конкурентами, и за деньги и за умы агентов, занимающихся вложениями в исследования. Исследователи нейронных сетей существенно игнорировались исследовательскими фондами в течение нескольких лет. Тем не менее, немногие продолжали задумываться о них, и к середине 80-х они окончательно добились места под солнцем. Сложно сказать точно, откуда взялся неожиданный интерес к нейронным сетям, но несомненно одним содействующим фактором стали продолжительные неудачи в искусственном интеллекте. Люди прогнозировали насчет альтернатив к ИИ и нашли одну в искусственных нейронных сетях.

Нейронные сети были подлинным улучшением после ИИ-подхода, потому что их архитектура базируется, хотя и очень слабо, на реальных нейронных сетях. Вместо программирования компьютеров, исследователи нейронных сетей, известные также как коннекционисты, интересовались изучением того, какое поведение может быть продемонстрировано связыванием группы нейронов вместе. Мозг состоит из нейронов; таким образом, мозг – это нейронная сеть. Это факт. Надежда коннекционистов была в том, что эфемерные свойства интеллекта могли бы проясниться изучением того, как нейроны взаимодействуют, и некоторые из проблем, которые не решались с помощью ИИ, могли бы быть решены репликацией соответствующего соединения между популяциями нейронов. Нейронные сети отличаются от компьютера тем, что в них нет процессора, и они не хранят информацию в централизованной памяти. Знания и память сети распределены по ее соединениям – прямо как в реальном мозге.

На первый взгляд, нейронные сети казались наилучшим приложением моих интересов. Но я быстро развеял свои иллюзии в этой области. За это время у меня сформировалось мнение, что три вещи были существенными для понимания мозга. Первым моим критерием было включение времени в функции мозга. Реальный мозг быстро обрабатывает потоки информации. Нет ничего статического в потоках информации, поступающих в мозг и из мозга.

Вторым критерием была важность обратных связей. Нейроанатомы давно осведомлены о том, что мозг насыщен обратными связями. Например, в контуре между неокортексом и нижележащей структурой, называемой таламусом, количество соединений, идущих обратно (в сторону входа) превышает количество идущих в прямом направлении более чем в десять раз! То есть, на каждое волокно, передающее информацию в направлении неокортекса приходится десять волокон, передающих информацию в направлении органов чувств. Обратные связи преобладают также и в большинстве соединений внутри неокортекса. Никто не понял конкретную роль этих обратных связей, но из опубликованных исследований ясно, что они есть везде. Я полагаю, что это должно быть важным.

Третьим критерием было то, что теория или модель мозга должна учитывать физическую архитектуру мозга. Неокортекс не является простой структурой. Как мы увидим далее, он организован как повторяющаяся иерархия. Ни одна нейронная сеть, не обладающая такой структурой, определенно не будет работать как мозг.

Но как только феномен нейронных сетей покинул сцену, ее занял класс очень простых моделей, которые не удовлетворяли ни одному из этих критериев. Большинство нейронных сетей состоят из небольшого количества нейронов, соединенных в три слоя. Паттерн (вход) подается на первый слой. Эти входные нейроны соединены со следующим слоем нейронов, так называемым скрытым слоем. Скрытый слой затем соединяется с последним слоем нейронов, выходным. Соединения между нейронами имеют различный вес, подразумевая, что активность одного нейрона может усиливать активность другого нейрона и ослаблять активность третьего в зависимости от веса соединения. Изменяя эти веса, нейронная сеть обучается отображать входные паттерны на выходные.

Такие простые нейронные сети обрабатывали только статические паттерны, не использовали обратных связей и не были чем-то похожим на мозг. Наиболее общий тип нейронных сетей, называемых сетями «обратного распространения», обучались путем распространения ошибки от выходных слоев обратно к входным. Вы могли бы подумать, что это одна из форм обратной связи, но это не так. Обратное распространение ошибки возникало только во время фазы обучения. Когда нейронная сеть работала в обычном режиме после того, как она была обучена, информация распространялась только в одну сторону. От выходов к входам не было обратных связей. И у этих моделей не было времени. Статический входной паттерн преобразовывался в статический выходной. Затем подставлялся другой входной паттерн. В сети не было какой либо истории или записей о том, что происходило даже чуть раньше. И наконец архитектура этих нейронных сетей была тривиальной по сравнению со сложной и иерархической структурой мозга.

Я думал, что это направление быстро перейдет к более реалистичным сетям, но этого не произошло. Поскольку эти простые нейронные сети были способны на интересные вещи, казалось, что исследования остановятся прямо там на годы. Был найден новый и интересный инструмент, и неожиданно тысячи ученых, инженеров и студентов бросились получать гранты, зарабатывать ученые степени и писать книги о нейронных сетях. Были созданы компании, использующие нейронные сети для предсказания курсов на фондовых рынках, обработки заявок на получение кредитов, проверки подписей и выполнения сотен других приложений по классификации паттернов. Хотя намерения исследователей в этой области могли бы быть более общими, эта область стала доминирующей стараниями людей, которые не были заинтересованы в понимании того, как работает мозг или что такое интеллект.

Популярная пресса не понимала различий. Газеты, журналы и телевизионные научные программы представляли нейронные сети «мозгоподобными», или работающими по тем же принципам, что и мозг. В отличие от ИИ, где все должно было быть запрограммировано, нейронные сети обучались на примерах, что казалось более интеллектуальным. Наиболее заметной демонстрацией был NetTalk. Эта нейронная сеть обучалась отображать последовательности букв в произносимые звуки. Как только сеть была обучена на печатных примерах, она начала произносить «компьютерным» голосом читаемые слова. Легко было вообразить, что чуть больше времени – и нейронная сеть начнет общаться с людьми. NetTalk была ошибочно провозглашена в национальных новостях как машина, обучающаяся чтению. NetTalk была отменной демонстрацией, но то, что она действительно делала, было тривиальным. Она не читала, она не понимала и не имела большого практического значения. Она просто сопоставляла подходящие паттерны букв предопределенным звуковым паттернам.

Позвольте привести вам аналогию, чтоб показать, как далеки были нейронные сети от реального мозга. Вообразите, что вместо того, чтоб пытаться понять, как работает мозг, мы попытаемся понять, как работает цифровая вычислительная машина. Через годы изучения мы откроем, что все в компьютере состоит из транзисторов. В компьютере есть сотни миллионов транзисторов и они соединены между собой определенным сложным образом. Но мы не понимаем, как работает компьютер или почему транзисторы соединены именно так. Итак, однажды мы решаем соединить всего лишь несколько транзисторов, чтоб увидеть, что происходит. Вот смотрите, мы обнаружили, что несколько транзисторов, когда соединяются вместе определенным образом, становятся усилителем! Слабый сигнал, поданный на один конец, усиливается на другом конце (подобным образом делают усилители в радиоприемниках и телевизорах). Это важное открытие, и немедленно вырастает индустрия, выпускающая транзисторные радиоприемники, телевизоры и другие электронные приборы на транзисторных усилителях. Все это хорошо, но это ничего не говорит нам о том, как работает компьютер. Хотя и усилитель и компьютер сделаны из транзисторов, они не имеют почти ничего общего. Точно так же и реальный мозг и трехслойная нейронная сеть построены из нейронов, но не имеют почти ничего общего.

Летом 1987 я получил опыт, который еще больше охладил мой и так невысокий энтузиазм относительно нейронных сетей. Я пришел на конференцию по нейронным сетям, где я увидел презентацию, устроенную компанией, называемой Nestor. Nestor пыталась продать приложение на нейронной сети для распознавания рукописных символов на подложке. Она предлагала лицензию на программу за один миллион долларов. Это привлекло мое внимание. Хотя Nestor провела улучшение алгоритма ее нейронной сети и рекламировала ее как еще один большой прорыв, я чувствовал, что проблема распознавания рукописных символов могла бы быть решена более простым, более традиционным путем. Я пришел домой той ночью, размышляя о проблеме, и за два дня разработал распознаватель рукописных символов который был быстрым, маленьким и гибким. Мое решение не использовало нейронную сеть и оно работало совершенно не так, как мозг. Хотя эта конференция разожгла мой интерес в разработке компьютеров со стилусом (в конечном счете приведший к проекту PalmPilot десять лет спустя), это также убедило меня, что нейронные сети были не таким уж большим улучшением по сравнению с традиционными методами. Распознаватель рукописных символов, который я создал, пригодился в конечном счете для системы текстового ввода, названной Graffiti, использованной в первых сериях продукции Palm. Я думаю, компания Nestor ушла из бизнеса.

Слишком много для простых нейронных сетей. Большинство их возможностей легко воспроизводились другими методами и в итоге шумиха в СМИ спала. Наконец, исследователи нейронных сетей перестали объявлять свои модели интеллектуальными. Ведь это были очень простые нейронные сети и делали гораздо меньше, чем ИИ-программы. Я не хочу оставить у вас впечатление, что все нейронные сети являются простыми трехслойными. Некоторые исследователи продолжали изучать нейронные сети различных конфигураций. Сегодня термин *нейронные сети* используется для описания различных множеств моделей, некоторые из них являются биологически более точными, некоторые – нет. Но почти ни одна из них не является попыткой охватить полностью функции или архитектуру неокортекса.

По моему мнению, большинство фундаментальных проблем с большинством нейронных сетей – это особенности, которые они разделяют с ИИ-программами. И те и другие обременены тем, что фокусируются на поведении. Называют ли они это поведение «ответами», «паттернами» или «выходными данными», и ИИ и нейронные сети полагают, что интеллект заключается в поведении, которое программа или нейронная сеть демонстрирует после обработки заданных входных данных. Наиболее важным атрибутом компьютерной программы или нейронной сети является то, выдает ли она правильные или желаемые результаты. Как навязано Аланом Тьюрингом, интеллект эквивалентен поведению.

Но суть интеллекта не только в действии или интеллектуальном поведении. Поведение – это проявление интеллекта, но не центральная характеристика или первичное определение интеллекта. Рефлексия доказывает это: вы можете быть интеллектуальным просто лежа в темноте, размышляя и понимая. Игнорирование того, что происходит в вашей голове и фокусирование вместо этого на поведении было большой преградой на пути понимания интеллекта и построения интеллектуальных машин.

\* \* \*

**П**режде, чем мы исследуем новое определение интеллекта, я хочу рассказать вам о других коннекционистских подходах, которые подошли гораздо ближе к описанию того, как работает мозг. Проблема в том, что слишком мало людей, кажется, осознали важность этих исследований.

Тогда как нейронные сети привлекли основное внимание, небольшая отколовшаяся группа теоретиков по нейронным сетям строила сети, которые не фокусировались на поведении. Названные самоассоциативной памятью, они были также построены из простых «нейронов», которые были соединены друг с другом и возбуждались, когда достигали определенного порога. Но они были соединены по-другому, с использованием большого количества обратных связей. Вместо того, чтоб передавать информацию только вперед, самоассоциативная память возвращала свои выходные данные на вход – что-то вроде того, как если б вы звонили самому себе по телефону. Эти циклы на обратных связях привели к некоторым интересным свойствам. Когда паттерн активности был приложен к искусственным нейронам, они формировали память на этот паттерн. Самоассоциативные сети ассоциировали паттерны с самими собой, отсюда термин *самоассоциативная память*.

Результаты такой записи на первый взгляд кажутся нелепыми. Чтоб выбрать паттерн, хранящийся в такой памяти, вы должны задать паттерн, который хотите выбрать. Это все равно, что вы пойдете к торговцу и попросите связку бананов. Когда торговец спросит вас, как вы собираетесь расплачиваться, вы скажете, что будете платить бананами. Вы можете спросить: «Какой в этом смысл?». Но автоассоциативная память обладает несколькими важными свойствами, которые обнаружены в реальном мозге.

Наиболее важным свойством является то, что вам не обязательно иметь целиком тот паттерн, который вы хотите выбрать. У вас может быть только часть паттерна, или у вас может быть немного искаженный паттерн. Самоассоциативная память может выбрать корректный паттерн, такой, каким он был сохранен изначально, даже если вы начали с искаженной версии. Это все равно, что вы придете к торговцу с наполовину съеденными коричневыми бананами и получите целые зеленые бананы назад. Или придете в банк с порванной и нечитаемой банкнотой, а кассир вам скажет: «я думаю, это испорченная 100-долларовая купюра, дайте ее мне, а я вам дам новую хрустящую 100-долларовую купюру».

Второе свойство, отличное от других нейронных сетей, это то, что самоассоциативная память может быть спроектирована для хранения последовательностей паттернов, или временных паттернов. Это свойство достигается добавлением временных задержек к обратным связям. С такими задержками вы можете подать на автоассоциативную память последовательность паттернов, похожих на мелодию, и она может вспомнить всю последовательность. Я мог бы подать несколько первых нот «В лесу родилась елочка» и память вернула бы песню целиком. Когда подставляется часть последовательности, память может вспомнить оставшуюся часть. Как мы увидим далее, люди обучаются практически любым вещам, как последовательностям паттернов. Я предполагаю, что мозг использует контуры, аналогично тому, как это делает автоассоциативная память.

Автоассоциативные воспоминания указывают на потенциальную важность обратных связей для входных данных, изменяющихся во времени. Но подавляющее большинство ученых в области ИИ, нейронных сетей, и познания игнорируют время и обратные связи.

Ученые в области нейронаук в целом не улучшили ситуацию. Они также осведомлены об обратных связях – они были теми, кто их открыл – но у большинства из них нет теории (не дальше неопределенных разговоров о «фазах» и «модуляциях»), объясняющей, для чего они нужны мозгу в таком большом количестве. А время занимает небольшую или не центральную роль в большинстве их идей о функционировании мозга в целом. Они намерены отобразить мозг в терминах того, где происходят вещи, но не когда или как паттерны возбуждения взаимодействуют во времени. Частично это предубеждение исходит из ограниченности нашей экспериментальной аппаратуры. Одна из предпочитаемых технологий декады 90-х, известной как Декада Мозга, это функциональное отображение. Аппараты функционального отображения могут получать картины мозговой активности человека. Однако они не способны отслеживать быстрые изменения. Исследователь просит испытуемого сконцентрироваться на одной задаче снова и снова, так же как если бы фотограф просил его оставаться неподвижными, только это ментальный фотограф. Результатом является множество данных о том, *где* возникает определенные задачи в мозгу, но очень мало данных о том, как действительно потоки данных протекают через мозг. Функциональное отображение дает возможность узнать, *где* что-то происходит в заданный момент, но не может ухватить, как активность мозга меняется во времени. Ученые хотели бы получить такие данные, но практически нет аппаратуры для этого. Таким образом большинство ученых в области когнитивных нейронаук попадаться на удочку парадигмы ввода-вывода. Вы подставляете фиксированные входные данные и видите, какие получились выходные. Диаграммы связей коры головного мозга имеют тенденцию отображать блок-схему, которая начинается в первичных сенсорных областях, куда сходятся зрение, слух и осязание, затем проходят в высшие аналитические отделы, отделы планирования и в моторные области и затем передают указания мускулам. Сначала вы чувствуете, затем действуете.

Я не хочу намекать, что все игнорировали время и обратные связи. Это настолько большое поле деятельности, что виртуально у каждой идеи есть свои приверженцы. В последние годы вера в важность обратных связей, времени и предсказания последовательностей находится на подъеме. Но грохот ИИ и классических нейронных сетей заглушил другие подходы и недооцененными на многие годы.

\* \* \*

**Н**есложно понять, почему люди – дилетанты и специалисты – считали, что интеллект определяется поведением. Несколько сотен лет люди соотносили способности мозга с часовым механизмом, с насосами и трубами, затем с паровой машиной и, в последнее время, с компьютером. Десятилетия научной фантастики плавали в идеях ИИ, от трех законов робототехники Айзека Азимова до C3PO из Звездных Войн. Идея интеллектуальных машин укоренилась в нашем воображении. Все машины, сделанные людьми или воображаемые, предназначены для того, чтоб что-то делать. У нас нет машин, которые думают, у нас есть машины, которые делают. Даже когда мы наблюдаем за людьми, мы фокусируемся на их поведении, а не на их скрытых мыслях. Следовательно, интуитивно кажется очевидным, что интеллектуальное поведение должно быть метрикой интеллектуальной системы.

Однако, смотря на историю науки, мы увидим, что наша интуиция часто является громадным препятствием на пути к истине. Система научных взглядов трудна для понимания, не потому что она сложная, а потому что интуитивные но некорректные предположения удерживают нас от того, чтоб увидеть правильный ответ. Астрономы до Коперника (1472-1543) ошибочно предполагали, что земля покоится в центре вселенной, потому что она ощущается, как неподвижная и кажется, что является центром вселенной. Было интуитивно очевидным, что все звезды являются частью гигантской вращающейся сферы, а мы находимся в ее центре. Чтобы предположить, что Земля вращается, что ее поверхность движется со скоростью около тысячи миль в час, и что Земля несется сквозь пространство – не говоря уж о том, что звезды удалены на триллионы миль – вы должны представить себя в качестве лунатика. Но это обернулось корректной системой взглядов. Простой для понимания, но интуитивно некорректной.

До Дарвина (1809-1882), казалось очевидным, что все виды животных неизменны в их форме. Крокодил не скрещивается с колибри; они различны и несовместимы. Идея о эволюции видов пришла в противоречие не только с религиозными учениями, но и со здравым смыслом. Эволюция предполагает, что у нас один общий предок со всеми живыми организмами на этой планете, включая червей и домашние растения на вашей кухне. Сейчас мы знаем, что это правда, но интуиция говорит обратное.

Я упомянул эти известные примеры, потому что я уверен, что поиски интеллектуальных машин были обременены интуитивными предположениями, которые препятствуют нашему прогрессу. Когда вы спросите себя, «Чем занимается интеллектуальная система?», интуитивно очевидно размышлять в терминах поведения. Мы демонстрируем человеческий интеллект посредством речи, письменности, действий, верно? Да, но только с одной точки зрения. Интеллект – это что-то, что происходит в вашей голове. Поведение – это только дополнительный ингредиент. Это не является интуитивно очевидным, но не сложно для понимания.

\* \* \*

**В**есной 1986 года, как только я сел за мой стол после целого дня чтения научной литературы, написания истории интеллекта, и рассмотрения эволюции ИИ и нейронных сетей, я оказался за обдумыванием деталей. Был нескончаемый поток вещей, которые надо было изучить и прочитать, но я не достиг какого либо прояснения в понимании того, как действительно мозг работает в целом, или даже что он делает. Так было потому, что область нейронаук погрязла в деталях. И это до сих пор так. Каждый год публикуются тысячи исследовательских отчетов, но они тяготеют к добавлению новых данных в кучу, вместо того, чтоб организовать их. До сих пор нет общей теории, нет системы взглядов, объясняющей, что ваш мозг делает и как он это делает.

Я начал воображать, на что должно быть похоже решение этой проблемы. Должно ли оно быть крайне сложным из-за сложности мозга? Должно ли оно занять сотни страниц сплошной математики для того, чтобы описать работу мозга? Должны ли мы отобразить сотни или тысячи отдельных контуров, прежде чем мы сможем понять что-то полезное? Я так не думаю. История показывает, что лучшие решения научных проблем просты и элегантны. Тогда как детали забываются и дорога к конечной теории может быть трудной, конечная концептуальная модель в общем проста.

Без глубинного объяснения того, в каком направлении вести исследования, ученым в области нейронаук не следует больше продолжать пытаться объединить все имеющиеся детали в согласованную картину. Мозг невероятно сложный, обширный и устрашающий клубок нервных клеток. На первый взгляд он похож на стадион, полный приготовленных спагетти. Так же его можно описать, как кошмар электрика. Но при более близком и аккуратном рассмотрении мы увидим, что мозг – не беспорядочная куча. В нем много структурной организации – но слишком много для того, чтоб можно было надеяться постичь интуитивно его работу как целого, точно также мы могли бы увидеть, как осколки разбитой вазы собираются обратно вместе. Наша неудача не от недостатка данных или даже правильной части данных; все что нам нужно – это смена перспективы. С правильной моделью детали станут значимыми и управляемыми. Рассмотрим следующую причудливую аналогию, чтоб ухватить суть того, что я имею в виду.

Вообразите, что тысячелетия спустя человечество пришло к своему закату, и исследователи из удаленных внеземных цивилизаций приземляются на Землю. Они хотят понять, как мы жили. В особенности они озадачены сетью наших дорог. Для чего эти причудливые искусные структуры? Они начинают классифицировать все, и со спутников и с земли. Они дотошные археологи. Они записывают расположение каждого случайного фрагмента асфальта, каждый дорожный указатель, который опрокинулся и был унесен под откос эрозией, любую деталь, которую им удается найти. Они замечают, что некоторые дорожные сети отличаются от других; в некоторых местах они ветреные, узкие и в основном случайно ориентированные, в других - в виде красивой периодической сетки, в некоторых участках они становятся плотными и идут на сотни миль через пустыню. Они собирают горы деталей, но эти детали ничего не значат для них. Они продолжают собирать еще больше деталей в надежде найти какие-то новые данные, которые объяснят все. Они остаются в тупике надолго.

Это только до тех пор, пока они не скажет, «Эврика! Мне кажется, я вижу… эти создания не могли телепортировать сами себя, как можем мы. Они вынуждены были путешествовать с места на место, возможно на мобильных платформах хитрой конструкции». С этой точки зрения многие детали становятся на свои места. Маленькие извилистые сети улиц остались с ранних времен, когда средства передвижения были очень медленными. Широкие длинные дороги были сделаны для путешествий на длинные дистанции на высоких скоростях, предполагается, наконец, объяснение, для чего на знаках были нарисованы различные числа. Ученые начинают отличать жилые зоны от индустриальных, понимать способы, которыми могли бы взаимодействовать коммерческие и транспортные инфраструктуры, и т.д. Множество деталей, которые они каталогизировали, оказались не совсем существенными, просто из-за катастроф или требований местной географии. Остается то же самое количество сырых данных, но они больше не представляют из себя головоломку.

Мы можем быть уверены, что прорыв подобного рода позволит нам понять, к чему относятся все детали мозга.

\* \* \*

**К** несчастью, не все верят, что мы можем понять, как работает мозг. Впечатляющее количество людей, включая некоторых нейрофизиологов, верят, что так или иначе мозг и интеллект находятся за пределами объяснимого. А некоторые верят, что даже если мы сможем понять их, будет невозможно построить машину, которая будет работать подобным образом, что интеллект требует человеческого тела, нейронов и, возможно, каких-то новых и непостижимых законов физики. Когда я слышу подобные аргументы, я представляю мудрецов из прошлого, которые выступали против изучения небес или против вскрытия трупов, чтоб увидеть, как работают наши тела. «Не утруждайте себя изучением этого, это не приведет ни к чему хорошему, и даже если вы сможете понять, как это работает, мы ничего не сможем сделать с этими знаниями». Аргументы, подобные этим, ведут нас к направлению философии, называемому функционализмом, нашей последней остановке в краткой истории наших размышлений над мышлением.

Согласно функционализму, интеллект или разум безоговорочно является свойством организации, и по сути неважно, организации чего именно. Разум существует в любой системе, чьи составляющие части имеют правильные причинные отношения друг с другом, но эти части могут быть нейронами, силиконовыми чипами или чем-нибудь еще. Ясно, что эта точка зрения - стандартный выход для любого потенциального проектировщика интеллектуальных машин.

Рассмотрим ситуацию: будет ли игра в шахматы менее реальной, если при игре в нее вместо коней поставить солонки? Ясно, что нет. Солонка функционально эквивалентна «реальному» коню по тому, как она двигается по доске и взаимодействует с другими фигурами, таким образом это будет действительно игра в шахматы, а не просто их симуляция. Или, например, будет ли эта фраза той же самой, если я пройдусь по ней курсором, удаляя каждый символ, потом заново его печатая? Или возьмем пример ближе к сути, рассмотрим тот факт, что каждые несколько лет ваше тело замещает большинство составляющих его атомов. Несмотря на это, вы остаетесь самим собой во всех смыслах, которые касаются вас. Один атом ни чем не хуже другого, если он играет ту же функциональную роль в вашей молекулярной структуре. Та же самая история должна относиться и к мозгу: если какой-то сумасшедший ученый решит заместить каждый ваш нейрон функционально эквивалентным микроскопическим машинным заменителем, вы должны уйти с процедуры, чувствуя себя не менее тем же самым, каким вы были вначале.

Следуя этому принципу, искусственная система, использующая ту же самую функциональную архитектуру, что и интеллектуальный живой мозг, должна быть точно так же интеллектуальной – и не просто с какими то натяжками, а действительно, истинно интеллектуальной.

Сторонники ИИ, коннекционисты и Я – мы все функционалисты, до тех пор, пока мы верим, что по существу нет ничего специального или магического в мозгу, что делает его интеллектуальным. Все мы верим, что мы способны построить интеллектуальные машины как-нибудь и когда-нибудь. Но есть различные интерпретации функционализма. Тогда как я утверждаю, что я вижу центральную неудачу ИИ и коннекционистской парадигмы – ошибочность подхода «ввод-вывод» - есть более ценное высказывание насчет того, почему мы еще не способны разработать интеллектуальную машину. Пока сторонники ИИ принимают то, что я рассматриваю как бескомпромиссное обречение на провал, коннекционисты, с моей точки зрения, в основном всего лишь застенчивы.

ИИ-исследователи спрашивают, «Почему мы, инженеры, должны быть ограничены решениями, на которые наткнулась эволюция?». В принципе, в этом есть свой резон. Биологические системы, подобные мозгу и генетическому аппарату, печально известны своей неэлегантностью. Общей метафорой является машина Руба Голдберга, названная так после «великой депрессии» карикатуристами, которые нарисовали комически сверхсложное приспособление для выполнения тривиальных задач. У разработчиков ПО есть подобный термин, клудж, для обозначения программ, которые написаны без предусмотрительности и наполнены обременительной, ненужной сложностью, часто приводящей к тому, что программа становится непонятной даже программисту, написавшему ее. Исследователи ИИ боятся, что аналогично и мозг - беспорядочный несколько-сот-миллионолетний клудж, битком набитый неэффективным и эволюционным «наследственным кодом». Если так, удивляются они, почему бы просто не выбросить без сожаления эту кутерьму и не начать с нуля?

Большинство философов и когнитивных психологов благожелательны к этой точке зрения. Им нравится метафора разума, как программы, которая работает в мозгу, органическом аналоге компьютера. В компьютере аппаратный и программный уровни четко разделяются друг от друга. Одна и та же программа может выполняться на любой Универсальной Машине Тьюринга. Вы можете запустить WordPerfect на PC, Макинтоше, или на суперкомпьютере Cray, например, даже если все три системы имеют различные аппаратные конфигурации. И аппаратура не имеет никакого значения для вашего обучения, если вы пытаетесь изучить WordPerfect. По аналогии, следует мысль, мозг не может научить нас чему либо о разуме.

Защитники ИИ также любят указывать на исторические ситуации, в которых инженерные решения радикально отличались от природных версий. Например, как мы преуспели в построении летающих машин? Имитацией махательных движений у крылатых животных? Нет. Мы сделали это с помощью фиксированных крыльев и пропеллеров, а затем – с помощью реактивных двигателей. Это не так, как сделано в природе, но это работает – и работает гораздо лучше, чем машущие крылья.

Аналогично, мы сделали наземные транспортные средства, которые могут обогнать гепарда, не изготовлением четырехногих гепардоподобных бегающих машин, а изобретением колеса. Колеса – великолепный способ передвигаться по плоской местности, и то, что эволюция не наткнулась на эту определенную стратегию, не значит, что это блестящий путь для нас, чтоб обойти ее. Некоторые философы разума приняли расположение к метафоре «когнитивных колес», то есть, ИИ-решению некоторых проблем, которое хотя полностью отличается от того, что делает мозг, делает это действительно хорошо. Другими словами, программа, которая воспроизводит выходные данные, подобные (или превосходящие) человеческие решения задач каким то узким, но полезным способом, действительно хороший путь делать то, что делает мозг.

Я верю, что этот род интерпретаций функционализма – цель оправдывает средства – ведет ИИ исследователей в сторону. Как показал Серл в своей Китайской Комнате, эквивалентности поведения не достаточно. Поскольку интеллект внутреннее свойство мозга, мы должны смотреть внутрь мозга, чтоб понять, что такое интеллект. В наших исследованиях мозга, и в особенности неокортекса, нам необходимо быть осторожными в понимании того, какие детали всего лишь избыточная «замороженная случайность» эволюционного прошлого; несомненно, множество процессов в стиле Руба Голдберга перемешаны с важными свойствами. Но, как мы скоро увидим, там скрыта элегантность и великая мощь, опережающая наши лучшие компьютеры, ожидая, пока мы ее извлечем из нейронных цепей.

Коннекционисты интуитивно чувствовали, что мозг не просто компьютер, и что его секрет лежит в том, как ведут себя нейроны, соединенные вместе. Это было хорошим началом, но эта область лишь немного сдвинулась от своих ранних успехов. Хотя тысячи людей работали с трехслойными сетями, и многие продолжают, исследования над биологически-реалистичными нейронными сетями были и остаются редкими.

За половину столетия мы вкладывали все усилия чтобы попытаться запрограммировать интеллект в компьютерах. По пути мы пришли к текстовым процессорам, базам данных, видеоиграм, Интернету, мобильным телефонам и анимировали динозавров на компьютере. Но интеллектуальных машин до сих пор нигде не видать. Чтоб добиться успеха, мы должны будем достаточно основательно подглядеть у природного механизма интеллекта, у неокортекса. Мы должны выделить интеллект из мозга. Другой дороги у нас не будет.

# 3. Человеческий Мозг

**И**так, что делает человеческий мозг таким непохожим на программирование ИИ и нейронные сети? Что такого необычного в структуре мозга, и почему оно так значимо? Как мы увидим в следующих нескольких главах, в архитектуре мозга есть много чего, что может сказать о том, как действительно работает мозг и почему он в корне отличается от компьютера.

Давайте начнем наше знакомство с мозгом в целом. Вообразите, что мозг лежит на столе, и мы препарируем его вместе. Первое, что вы заметите, это то, что внешняя поверхность мозга кажется весьма однородной. Розовато-серый, он похож на гладкую цветную капусту с несколькими гребнями и впадинами, называемыми извилинами и бороздами. Он мягкий и желеобразный на ощупь. Это неокортекс, тонкий слой нервной ткани, который окутывает большинство более старых частей мозга. Мы собираемся сфокусировать большую часть нашего внимания на неокортексе. Практически все, о чем мы думаем, как о интеллекте – восприятие, язык, воображение, способности к математике, рисованию, музыке, планированию – происходит здесь. Ваш неокортекс читает эту книгу.

Сейчас я должен признаться, что я неокортикальный шовинист. Я знаю, что я встречу некоторое сопротивление на этой позиции, поэтому позвольте мне взять минуту на защиту моего подхода прежде, чем мы заберемся слишком далеко. У каждой части мозга есть свое сообщество ученых, изучающих ее, и предположение, что мы сможем добраться до оснований интеллекта пониманием только неокортекса, конечно же вызовет возмущение со стороны сообществ обиженных исследователей. Они скажут что-то вроде: «Вам не удастся понять неокортекс без понимания области X, потому что эти две области мозга очень сильно взаимосвязаны, и вам необходима область X для того-то и того-то». Я не говорю, что не согласен. При условии, что мозг состоит из множества частей и большинство из них критически важны для человека. (Странно, исключением является часть мозга с достаточно большим числом нейронов, мозжечок. Если вы родились без мозжечка или если он поврежден, вы можете вести почти нормальную жизнь. Однако это неверно для большинства других отделов мозга; большинство из них требуются для базовых жизненных функций или способностей к ощущениям).

Моим контраргументом является то, что я не заинтересован в построении человека. Я хочу понять интеллект и построить интеллектуальную машину. Быть человеком и быть интеллектуальным – различные вещи. Интеллектуальной машине не нужны сексуальные желания, голод, пульс, мускулы, эмоции или человекоподобное тело. Человек нечто большее, чем интеллектуальная машина. Мы биологические создания со всеми необходимым и иногда нежелательным багажом, которые происходят из многих этапов эволюции. Если вы хотите построить интеллектуальную машину, которая ведет себя как человек – то есть проходит Тест Тьюринга во всех случаях – тогда вы возможно должны будете воспроизвести множество из тех вещей, которые делают человека человеком. Но как мы увидим дальше, чтобы построить машину, которая несомненно интеллектуальна, но не является в точности человеком, мы можем сфокусироваться на той части мозга, которая строго соотносится с интеллектом.

Тем, кого возможно задевает мое исключительное внимание к неокортексу, позвольте сказать, что я согласен, что другие структуры мозга, такие как мозговой ствол, базальные ганглии, несомненно важны для функционирования человеческого неокортекса. Без вопросов. Но я надеюсь убедить вас, что все основные аспекты интеллекта возникают в неокортексе, где также важную роль играют два других отдела – таламус и гиппокамп, что мы обсудим дальше по книге. В течение продолжительного периода нам необходимо будет понять функциональную роль всех отделов мозга. Но я верю, что эти вещи было бы лучше адресовать в контексте хорошей общей теории функций неокортекса. Но давайте вернемся к неокортексу, или как его сокращенно называют, к кортексу (коре).

Возьмем шесть визиток или игральных карт и сложим их в стопку. (Это действительно поможет, если вы сделаете это, а не просто вообразите). Сейчас вы держите модель кортекса. Ваши шесть визиток примерно 2 миллиметра в толщину и должны дать вам ощущение того, как тонок кортикальный слой. Также как и стопка визиток, неокортекс примерно 2 миллиметра в толщину и содержит 6 слоев, каждый имитируется примерно одной картой.

Развернутый, неокортикальный слой человека примерно размером с обеденную салфетку. Кортикальные слои других животных меньше: у крысы – размером с почтовую марку; у обезьяны – размером с почтовый конверт. Но несмотря на размеры, большинство из них состоят из шести слоев, аналогично той стопке из шести визиток. Люди умнее потому что наш кортекс относительно размеров тела покрывает большую площадь, а не потому, что слои толще или содержат специальные «умные» нейроны. Его размеры впечатляющи, так как он окружает и обертывает большинство других частей мозга. Чтобы приспособиться к большим размерам мозга, природа модифицировала нашу общую анатомию. Человеческие женщины развиваются с широким тазом, чтоб дать возможность родиться ребенку с большой головой, свойством, которое некоторые палеоантропологи считают эволюционировавшим совместно со способностью ходить на двух ногах. Но это еще не все, также эволюция скомкала неокортекс, запихнув его в наши черепа как мятый лист бумаги в рюмке для бренди.

Ваш неокортекс заполнен нервными клетками, или нейронами. Они так плотно упакованы, что никто точно не знает, сколько же в нем клеток. Если вы нарисуете крошечный квадрат со стороной один миллиметр сверху на стопке ваших визиток, вы обозначите положение приблизительно сотни тысяч (100 000) нейронов. Вообразите попытку посчитать точное число в таком крошечном пространстве; это даже виртуально невозможно. Тем не менее, некоторые анатомы предсказывают, что в среднем человеческий неокортекс содержит порядка тридцати миллиардов нейронов (30 000 000), но никого не удивит, если в действительности окажется больше или меньше.

Эти тридцать миллиардов клеток и есть Вы. Они содержат практически все ваши воспоминания, знания, мастерство и накопленный жизненный опыт. После 25 лет размышления над мозгом, я до сих пор нахожу этот факт удивительным. То что тонкий слой клеток видит, чувствует и создает наше мировоззрение – это на грани невероятного. Тепло летних дней и наши мечты о лучшем мире каким-то образом являются созданием этих клеток. Через много лет после публикации статьи в *Scientific American,* Френсис Крик написал книгу о мозге, названную Ошеломительная Гипотеза. Ошеломительная гипотеза состояла в том, что разум – это творение клеток в мозгу. Нет ничего больше, ни магии, не специального соуса, только нейроны и танец информации. Я надеюсь вы ощутили, как невероятна такая постановка дела. Получается, что существует большая философская воронка между набором клеток и нашим сознательным опытом, хотя разум и мозг – едины. Называя это гипотезой, Крик просто соблюдал политкорректность. То, что нейроны в нашем мозгу создают разум – это факт, а не гипотеза. На необходимо понять, что эти тридцать миллиардов нейронов делают, и как они это делают. К счастью, кортекс не просто бесформенный комок ячеек. Мы можем глубже поискать в его структуре идеи о том, как он дает начало человеческому разуму.

\* \* \*

**Д**авайте вернемся к нашему столу для препарирования и еще раз взглянем на мозг. Для невооруженного глаза на неокортексе практически нет каких либо ориентиров. Точнее, есть немного, такие как гигантская борозда, разделяющая большие полушария, и заметная борозда, разделяющая передние и задние отделы мозга. Но повсюду, куда бы вы не посмотрели, извилистая поверхность практически одинакова. Нет никаких видимых разделительных линий или цветовых пометок, обозначающих области, специализирующиеся на различной сенсорной информации или различных типах мыслей.

Однако людям давно уже известно, что где-то границы все-таки есть. Даже до того, как нейрофизиологи смогли разглядеть что-нибудь полезное об устройстве кортекса, они уже знали, что некоторые ментальные функции локализованы в определенных областях мозга. Если боксерский удар поражает правую затылочную долю Джо, он может потерять, представьте себе, способность к восприятию чего-либо на левой стороне своего тела или в левой половине пространства вокруг себя. Удар в левую переднюю область, известную как область Брока, наоборот, компрометирует его способность использовать правила грамматики, хотя его словарный запас и способность понимать значение слов не изменятся. Удар в так называемую веретенообразную извилину может вырубить способность распознавать лица – Джо не может узнать свою мать, своих детей или даже свое собственное лицо на фотографии. Замечательные нарушения наподобие этих дали первым нейрофизиологам представление о том, что кортекс состоит из множества функциональных областей.

За прошедшее столетие нам многое стало известно о функциональных областях, но еще много предстоит открыть. Каждый из этих регионов частично независим и, кажется, специализирован для определенных аспектов восприятия или мышления. Физически они организованы в виде лоскутного одеяла, которое совсем немного отличается у разных людей. Они редко бывают четко очерчены. Функционально они организованы в ветвящуюся иерархию.

Упоминание об иерархии критично, так что я хочу выделить время на то, чтоб основательно определить ее. Я буду ссылаться на нее в течение всей книги. В иерархических системах некоторые элементы располагаются «выше» или «ниже» других. В иерархии бизнеса, например, менеджеры среднего уровня выше почтового клерка и ниже вице-президента. Это не имеет ничего общего с физическим положением; даже если менеджер работает этажом ниже клерка, он все равно остается «выше» по иерархии. Я акцентировал на этом внимание для того, чтобы прояснить, что я имею в виду, когда говорю о том, что некоторые функциональные области выше или ниже других. Их физическое расположение в мозгу не имеет значения. Все функциональные области кортекса располагаются на одной и той же свернутой кортикальной поверхности. Что делает один регион выше другого – это то, как они соединены друг с другом. В кортексе нижние области передают информацию к верхним областям путем определенных межнейронных соединений, тогда как верхние передают информацию нижним через другие соединения. Есть также латеральные соединения между областями, которые в различных ветвях иерархии, подобно тому, как менеджер среднего уровня общается с сотрудником из другого офиса этой же компании в другом штате. Детальная схема соединений кортекса обезьяны была разработана двумя учеными, Дэниелом Феллеманом и Дэвидом ван Эссеном. Схема показывает множество областей, соединенных вместе в сложную иерархию. Мы можем предполагать, что человеческий кортекс имеет похожую структуру.

Самые нижние функциональные области, - это первичные сенсорные области, куда первоначально приходит информация в кортекс. Эти области обрабатывают информацию в ее наиболее сырой форме, на самом базовом уровне. Например, визуальная информация попадает в кортекс через первичную визуальную область, называемую для краткости V1. V1 обрабатывает низкоуровневые визуальные элементы, такие как крошечные краевые сегменты, мелкомасштабные компоненты движения, бинокулярные несоответствия (для стереоскопического зрения) и базовую цветовую и контрастную информацию. V1 затем передает информацию верхним областям, таким как V2, V4 и IT (расскажем о ней позже) и помимо этого многим другим областям. Каждая из этих областей работает с более специализированными или абстрактными аспектами информации. Например, нейроны в V4 отвечают на объекты среднего уровня сложности, такие как форма звезды в различных цветах, например красном и синем.. Другая область, MT, специализируется на движении объектов. В верхних эшелонах визуального кортекса есть области, отвечающие за представление ваших визуальных воспоминаний всех сортов объектов типа лиц, животных, инструментов, частей тела и т.п.

Другие ваши чувства имеют похожую иерархию. В вашем кортексе есть первичная слуховая область, называемая A1, и иерархия слуховых областей над ней; в нем есть первичная соматосенсорная область, называемая S1, и иерархия соматосенсорных областей над ней. В итоге сенсорная информация попадает в «ассоциативные области», так называют иногда области кортекса, которые получают информацию более чем с одного органа чувств. Например, в вашем кортексе есть область, получающая информацию от зрения и осязания. Благодаря ассоциативным областям вы понимаете, что то, что вы видите муху, ползающую по вашей руке, и что ощущаете щекотку, вызвано одной и той же причиной. Большинство из этих областей получают сильно обработанную информацию от нескольких чувств, и их функция остается непонятной. Я еще много скажу про кортикальную иерархию позже в этой книге.

Так же есть другой набор областей в передних долях мозга, создающих двигательную (моторную) информацию. Моторная система кортекса также иерархически организована. Самая нижняя область, M1, посылает соединения в позвоночник и напрямую управляет мускулатурой. Области выше M1 посылают более сложные моторные команды в M1. Иерархия моторных областей и иерархия сенсорных областей выглядят удивительно похоже. Кажется, что они объединялись одним и тем же образом. Считается, что в моторных областях информация распространяется вниз по иерархии к M1, чтоб привести в действие мускулатуру, а в сенсорных – вверх по иерархии от органов чувств. Но в действительности информация распространяется в обоих направлениях. Что считается обратной связью в сенсорных регионах – в моторных регионах является выходной информацией, и наоборот.

Большинство описаний мозга базируется на блок-схемах, которые отображают очень упрощенную точку зрения на иерархию. То есть, входные потоки (визуальные, акустические, осязательные) попадают в первичные сенсорные области и обрабатываются по мере продвижения вверх по иерархии, затем передаются через ассоциативные области, затем попадают в передние доли кортекса и, наконец, передаются обратно вниз к моторным областям. Я не говорю, что такая точка зрения полностью ошибочна. Когда вы читаете вслух, визуальная информация несомненно приходит в V1, поднимается к ассоциативным областям, проделывает путь через фронтальный кортекс и возбуждает мускулатуру вашего рта и гортани формировать звуки речи. Но это еще не все. Не все так просто. С упрощенной точки зрения, которой я не советую придерживаться, процессы в основном рассматриваются как информация, текущая в одном направлении, как изделие на сборочном конвейере. Но информация в кортексе всегда течет также и в обратном направлении, и гораздо больше проекций, передающих вниз по иерархии, чем вверх. Когда вы читаете вслух, высшие области вашего кортекса посылают гораздо больше сигналов «вниз» к первичному визуальному кортексу, чем ваши глаза получают с напечатанной страницы! В последующих главах мы разберем, что делают эти обратные связи. Сейчас же я хочу убедить в одном факте: хотя иерархия действительно существует, необходимо всегда быть начеку, чтоб не думать, что информация течет только в одном направлении.

Вернемся к столу для препарирования, предположим, что мы настроили мощный микроскоп, взяли тонкий срез от кортикального слоя, подкрасили некоторые клетки, и взглянули на наше творчество через окуляр. Если мы подкрасили все клетки в нашем срезе, мы должны увидеть плотную черную массу, потому что клетки так плотно упакованы и перемешаны. Но если мы используем краску, которая пометит гораздо меньше клеток, мы увидим 6 слоев, о которых я упоминал. Эти слои сформированы изменением плотности клеточных тел, типов клеток и их соединений.

У всех нейронов есть общие черты. Помимо тела клетки, округлой части, которую вы представляете при упоминании клеток, у них также есть ветвящиеся, похожие на провода структуры, называемые аксонами и дендритами. Когда аксон одного нейрона соприкасается с дендритом другого, они формируют маленькое соединение, называемое синапсом. Синапс – это где нервный импульс с одной клетки воздействует на поведение другой. Нервные сигналы, или спайки, приходящие на синапс, могут сделать более вероятным появление спайка на приемной клетке. Некоторые синапсы имеют обратный эффект, делая генерацию спайка на приемной клетке менее вероятной. Таким образом, синапсы бывают возбуждающими и тормозными. Синаптическая сила может изменяться в зависимости от поведения двух клеток. Простейшая форма синаптических изменений – это когда оба нейрона генерируют спайк приблизительно в один и тот же момент, сила соединения между нейронами увеличивается. Я скажу больше про этот процесс, называемый Хеббиановским обучением, чуть позже. В дополнение к изменениям синаптической силы, очевидно, могут формироваться и новые синапсы между двумя нейронами. Это может происходить всегда, хотя научная очевидность противоречива. Независимо от деталей того, как синапсы меняют свою силу, определенно формирование и усиление синапсов – это то, что вызывает сохранение воспоминаний.

Хотя существует множество типов нейронов в неокортексе, один распространенный класс включает восемь из десяти ячеек. Это пирамидальные нейроны, называемые так, потому что из тело немного похоже на пирамиду. За исключением верхнего слоя шестислойного кортекса, который содержит мили аксонов, но очень мало клеток, каждый слой содержит пирамидальные клетки. Каждый пирамидальный нейрон соединяется с множеством других нейронов в непосредственном окружении, и каждый посылает длинный аксон вбок к более отдаленным областям кортекса или вниз, к нижележащим структурам мозга, например к таламусу.

Типичная пирамидальная клетка имеет несколько тысяч синапсов. Опять же, очень трудно узнать точно, сколько, по причине их высокой плотности и маленьких размеров. Количество синапсов изменяется от клетки к клетке, от слоя к слою и от области к области. Если б мы заняли консервативную позицию, что каждый нейрон имеет одну тысячу синапсов (действительное число синапсов оценивается ближе к пяти или десяти тысячам), то наш неокортекс должен был бы иметь примерно тридцать триллионов синапсов в сумме. Это астрономически большое число, намного за пределами наших интуитивных возможностей. Это несомненно достаточно, чтоб сохранить все вещи, которые мы когда либо узнали в течение жизни.

\* \* \*

**С**огласно преданию, Альберт Эйнштейн сказал однажды, что путь к специальной теории относительности был прямым, почти легким. Он следовал естественным путем из единственного наблюдения: что скорость света постоянна для всех наблюдателей, даже если наблюдатели движутся с различными скоростями. Это противоречит интуиции. Это все равно, что сказать, что скорость брошенного мяча всегда одна и та же, независимо от того, как сильно он был брошен, или как быстро бегут бросающий и наблюдающий. Все видят мяч, движущийся с одинаковой скоростью относительно них в любой обстановке. Кажется, что такое не может быть истиной. Но было доказано, что это истинно для света, и Эйнштейн сообразительно задался вопросом, каковы следствия этого странного факта. Он методично обдумал все следствия постоянности скорости света, и это привело его к еще более странным предсказаниям специальной теории относительности, таким, как замедление времени при увеличении скорости, что энергия и масса –фундаментально одно и то же. Книги по теории относительности повторяют его цепочку рассуждений на обыденных примерах с поездами, пулями, вспышками света и т.д. Теория не сложна, но она определенно противоречит интуиции.

В нейрофизиологии есть аналогичное открытие – факт о неокортексе, который является настолько неожиданным, что некоторые ученые отказываются верить в него и большинство оставшихся игнорируют его, потому что не знают, что с ним делать. Но это факт настолько важный, что если вы аккуратно и методически исследуете его следствия, это раскроет секрет того, что делает неокортекс и как он работает. В этом случае, неожиданное открытие пришло из базовой анатомии самого кортекса, но потребовался необычайно догадливый разум, чтоб распознать его. Это был Вернон Монткастл, нейрофизиолог из университета Джона Хопкинса в Балтиморе. В 1978 году он опубликовал статью, названную «Организационные принципы Церебральных Функций». В этом документе Монткастл указал, что неокортекс удивительно однороден по виду и структуре. Области неокортекса, которые оперируют слуховой информацией, похожи на области, оперирующие с осязанием, управлением мускулатурой, языковую область Брока, практически как любые области неокортекса. Монткастл предположил, что поскольку эти области выглядят одинаково, они действительно выполняют одну и ту же базовую операцию! Он предположил, что кортекс использует один и тот же вычислительный инструмент для всего, чем он занимается.

Все анатомы в то время и за десятилетие до МонтКастла уже знали, что кортекс выглядит одинаково везде; это непреложный факт. Но вместо того, чтоб задаться вопросом, что это могло бы означать, они тратили свое время на поиски различий между областями кортекса. Но они не находили различий. Они полагали, что если одна область используется для языка, а другая – для зрения, то вероятнее всего должны быть различия между этим областями. Если вглядеться достаточно близко – они найдутся. Области кортекса отличаются по толщине, плотности клеток, относительной пропорции типов клеток, длине горизонтальных соединений, плотности синапсов и по многим другим параметрам, которые сложно найти. В одной из наиболее изученных областей, первичной визуальной области V1, действительно есть некоторые дополнительные подразделения в одном из ее слоев. Ситуация аналогична работе биологов начала 19-го века. Они тратили свое время на поиски сиюминутных различий между видами. Успехом для них было найти, что две мыши, выглядящие практически идентично, в действительности принадлежат к различным видам. Многие годы Дарвин следовал тем же самым курсом, нередко изучая моллюсков. Но Дарвина в конце концов осенило вопросом, как такие виды могут быть настолько похожи. Их похожесть была более удивительной и интересной, гораздо больше чем их различия.

Монткастл проводил подобные наблюдения. Работая среди анатомов, отыскивающих различия между областями кортекса, он показал, что несмотря на различия, неокортекс удивительно однороден. Одни и те же слои, типы клеток и соединения существуют по всему кортексу. Он везде похож на шесть визиток. Отличия настолько тонки, что опытные анатомы не могли с ним согласиться. Следовательно, Монткастл утверждал, что все области кортекса выполняют одну и ту же операцию. То, каким образом области кортекса соединены друг с другом и с другими частями центральной нервной системы, является причиной того, что визуальные области являются визуальными, моторные – моторными.

Фактически, Монткастл утверждал, что причина того, что области кортекса выглядят слегка различными, заключается в том, к чему они подсоединяются, а не в том, что отличаются их основные функции. Он сделал вывод, что есть одна общая функция, общий алгоритм, который выполняют все области кортекса. Зрение не отличается от слуха, не отличается от моторного управления. Он предположил, что наши гены указывают, как области кортекса соединяются, что очень индивидуально для функций и видов, но кортикальная ткань сама по себе выполняет всегда одно и то же.

Давайте взглянем на это сиюминутно. Для меня зрение, слух и осязание кажутся совершенно различными. У них фундаментально разные качества. Зрение охватывает цвет, текстуру, форму, глубину. Слух работает с высотой, ритмом и тембром. Они кажутся совершенно разными. Как они могут быть одним и тем же? Монткатстл говорит – они не одинаковы, но способ, которым кортекс обрабатывает сигналы от уха такой же, как и способ, которым он обрабатывает сигналы от глаза. Он так же говорит, что моторное управление работает на тех же принципах.

Ученые и инженеры в большинстве своем игнорировали предложение Монткастла. Когда они пытались понять зрение или сделать «видящий» компьютер, они выдумывали словарь и технологии, специфические для зрения. Они говорили о краях, текстурах, трехмерном представлении. Если они хотели понять язык, они строили алгоритмы, базирующиеся на правилах грамматики, синтаксисе и семантике. Но если Монткастл прав, эти подходы не такие, как решает эти проблемы мозг, и скорее всего будут безуспешными. Если Монткастл прав, алгоритм кортекса должен быть выражен независимо от любых специфических функций или чувств. Мозг использует одни и те же процессы, чтобы видеть и слышать. Мозг занимается чем-то универсальным, что может быть применено к любому типу сенсорных или к моторной системе.

Когда я впервые прочитал статью Монткастла, я чуть не упал со стула. Это был Розеттский камень нейронауки – единственная статья и единственная идея, объединяющая все многообразные и поразительные возможности человеческого разума. Она объединяла их под единым алгоритмом. Одним махом она показала ошибочность всех предыдущих попыток понять и воплотить в технике человеческое поведение как многообразие возможностей. Я надеюсь вы сможете оценить, насколько радикальным и с чудесно элегантным является предложение Монткастла. Лучшие идеи в науке всегда просты, элегантны и неожиданны, и это одна из лучших. По моему мнению это было, есть и будет оставаться в перспективе одним из самых важных открытий нейронауки. Невероятно, однако многие ученые и инженеры либо продолжают отказываться верить в это, игнорируют это, либо вообще не знают об этом.

\* \* \*

**Ч**астично такое пренебрежение произрастает из скудности инструментов для изучения того, как информация распространяется внутри шестислойного кортекса. Имеющиеся у нас инструменты оперируют на более грубом уровне и в основном нацелены на нахождении того, где (вместо когда и как) в кортексе возникают различные способности. Например, большинство нейрофизиологов неявно пропагандируют в популярной прессе идею, что мозг – это набор высокоспециализированных модулей. Техника функционального отображения, наподобие функциональной магниторезонансной томографии (MRI) и позитрон-эмисиионной томографии (PET), сфокусирован в основном исключительно на картировании мозга и функциональных областей, о которых я упоминал ранее. Обычно в таких экспериментах подопытный лежит головой в сканере и выполняет определенный вид ментальных или моторных задач. Это может быть видеоигры, спряжение глаголов, чтение высказываний, рассматривание лиц, называние картинок, воображение чего-либо, запоминание списков, принятие финансовых решений и т.д. Сканер детектирует, какие области мозга при выполнении этих задач более активны, чем в обычном состоянии, и рисует цветные точки на изображении мозга субъекта, чтоб точно указать местоположение. Эти области, предположительно, центральные для данной задачи. Произведены тысячи экспериментов по функциональному отображению, и еще тысячи предстоит сделать. Таким образом мы постепенно строим картину того, где происходят определенные функции в типичном взрослом мозге. Легко сказать «это область, ответственная за распознавание лиц, эта – за математические способности, эта – за музыкальные» и т.д. Поскольку мы не знаем, как мозг выполняет эти задачи, естественно предположить, что мозг решает различные задачи различными способами.

Но так ли это? Растущий и зачаровывающий массив данных подтверждает предположение Монткастла. Несколько лучших примеров демонстрируют крайнюю гибкость неокортекса. Любой человеческий мозг, правильно питающийся и помещенный в правильное окружение, может изучить тысячи языков. Один и тот же мозг может выучить язык жестов, письменный язык, музыкальный язык, математический язык, компьютерные языки, язык тела. Он может научиться выживать в прохладном северном климате или в обжигающей пустыне. Он может стать экспертом по шахматам, рыбной ловле, сельскому хозяйству или теоретической физике. Рассмотрим факт, что у вас есть специальная визуальная область, которая предположительно специально предназначена представлению письменных букв и цифр. Значит ли это, что вы родились с языковой областью, готовой к обработке букв и цифр? Вряд ли. Письменный язык существует совсем недолго для того, чтоб попасть в наши гены, чтоб задействовать специальный механизм для этого. Таким образом, кортекс все равно разделяется на функциональные области, специфичные для конкретных задач в детстве, основываясь исключительно на опыте. Человеческий мозг имеет невероятную емкость для изучения и адаптации к тысячам обстановок, которые не существовали ранее. Это указывает на черезвычайную гибкость системы, а не на то, что в ней тысячи решений для тысяч задач.

Нейрофизиологи также обнаружили, что соединения в неокортексе поразительно «пластичны», что означает, что неокортекс может изменять и перекоммутировать себя в зависимости от типа информации, поступающей в него. Например, мозг новорожденного хорька можно хирургически переделать так, что глаза животного будут посылать свои сигналы в области, в норме развивающиеся для слуха. Неожиданный результат в том, что хорек развивает функционирующий визуальный канал в слуховой части своего мозга. Другими словами, он видит той мозговой тканью, которая в норме слышит звуки. Аналогичные эксперименты были проведены с другими органами чувств и областями мозга. Например, часть визуального кортекса крысы может быть пересажена при рождении в область, где обычно представлено осязание. Когда крыса вырастает, трансплантированная ткань обрабатывает осязание, а не зрение. Клетки не рождаются специализированными к зрению, осязанию или слуху.

Каждая часть человеческого неокортекса пластична. Взрослые люди, родившиеся глухими, обрабатывают визуальную информацию в тех областях, которые в норме становятся слуховыми областями. А люди с врожденной слепотой используют тыльную часть своего кортекса, которая обычно становится визуальной, для чтения азбуки Брайля. Поскольку азбука Брайля затрагивает осязание, вы могли бы подумать, что она должна в первую очередь активироваться в областях осязания, но, очевидно, ни одна область кортекса не бездействует. Визуальный кортекс, не получающий информацию от глаз, как это «предполагалось», переназначается на просеивание других паттернов – в данном случае от других кортикальных областей.

Все это приведено, чтобы показать, как области мозга развивают специализированные функции базируясь в основном на типе информации, поступающей в него во время развития. Кортекс не запрограммирован жестко для выполнения различных функций с использованием различных алгоритмов, как и земная поверхность не предопределена для того, чтоб прийти к современному расселению национальностей. Организация вашего кортекса, подобно политической географии на земном шаре, могла бы пойти по-другому, если б ранее были заданы другие условия.

Гены предопределяют общую архитектуру кортекса, включая то, какие области соединяются, но внутри этой структуры система черезвычайно гибкая.

Монткастл был прав. Есть единый мощный алгоритм, реализованный в каждой области кортекса. Если вы соедините области кортекса в подходящую иерархию и обеспечите потоками информации, он будет изучать свое окружение. Таким образом, нет причин, по которым интеллектуальные машины будущего должны иметь тот же самый набор органов чувств или возможностей, что и человек. Кортикальный алгоритм может быть использован другим путем, с другими органами чувств в машинном кортексе, так что подлинный гибкий интеллект возникнет не только в биологическом мозгу.

\* \* \*

**Д**авайте перейдем к вопросу, который посвящен предположению Монткастла и такой же неожиданный. Информационные потоки, поступающие в ваш кортекс, в основном идентичны. Вы, возможно, снова думаете, что ваши чувства полностью отдельные сущности. В конце концов, звук передается волнами давления по воздуху, зрение передается светом, осязание передается через давление на вашу кожу. Слух кажется временным, зрение кажется в основном в виде отдельных изображений, осязание – в основном пространственным. Что может быть более различным, чем звук блеющего козла и изображение яблока?

Но давайте взглянем более пристально. Визуальная информация посылается в ваш мозг через миллионы волокон в оптическом нерве. После короткой передачи через таламус она попадает в первичный визуальный кортекс. Звуки переносятся через тридцать тысяч волокон слухового нерва. Они передаются через некоторые более старые области мозга и попадают в первичную слуховую кору. Ваш спинной мозг переносит информацию о прикосновениях и внутренних ощущениях в ваш мозг через другие миллионы волокон. Они принимаются вашим первичным соматосенсорным кортексом. Это основные каналы попадания информации в ваш мозг. Они являются тем, как вы ощущаете мир.

Вы можете представлять эти каналы как связку электрических проводов или связку оптических волокон. Вы могли бы увидеть индикаторы, организованные оптическими волокнами, где цветные точки появляются на конце каждого волокна. Каналы попадания информации в мозг похожи на это, только волокна называются аксонами, и они переносят нервные сигналы, называемые «потенциалами действия» или «спайками», имеющие электрохимическую природу. Органы чувств, подающие эти сигналы, различны, но как только они превращаются в потенциалы действия, они становятся одним и тем же – просто паттернами.

Если вы взглянете, например, на собаку, множество паттернов поступит через волокна вашего оптического нерва в визуальную часть вашего кортекса. Если вы слышите собачий лай, другое множество паттернов поступит по слуховому нерву в слуховую область вашего мозга. Если вы гладите собаку, множество осязательных паттернов поступает из вашей руки по волокнам в спинном мозге в часть вашего мозга, оперирующую с осязанием.. Каждый паттерн – от взгляда на собаку, от звука ее лая, от прикосновения к собаке – расценивается по-разному, потому что приходит различными путями кортикальной иерархии. Важно где кабели приходят в мозг. Но на абстрактном уровне сенсорного ввода, они все по существу одно и то же, и все обрабатываются похожим образом в шестислойном кортексе. Слышите ли вы звук, видите ли свет, чувствуете ли давление – в мозгу нет фундаментальных различий между этими типами информации. Потенциал действия есть потенциал действия. Эти мгновенные спайки идентичны, несмотря на то, чем исходно они были вызваны. Все, что знает ваш мозг – это паттерны.

Ваше восприятие и знания о мире построены из этих паттернов. Внутри вашей головы нет света. Там темно. Внутрь головы не приходят звуки. Там тихо. Фактически, мозг – это только часть вашего тела, которая не ощущает сама по себе. Хирург мог бы ткнуть пальцем в ваш мозг и вы бы не почувствовали этого. Вся информация, попадающая в ваш разум, приходит как пространственные и временные паттерны на аксонах.

Что точно я имел в виду под пространственными и временными паттернами? Давайте взглянем на наши основные чувства по очереди. Зрение переносит и пространственную и временную информацию. Пространственные паттерны это одновременные паттерны во времени; они создаются, когда множество рецепторов одного и того же органа чувств стимулируются одновременно. В зрении чувствительным органом является сетчатка. Изображение попадает в зрачок, переворачивается хрусталиком, падает на сетчатку и создает пространственный паттерн. Этот паттерн передается в ваш мозг. Люди склонны думать, что маленькое перевернутое изображение мира поступает в ваши визуальные области кортекса, но это работает не так. Нет картинки. Это больше не картинка. Фундаментально это просто паттерн электрической активности. Его картинкоподобное качество теряется очень быстро по мере того, как ваш кортекс обрабатывает информацию, передавая компоненты паттерна вверх и вниз между различными областями, сдвигая его, фильтруя его.

Зрение так же соотносится и с временными паттернами, это обозначает, что паттерны, попадающие в ваши глаза, постоянно изменяются во времени. Но тогда как пространственный аспект зрения очевиден, его временной аспект менее очевидный. Примерно три раза в секунду ваши глаза совершают неожиданные движения, называемые саккадами. Они фиксируются на одной точке, затем неожиданно перепрыгивают к другой. Каждый раз, когда ваши глаза движутся, изображение на вашей сетчатке изменяется. Это обозначает, что паттерны, передаваемые в ваш мозг, также полностью меняются с каждой саккадой. Это только в простейшем возможном случае вы сидите и смотрите на неподвижную сцену. В реальной жизни вы постоянно вертите головой, двигаете телом и ходите в постоянно изменяющемся окружении. Ваше сознательное впечатление – это стабильный мир, полный объектов и людей, которые легко отслеживать. Но это впечатление возможно только благодаря способности вашего мозга оперировать с потоком сетчаточных изображений, паттерны которых никогда точно не повторяются. На самом деле зрение, расцененное как паттерны, попадающие в мозг, течет подобно реке. Зрение больше похоже на песню, чем на рисунок.

Большинство исследователей зрения игнорируют саккады и быстрые изменения визуальных паттернов. Работая с обездвиженными животными, они изучают, как возникает зрение, когда бессознательное животное фиксирует взгляд в точке. Поступая так, они уходят от временного измерения. Нет ничего ошибочного в этом принципе; исключение переменных – базовый элемент научной методологии. Но они выбрасывают центральный компонент зрения, из которого оно состоит. Время должно занимать центральное место в нейрофизиологическом учете зрения.

Говоря о слухе, мы думаем о временном аспекте звука. Для нас интуитивно очевидно, что звуки, речь и музыка изменяются во времени. Вы не можете услышать всю песню сразу же, также как не можете мгновенно услышать произнесенное предложение. Песня существует только во времени. Таким образом, мы обычно не думаем о звуке, как о пространственном паттерне. В некотором смысле это инверсия случая со зрением: временной аспект сразу же очевиден, но пространственный аспект менее очевиден.

Слух также имеет пространственный компонент. Вы конвертируете звуки в потенциалы действия в свернутом кольцами органе в ухе, называемом улиткой. Крошечный, непрозрачный, свернутый в спираль и встроенный в самую твердую кость тела височную кость, улитка была дешифрована более чем полстолетия назад Венгерским физиологом Джорджем фон Бекси. Строя модель внутреннего уха, фон Бекси открыл, что каждая компонента слышимого вами звука заставляет вибрировать участок улитки. Высокочастотные тоны заставляют вибрировать жесткое основание улитки. Низкочастотные тоны заставляют вибрировать свободную и самую внешнюю часть улитки. Средние частоты вызывают вибрацию в промежуточных сегментах. Каждая площадка улитки усеяна нейронами, которые возбуждаются, когда их встряхивает. В повседневной жизни ваши улитки вибрируют на большом количестве частот одновременно. Таким образом в каждый момент времени имеется новый пространственный паттерн возбуждений по длине каждой улитки. Каждый момент времени новый пространственный паттерн пробегает по слуховому нерву. Снова мы видим, как эта сенсорная информация сводится к пространственно-временным паттернам.

Люди обычно не думают об осязании, как о пространственном феномене, но это во всех отношениях как временной, так и пространственный паттерн. Вы можете провести эксперимент, чтоб самим убедиться. Попросите друга сложить руку в виде чашки, ладонью вверх, и закрыть глаза. Поместите маленький обычный предмет в его ладонь – кольцо, ластик, или еще что-либо – и попросите его определить что это, не двигая частями руки. У него не будет другой подсказки кроме как вес или, возможно, большой размер. Затем попросите его продолжать держать глаза закрытыми, и двигать пальцами по объекту. Скорей всего, он сразу же определит. Позволив пальцам двигаться, вы добавили время к сенсорному восприятию осязания. Есть прямая аналогия между фовеальной областью в центре вашей сетчатки и вашими пальцами, и то и другое обладают высоким разрешением. Таким образом, осязание также подобно песне. Ваша способность использовать осязание сложным образом, такое как застегивание рубашки или отпирание двери в темноте, зависит от изменяющихся во времени паттернов чувства осязания.

Мы учим наших детей, что у человека пять чувств: зрение, слух, осязание, обоняние и вкус. В действительности у нас их больше. Зрение больше похоже на три чувства: чувство движения, ощущение цвета и ощущение освещенности. В осязании присутствуют давление, температура, боль и колебания. У нас также есть целая система сенсоров, говорящих нам о положении нашего тела и о углах сгиба суставов. Она называется проприоцептивной системой (*proprio*- имеет тот же латинский корень, что и *proprietary* – «собственность» и *property* – «свойство»). Вы не сможете двигаться без них. У нас также есть вестибулярный аппарат во внутреннем ухе, который дает нам ощущение равновесия. Некоторые из этих чувств богаче и более очевидны для нас, чем другие, но все они поступают в наш мозг как потоки пространственных паттернов, распределенные во времени по аксонам.

Ваш кортекс в действительности не знает и не чувствует мир напрямую. Единственная вещь, которую знает кортекс – это потоки паттернов на входных аксонах. Ваше восприятие мира создано из этих паттернов, включая ваше ощущение самого себя. Фактически, ваш мозг не может напрямую узнать, где кончается ваше тело и начинается мир. Нейрофизиологи, изучающие отображение тела, обнаружили, что наше ощущение самих себя более гибкое, чем нам кажется. Например, если я дам вам маленькие грабли и заставлю использовать их для того, чтоб брать и хватать вместо того, чтоб делать это руками, скоро вы почувствуете, что они стали частью вашего тела. Ваш мозг изменит его ожидания, приспосабливаясь к новым паттернам тактильной информации. Грабли станут буквально встроенными в образ вашего тела.

\* \* \*

**И**дея, что паттерны от различных чувств эквивалентны внутри вашего мозга, является совершенно неожиданной, и хотя она вполне понятна, она до сих пор не оценена по достоинству. Вот еще несколько примеров. Первый вы можете воспроизвести в домашних условиях. Все что вам необходимо – это напарник, устойчивая картонная ширма и муляж руки. Если вы производите этот эксперимент в первый раз, было бы идеальным, если бы у вас была резиновая рука, наподобие тех, что продаются на распродажах перед Хеллоуином, но также сработает, если вы просто обведете вашу руку на листе чистой бумаги. Положите вашу настоящую руку на поверхность стола в нескольких дюймах от фальшивой и выровняйте их одинаково (чтобы кончики пальцев были в одном направлении, ладони либо обе вверх, либо обе вниз). Затем поставьте ширму между двумя руками, так чтобы вы видели только фальшивую руку. Пока вы пристально глядите на фальшивую руку, работа вашего напарника – одновременно постукивать по обеим рукам в соответствующих точках. Например, ваш напарник мог бы постукивать по обоим мизинцам от сустава к ногтю с одной и той же скоростью, затем сделать три быстрых постукивания по второму суставу обоих указательных пальцев с одинаковыми промежутками, затем постукивать по окружности по тыльной стороне каждой руки и т.п. Через некоторое время области вашего мозга, где сходятся визуальные и соматосенсорные паттерны – одна из тех ассоциативных областей, о которых я упоминал ранее в этой главе - станут сбиты с толку. Вы действительно будете чувствовать прикосновения к фальшивой руке, как будто она ваша собственная.

Другой замечательный пример «эквивалентности паттернов» называется сенсорной подстановкой. Это может совершить революцию в жизни людей, потерявших зрение в детстве, и может когда-нибудь стать благом для людей, родившихся слепыми. Это также может породить новую технологию человеко-машинного интерфейса для остальных.

Понимая, что для мозга паттерны это все, Пол Бах-и-Рита, профессор биомедицинской инженерии из Университета Висконсина, разработал метод для отображения визуальных паттернов на поверхности языка. Надевая это отображающее устройство, слепые люди обучались «видеть» через ощущения на поверхности своего языка.

Вот как это работает. Человек надевает маленькую камеру на голову и чип на язык. Визуальные картинки транслируют пиксель за пикселем в точечные нажимы на языке. Визуальная сцена, которая может быть изображена как сотни пикселей на телевизионном экране, может быть превращена в паттерн из сотен точечных нажимов на поверхность языка. Мозг быстро обучается правильно интерпретировать эти паттерны.

Одним из первых, кто испробовал это устройство, является Эрик Вейхенмайер, атлет мирового класса, который ослеп в возрасте тринадцати лет, и который читает лекции о том, что слепые люди не должны сдаваться. В 2002 году Вейхенмайер взобрался на гору Эверест, став первым слепым человеком, не только достигнувшим, но и впервые предпринявшим такую цель.

В 2003 году Вейхенмайер испробовал наязычное приспособление и увидел изображение впервые со своего детства. Он смог разглядеть катящийся к нему по полу мяч, достать напиток со стола и сыграть в игру «Камень, Ножницы, Бумага». Позже он прогуливался по коридору, видел открывающиеся двери, изучил дверь и окно, и заметил, что на них есть значки. Изображения, первоначально воспринимаемые как прикосновения к языку, вскоре стали восприниматься как изображения в пространстве.

Эти примеры еще раз показывают, что кортекс черезвычайно гибкий, и что информация, поступающая в мозг, всего лишь паттерны. Не важно, откуда пришли паттерны; пока они коррелируют во времени определенным образом, мозг может воспринимать их как ощущения.

\* \* \*

**В**се это не должно быть неожиданным, если мы примем точку зрения, что все, о чем знает мозг. Мозг это машина для паттернов. Я не говорю, что неправильно выражать функции мозга в терминах слуха, зрения, но на более фундаментальном уровне суть дела – паттерны. Не важно, насколько различной кажется активность разных областей кортекса, в них работает один и тот же базовый кортикальный алгоритм. Кортексу не важно, исходят ли паттерны от зрения, слуха или других чувств. Ему не важно, приходит информация от одного органа чувств или от четырех. Также ему было бы не важно, если бы вы воспринимали мир с помощью сонара, радара или магнитных полей, или если б у вас были щупальца вместо рук, или даже если б вы жили в четырех измерениях, а не в трех.

Это означает, что вам не нужен какой либо из ваших органов чувств или определенная комбинация чувств, чтоб быть интеллектуальным. У Хелен Келлер не было ни зрения, ни слуха, хотя она изучила язык и стала более опытным писателем, чем большинство видящих и слышащих людей. Это был очень интеллектуальный человек без двух основных чувств, невероятная гибкость мозга позволила ей воспринимать и понимать мир как и людям с пятью органами чувств.

Этот вид замечательной гибкости человеческого мозга дает мне надежду, что мы воссоздадим технологии, подсмотренные у мозга. Когда я думаю о построении интеллектуальных машин, я удивляюсь, зачем их привязывать к обычным органам чувств? Как только мы сможем расшифровать неокортикальный алгоритм и создать науку о паттернах, мы сможем применить его к любой системе, которую мы захотим сделать интеллектуальной. И одним из замечательных свойств схемы, подсмотренной в неокортексе, является то, что нам не нужно будет специально хитро программировать ее. Также, как слуховой кортекс стал визуальным кортексом у «перекоммутированого хорька», также, как визуальный кортекс находит альтернативное применение у слепых людей, точно также система, работающая на неокортикальном алгоритме, будет интеллектуальной независимо от того, какие типы паттернов мы выберем для нее. Нам все еще надо будет поломать голову над настройкой различных параметров системы, и нам необходимо будет натренировать и обучить ее. Но миллиарды нейронных деталей, дающих мозгу возможность быть сложным, создающим мысли, позаботятся сами о себе таким же естественным образом, как это происходит у наших детей.

В конечном счете, идея, что паттерны это фундаментальная валюта интеллекта, ведет к одному интересному философскому вопросу. Когда я сижу в комнате с моими друзьями, откуда я знаю, что они там или даже что они реальны? Мой мозг получает паттерны, которые согласуются с паттернами, которые я получил в прошлом. Эти паттерны соответствуют людям, которых я знаю, их лицам, их голосам, их обычному поведению, и всем видам фактов о них. Я научился ожидать, что эти паттерны возникают вместе определенным образом. Но когда вы переходите к ним, это всего лишь модель. Все наши знания о мире – это модель, основанная на паттернах. Уверены ли мы, что мир реален? Весело и необычно размышлять об этом. Некоторые фантастические книги и фильмы исследуют эту тему. Это не для того, чтоб сказать, что люди или объекты не реальны. Они реальны. Но наша уверенность в существовании мира базируется на соответствии паттернов и том, как мы их интерпретируем. Нет такой вещи, как непосредственное восприятие. У нас нет сенсоров «человека». Вспомните, мозг находится в темной и тихой коробке, и не знает ни о чем, кроме распределенных во времени паттернов, поступающих по входным волокнам. Ваше восприятие мира создано из этих паттернов, больше ничего. Существование может быть объективным, но пространственно-временные паттерны, текущие по связкам аксонов в наших мозгах, это все, с чем мы должны работать.

Эта дискуссия освещает иногда поднимаемые вопросы отношений между галлюцинациями и реальностью. Если мы можем воспринимать галлюцинации прикосновений, исходящие от резиновой руки и мы можем «видеть» симуляцией прикосновений к поверхности языка, «обманываемся» ли мы точно также, когда ощущаем прикосновения к собственной руке или видим своими собственными глазами? Можем ли мы верить, что мир таков, как кажется? Да. Мир действительно существует в абсолютной форме, близкой к той, в кокой мы его воспринимаем. Однако, наш мозг не может знать напрямую об абсолютном мире.

Мозг знает о мире через множество чувств, которые детектируют только часть абсолютного мира. Чувства создают паттерны, которые посылаются в кортекс и обрабатываются одними и теми же кортикальными алгоритмами для создания модели мира. Таким способом устная и письменная речь воспринимаются удивительно похоже, несмотря на то, что они совершенно различны на сенсорном уровне. Так же, модель мира Хелен Келлер близка к вашей и моей, несмотря на то, что у нее сильно сокращенный набор чувств. Через эти паттерны мозг создает модель мира, которая близка к реальным вещам, и затем, удивительно, держит ее в памяти. Память – вот что происходит с паттернами после того, как они попадают в кортекс – что мы и обсудим в следующей главе.

# 4. Память

**Когда** вы читаете эту книгу, идете по переполненной улице, слушаете симфонию, успокаиваете плачущего ребенка, ваш мозг набит пространственными и временными паттернами от всех ваших органов чувств. Мир это океан постоянно меняющихся паттернов, которые приходят, захватывая и врезаясь в ваш мозг. Почему вы ощущаете этот натиск? Паттерны приходят, проходят через различные структуры старого мозга, и, в конечном счете, попадают в неокортекс. Но что происходит с ними, когда они попадают в кортекс?

С времен начала индустриальной революции люди рассматривали мозг как некоторый тип машины. Они знали, что в голове нет шестерней и зубьев, но это было лучшей метафорой, которая у них была. Каким-то образом информация приходит в мозг и мозг-машина определяет, как должно реагировать тело. В компьютерную эру мозг стал рассматриваться как особый тип машины, программируемый компьютер. И как мы увидели в главе 1, исследователи ИИ уткнулись в эту точку зрения, аргументируя отсутствие прогресса тем, что компьютеры слишком слабы и медленны по сравнению с человеческим мозгом. Они говорят, что современный компьютер может быть эквивалентен только мозгу таракана, но когда мы сделаем компьютеры мощнее и быстрее, они станут такими же интеллектуальными, как люди.

С этой аналогией мозг - компьютер существует повсеместно игнорируемая проблема. Нейроны гораздо медленнее транзисторов в компьютере. Нейрон собирает информацию со своих синапсов и комбинирует эту информацию, чтоб решить, когда сгенерировать спайк для других нейронов. Обычный нейрон может сделать это и сбросить себя миллисекунд на пять, то есть примерно 200 раз в секунду. Это может показаться быстрым, но современные кремниевые компьютеры могут выполнять миллиард операций в секунду. Это означает, что базовая компьютерная операция в пять миллионов раз быстрее базовой операции вашего мозга. Это очень, очень большая разница. Так как же может быть, что мозг быстрее и мощнее, чем самые быстродействующие современные компьютеры? «Без проблем», говорят люди, поддерживающие аналогию «мозг это компьютер». «Мозг это параллельный компьютер. В нем миллиарды клеток, работающих одновременно. Этот параллелизм значительно увеличивает вычислительную мощь биологического мозга».

Я всегда чувствовал, что такой аргумент был хитростью, и простой мысленный эксперимент показывает почему. Это называется «правило в сто шагов». Человек может выполнять значительные задачи меньше чем за секунду. Например, я мог бы показать вам фотографию и попросить определить, изображена ли на ней кошка. Вашей задачей было бы нажать на кнопку, если там кошка, но не медведь или бородавочник или репа. Эта задача для компьютера сложная или невозможная на сегодняшний день, тогда как человек может решить ее достоверно за полсекунды или меньше. Но нейроны медленны, так что за полсекунды информация, поступающая в мозг, может пройтись только по цепочке длиной в сто нейронов. То есть, «компьютерное» решение подобной проблемы мозгом может быть в сто шагов или меньше, несмотря на то, сколько всего нейронов задействовано. С момента, когда свет попал в ваш глаз, до момента нажатия кнопки может быть задействована цепочка не длиннее ста нейронов. Цифровой компьютер, пытающийся решить ту же самую задачу, сделал бы миллиарды операций или шагов. Одной сотни компьютерных команд хватит только на то, чтоб переместить единичный символ на дисплей, не говоря о том, чтоб сделать что-то интересное.

Но если у меня есть миллионы нейронов, работающих совместно, не похоже ли это на параллельный компьютер? Конечно нет. И мозг и параллельный компьютер оперируют параллельно, но это все, что между ними общего. Параллельные компьютеры комбинируют множество скоростных компьютеров для работы над большой задачей, типа прогноза погоды на завтра. Чтоб предсказать погоду, вы должны вычислить физические условия во множестве точек планеты. Каждый компьютер может работать над отдельным местом в одно и то же время. Но даже если сотни или тысячи компьютеров будут работать параллельно, единичный компьютер все равно выполнит миллиарды или триллионы операций, прежде чем завершит задачу. Самый большой мыслимый параллельный компьютер не может сделать ничего полезного за сто шагов, не важно, насколько он большой или быстрый.

Вот аналогия. Предположим, я попрошу вас перенести одну сотню блоков через пустыню. Вы можете переносить только один камень одновременно, и пересечение пустыни потребует миллион шагов. Вы понимаете, что это займет у вас много времени, если действовать в одиночку, поэтому вы нанимаете сотню работников, которые будут работать параллельно. Теперь задача решается в сто раз быстрее, но она все также требует как минимум миллион шагов на пересечение пустыни. Наем еще большего количества рабочих – даже тысячи – не даст никакого выигрыша. Не важно, сколько рабочих вы наняли, задача не может быть решена за меньшее время, чем потребуется на миллион шагов. То же самое верно и для параллельных компьютеров. С некоторого момента, добавление новых компьютеров ничего не изменит. Компьютер, не важно, сколько в нем процессоров и как быстро они работают, не может «вычислить» ответ на сложную задачу за сотню операций.

Так как же мозг решает сложную задачу за сто шагов, которую параллельный компьютер даже теоретически не может решить за миллион или миллиард операций? Ответ в том, что мозг не «вычисляет» ответ на задачу; он достает ответ из памяти. По существу ответ был сохранен в памяти заранее. Всего несколько шагов требуется, чтоб достать что-то из памяти. Медленные нейроны не только достаточно быстры, чтоб сделать это, но они сами составляют эту память. Весь кортекс – это система памяти. Это совсем не компьютер.

\* \* \*

**П**озвольте показать на примере различие между *вычислением* ответа на задачу и *использование памяти* для решения той же самой задачи. Рассмотрим задачу ловли мяча. Кто-то бросает мяч вам, вы видите, как он движется к вам, и менее чем за секунду вы хватаете его. Это кажется несложным – до тех пор, пока вы не попытаетесь запрограммировать манипулятор робота, чтоб сделать то же самое. Как убедились на своем опыте множество аспирантов, это кажется практически невозможным. Когда инженер или компьютерщик энергично берется за эту задачу, он в первую очередь пытается вычислить полет мяча, чтоб определить, где он будет, когда достигнет манипулятора. Это вычисление требует решения набора уравнений того типа, что изучались вами на физике в институте. Затем, все шарниры манипулятора должны дружно передвинуть манипулятор в необходимое положение. Это требует решение другого набора математических уравнений, более сложного, чем первые. Наконец, эта операция в целом должна быть повторена множество раз, чтобы по мере приближения мяча робот получил наилучшую информацию о положении и траектории мяча. Если робот будет ждать вычисления точного положения прибытия мяча, прежде чем начнет движение, он не успеет поймать его. Он должен начать движение, как только получит малейшую информацию о положении мяча, и постоянно корректировать свое положение по мере приближения мяча. Компьютеру требуются миллионы операций, чтоб решить множество математических уравнений для поимки мяча. И хотя компьютер мог бы быть запрограммирован для решения этой задачи, «правило ста шагов» говорит нам, что мозг решает ее другим способом. Он использует память.

Каким образом вы ловите мяч, используя память? Ваш мозг хранит информацию о мышечных командах, необходимых для поимки мяча (вместе с другими заученными движениями). Когда мяч брошен, происходят три веши. Во-первых, соответствующие воспоминания автоматически вызываются образом мяча. Во-вторых, фактически вспоминается временная последовательность мышечных команд. И в-третьих, полученная информация корректируется по мере ее вспоминания для того, чтоб приспособить к определенному моменту, такому как фактическая траектория мяча и положение вашего тела. Память о том, как поймать мяч, не запрограммирована в вашем мозгу; она запоминается за годы постоянной практики, и сохраняется без вычислений в ваших нейронах.

Вы могли бы подумать, «подождите, каждая попытка поймать мяч слегка отличается. Вы только что сказали, что каждое воспоминание постоянно корректируется, чтоб приспособить к различным вариациям мяча в каждом конкретном броске… Разве это не требует решения тех же самых уравнений, которых мы попытались избежать?». Так может показаться, но природа решила задачу вариации другим, очень простым путем. Как мы увидим позже в этой главе, кортекс создает то, что называется *инвариантный образ*, который автоматически оперирует с вариациями в мире. В качестве полезной аналогии можно вообразить, что происходит, когда вы садитесь на водяную кровать: подушки и другие люди на кровати внезапно смещаются в новое положение. Кровать не рассчитывает, как высоко должен быть поднят каждый объект; физические свойства воды и пластиковой оболочки матраца автоматически заботятся о корректировке. Как мы увидим в следующей главе, архитектура шестислойного кортекса, мягко говоря, делает нечто подобное с информацией, проходящей через него.

\* \* \*

**Т**аким образом, неокортекс не похож на компьютер, параллельный или какой либо другой. Вместо вычисления ответов на задачи для их решения и формирования поведения неокортекс использует сохраненную информацию. У компьютеров также есть память в виде жесткого диска или чипов памяти; однако у неокортикальной памяти есть четыре атрибута, фундаментально отличающиеся от компьютерной памяти:

• **Неокортекс хранит последовательности паттернов.**

• **Неокортекс вспоминает паттерны автоассоциативно.**

• **Неокортекс хранит паттерны в инвариантной форме.**

• **Неокортекс хранит паттерны иерархически.**

Мы обсудим первые три различия в этой главе. Я введу концепцию иерархии неокортекса в главе 3. В главе 6 я опишу ее значимость и как она работает.

Когда в следующий раз вы будете рассказывать историю, остановитесь и подумайте, почему в один момент времени вы можете рассказывать только об одном аспекте истории. Вы не можете рассказать одновременно все, что происходило, не важно, как быстро вы говорите или я слушаю. Вам необходимо закончить одну часть истории прежде, чем вы сможете приступить к следующей. Это не только потому что разговорный язык последовательный; и письменный, и устный, и визуальный рассказ истории происходят последовательным образом. Все это потому, что рассказ хранится в вашей голове в последовательной форме и может быть вспомнен только в той же самой последовательности. Вы не можете вспомнить всю историю одновременно. Фактически в большинстве случаев невозможно думать о какой-либо сложной вещи не как о последовательности событий или мыслей.

Вы можете отметить также, что рассказывая историю некоторые люде не могут сразу же дойти до сути. Кажется, что они распространяются о совершенно несвязанных деталях. Это может раздражать. Вы можете крикнуть «Ближе к делу!». Но они описывают историю так, как будто она происходила с ними, во времени, и не могут рассказать ее по-другому.

Другой пример: я хотел бы, чтоб вы представили свой дом прямо сейчас. Закройте глаза и представьте его. В вашем воображении пройдите к входной двери. Вспомните, как она выглядит. Откройте входную дверь. Зайдите внутрь. Теперь посмотрите налево. Что вы видите? Посмотрите направо. Что там? Пройдите в ванную. Что справа? Что слева? Что в верхнем правом ящике? Какие вещи у вас в душе? Вы знаете все эти вещи плюс тысячи других и можете вспомнить их достаточно детально. Эти воспоминания хранятся в вашем кортексе. Вы могли бы сказать, что все эти вещи часть воспоминания о вашем доме. Но вы не сможете думать о них о всех одновременно. Очевидно, что это связанные воспоминания, но нет способа вспомнить все эти детали одновременно. У вас исчерпываюшие воспоминания о вашем доме; но чтоб вспомнить их, вы должны пройти через последовательность сегментов, в основном тем же самым путем, каким они известны вам из личного опыта.

Все воспоминания похожи на это. Вы должны пройти через временную последовательность так, как вы сделали бы в действительности. Один паттерн (подход к двери) затрагивает другой паттерн (прохождение через дверь), он затрагивает следующий (спуститься в холл или подняться по ступенькам), и так далее. Каждый – это последовательность, которую вы проходили раньше. Конечно, сознательным усилием я могу изменить порядок того, как я описываю вам мой дом. Я могу перепрыгнуть от цокольного ко второму этажу, если захочу сфокусироваться на вещах непоследовательно. Однако же, как только я начну описывать какую-нибудь выбранную комнату или вещь, я снова вернусь к последовательности. Действительно произвольного мышления не существует. Воспоминания почти всегда следуют путем ассоциаций.

Вы знаете алфавит. Попробуйте произнести его задом наперед. Вы не сможете, потому что вы обычно не слышали его в обратной последовательности. Если вы хотите узнать, что чувствуют дети, изучающие алфавит, попробуйте произнести его в обратной последовательности. Это тоже самое, с чем сталкиваются они. Это действительно трудно. Ваше знание алфавита – это последовательность паттернов. Оно не может быть запомнено или вспомнено непосредственно или в произвольном порядке. То же самое с днями недели, месяцами года, вашим телефонным номером и бесчисленным множеством других вещей.

Ваша знание песни – великолепный пример временной последовательности в памяти. Вспомните какую-нибудь мелодию. Я люблю приводить пример с «Somewhere over the Rainbow», но подойдет любая мелодия. Вы не сможете представить сразу всю песню целиком, только в последовательности. Вы можете начать с начала или, возможно, с припева, но затем вы будете воспроизводить нота за нотой. Вы не сможете вспомнить песню задом наперед, точно также, как и вспомнить всю мгновенно. Сначала вы услышали песню так, как она звучит во времени, и вы можете ее вспомнить только точно так же, как ее заучили.

Это применимо также и к сенсорной памяти низкого уровня. Вспомним ваши тактильные ощущения текстуры. Ваш кортекс хранит память о том, на что похоже ощущение, когда вы держите пригоршню гравия, проводите пальцами по бархату, нажимаете на клавиши фортепиано. Эта память точно так же основана на последовательностях, как алфавит или песни; только последовательности более короткие, охватывающие всего лишь доли секунд, а не несколько секунд или минут. Если я зарою вашу руку в ведре гравия, пока вы спите, когда вы проснетесь, вы не поймете, с чем соприкасается ваша рука, пока не пошевелите пальцами. Ваша память тактильной текстуры гравия основана на последовательностях паттернов ощущения нейронами вашей кожи давления и вибрации. Эти последовательности отличаются от тех, которые вы получили бы, если б ваша рука была погружена в песке или в пенопластовые шарики или в сухие листья. Когда вы сгибаете вашу руку, царапанье и перекатывание камешков должно создавать сигнальные последовательности паттернов гравия и вызывать соответствующие воспоминания в вашем соматосенсорном кортексе.

Когда вы в следующий раз будете выходить из душа, уделите внимание тому, как вы вытираетесь полотенцем. Я обнаружил, что каждый раз вытираюсь практически в одной и той же последовательности растираний, похлопываний и положений тела. С помощью забавных экспериментов я обнаружил, что моя жена также следует неизменным паттернам, когда выходит из душа. Возможно, с вами происходит то же самое. Если вы следуете последовательности, попробуйте изменить ее. Вы можете заставить себя сделать это, но вам необходимо оставаться сосредоточенным. Если вы отвлечетесь, вы снова вернетесь к заученным паттернам.

Вся память хранится в синаптических соединениях между нейронами. Фактически мы храним в кортексе очень большое количество вещей, но то, что в один момент времени мы можем вспомнить только крошечную долю этих сохраненных воспоминаний, вызвано тем, что только ограниченное количество синапсов и нейронов в вашей голове играют активную роль в вспоминании в один момент времени. Когда вы начинаете вспоминать, что есть у вас дома, одно множество нейронов становится активным, что ведет затем к активизации другого множества нейронов, и т. д. У неокортекса взрослого человека невероятно большая емкость памяти. Но несмотря на то, что мы помним так много вещей, мы можем вспомнить в один момент времени только небольшую часть, и можем сделать это только в последовательности ассоциаций.

Вот интересное упражнение. Попробуйте вспомнить детали вашего прошлого, где вы жили, какие места посещали, людей, которых вы знали. Я обнаружил, что я всегда могу вспомнить о вещах, о которых не думал в течение множества лет. Есть тысячи деталей воспоминаний, сохраненных в синапсах вашего мозга, которые редко используются. В любой момент времени мы вспоминаем только крошечную долю того, что мы знаем. Большая часть информации ожидает соответствующих стимулов, чтоб быть востребованной.

Компьютерная память не хранит последовательности паттернов. Это может быть сделано с помощью различных программных средств, но компьютерная память не делает этого автоматически. В отличие от этого, кортекс автоматически хранит последовательности. Это является неотъемлемым аспектом системы неокортикальной памяти.

\* \* \*

**Т**еперь давайте рассмотрим второе ключевое свойство нашей памяти, ее автоассоциативную природу. Как мы увидели в главе 2, этот термин просто обозначает, что паттерны ассоциируются с самими собой. Автоассоциативные системы памяти – это то, что может вспомнить паттерн полностью, когда заданы только частичные или искаженные данные. Это работает и для пространственных и для временных паттернов. Если вы видите, что из-за занавеса торчат ботинки вашего ребенка, вы автоматически воображаете его целиком. Вы восполняете пространственный паттерн по его частичной версии. Или вообразите, что вы видите человека, ждущего автобус, но можете наблюдать только его часть, потому что он частично загорожен кустом. Ваш мозг не сбивается с толку. Ваши глаза видят только часть тела, но ваш мозг дополняет оставшееся, создавая восприятие человека целиком настолько явно, что вы можете не заметить, что вы это только подразумеваете.

Вы также восполняете и временные паттерны. Если вы вспоминаете небольшую деталь того, что произошло очень давно, полная последовательность событий может заполнить ваш разум. Известная серия новелл Марселя Пруста, *Remembrance of Things Past,* начинается с воспоминания о том, как пахнет печенье – и отталкиваясь от этого, он выдает больше тысячи страниц. При общении мы часто не слышим некоторых слов, если мы находимся в шумном помещении. Никах проблем. Наш мозг заменяет пропущенное тем, что он ожидает услышать. Точно установлено, что мы действительно слышим не все слова из тех, которые воспринимаем. Некоторые люди вслух завершают предложения других, но мысленно каждый из нас делает это постоянно. И не только конец предложения, но и середину и даже начало. В большинстве случаев мы не осознаем, что постоянно завершаем паттерны, но это повсеместное и фундаментальное свойство того, как память хранится в кортексе. В любой момент часть может активизировать целое. В этом суть автоассоциативной памяти.

Ваш неокортекс это сложная биологическая автоассоциативная память. В любой момент прогулки каждый функциональный регион по существу неустанно ожидает, что поступят знакомые паттерны или фрагменты паттернов. Вы можете сильно задуматься о чем-либо, но как только появится ваша подруга, ваши мысли переключатся на нее. Это переключение не является чем-то, что вы выбираете сознательно. Появление вашей подруги всего лишь заставляет ваш мозг начать вспоминать паттерны, ассоциирующиеся с ней. Это неизбежно. После того, как вас прервут, вы часто спрашиваете, «О чем я только что думал?». Общение с друзьями за ужином следует путями ассоциаций. Разговор может начаться с еды перед вами, но салат вызовет ассоциированное воспоминание о салате на вашей свадьбе, что приведет к воспоминанию о еще чьей-нибудь свадьбе, что приведет к воспоминанию о том, где они провели медовый месяц, к политическим проблемам в прошлом и т.д. Мысли и воспоминания ассоциативно связаны, произвольно мысли никогда не возникают. Информация, поступающая в мозг, автоассоциативно связана сама с собой, заполняя настоящее и автоассоциативно связывая с тем, что должно последовать. Мы называем эти цепочки воспоминаний *мыслями*, и хотя их путь не предопределен, мы не можем полностью их контролировать.

\* \* \*

**С**ейчас мы можем рассмотреть третье важное свойство неокортикальной памяти: как она формирует то, что мы называем инвариантным представлением. В этой главе я поясню основные идеи инвариантного представления, а в главе 6 – детали того, как кортекс создает его.

Компьютерная память разработана для того, чтоб хранить информацию в точно таком же виде, как она представляется. Если вы копируете программу с компакт-диска на жесткий диск, каждый байт копируется со стопроцентной точностью. Единичная ошибка или отличие между двумя копиями может вызвать сбой программы. Память неокортекса отличается от этого. Наш мозг не помнит в точности, что он видит, слышит или чувствует. Мы не помним или не вспоминаем вещи со точно – не потому что кортекс и его нейроны небрежные или подвержены ошибкам, а потому что мозг помнит важные взаимосвязи мира независимо от деталей. Давайте рассмотрим несколько примеров, иллюстрирующих это.

Как мы видели в главе 2, простая модель автоассоциативной памяти существует десятилетия, и, как сказано выше, мозг вспоминает автоассоциативно. Но между автоассоциативной памятью, построенной исследователями нейронных сетей, и памятью кортекса есть большая разница. Искусственная автоассоциативная память не использует инвариантное представление и, следовательно, она ошибочна в некоторых базовых принципах. Вообразите, что у меня есть изображение лица, сформированное большим набором черных и белых точек. Эта картинка – паттерн, и если у меня есть искусственная автоассоциативная память, я могу хранить множество картинок в этой памяти. Наша искусственная автоассоциативная память надежна в том, что если я дам половину лица, или даже пару глаз, она распознает эту часть и корректно заполнит оставшуюся часть. Этот эксперимент в точности был проделан несколько раз. Но если я сдвину каждую точку на пять пикселей влево, память совершенно не сможет распознать лицо. Для искусственной автоассоциативной памяти это совершенно новый паттерн, потому что ни один из пикселей сохраненного паттерна и нового паттерна не выровнены. Вы и я, конечно же, без труда видим сдвинутый паттерн как то же самое лицо. Возможно, мы даже не заметим изменений. Искусственная автоассоциативная память не может распознать паттерны, если они сдвинуты, повернуты, масштабированы или трансформированы одним из тысяч других способов, тогда как наш мозг легко справляется с этими вариациями. Что позволяет нам воспринимать что-то как то же самое или неизменное, тогда как поступающие паттерны, представляющие его, новые или измененные? Давайте рассмотрим другой пример.

Возможно, сейчас вы держите в руках книгу. Когда вы двигаете книгу, или меняете освещение, или по-другому усаживаетесь на стуле, или фиксируете глаза на другой части страницы, паттерны освещенности, попадающие на вашу сетчатку, полностью меняются. Визуальная информация изменяется от момента к моменту и никогда не повторяется. Фактически, вы могли бы держать эту книгу сотни лет и ни разу паттерны на вашей сетчатке, и, следовательно, попадающие в ваш мозг, не повторились бы. Но у вас ни на секунду не возникало бы сомнений, что книга, которую вы держите, действительно та самая книга. Внутренний паттерн вашего мозга, представляющий «эту книгу», не меняется, хотя стимулы, информирующие вас находятся в постоянном изменении. Поэтому мы будем использовать термин *инвариантное представление*, ссылаясь на внутреннее представление в мозгу.

Другой пример, вспомните лицо подруги. Вы узнаете ее каждый раз, когда видите. Это происходит автоматически менее чем за секунду. Не имеет значения, что она в двух шагах от вас, трех или на другой стороне комнаты. Когда она близко, изображение ее лица занимает всю вашу сетчатку. Когда она далеко, ее изображение занимает маленькую часть вашей сетчатки. Она может быть лицом к вам, немного боком или в профиль. Она может улыбаться, прищуриваться или зевать. Вы можете видеть ее в ярком свете, в тени или под специфическим освещением дискотеки. Ее лицо может появляться в бесчисленных позициях и вариациях. В каждом случае паттерны света, попадающие на сетчатку, уникальны, но в каждом случае вы точно знаете, что смотрите именно на нее.

Давайте приоткроем завесу и взглянем на то, что происходит в вашем мозгу во время выполнения этого изумительного трюка. Из экспериментов нам известно, что если мы будем отслеживать активность нейронов визуальной области кортекса, называемой V1, паттерны активности будут отличаться при каждом различном взгляде на лицо. Каждый раз, когда лицо двигается или ваши глаза фиксируются в новой точке, паттерны активности в V1 изменяются, практически подобно изменению паттернов на сетчатке. Однако, если мы будем отслеживать активность клеток в области, ответственной за распознавание лиц – функциональной области на несколько уровней выше по кортикальной иерархии, чем V1 – мы обнаружим стабильность. То есть, некоторое множество клеток в области, ответственной за распознавание лиц, остается активным, пока лицо вашей подруги находится в поле вашего зрения (или даже если воображается в мысленном поле зрения), несмотря на его размер, положение, ориентацию, масштаб и выражение лица. Эта стабильность возбуждения нейронов – инвариантное представление.

Интроспективно эта задача кажется настолько простой, что едва ли стоит называть ее проблемой. Она работает автоматически, как дыхание. Она кажется тривиальной, потому что мы не осознаем, как это происходит. И в некотором смысле, она тривиальна, потому что наш мозг может решить ее очень быстро (вспомните правило ста шагов). Однако, проблема понимания того, как кортекс формирует инвариантное представление, остается одной из величайших тайн науки. Насколько сложна эта проблема, спросите вы? Настолько, что никто, даже используя самый мощный компьютер в мире, не способен решить ее. И это не от не от недостатка количества попыток.

Размышления по данной проблеме имеют древние корни. Они доходят до Платона, 23 столетия назад. Платон удивлялся, каким образом люди способны мыслить и узнавать о мире. Он обратил внимание, что реальные воплощения вещей и идеи всегда несовершенны, и всегда различны. Например, у вас есть понятие совершенной окружности, хотя вы никогда не видели такую в действительности. Все нарисованные окружности несовершенны. Даже если она нарисована с помощью циркуля, окружность представляется темной линией, тогда как линия настоящей окружности не имеет толщины. Как же вы тогда вообще смогли воспринять понятие идеальной окружности? Или возьмем более житейский случай, подумаем о понятии «собака». Любая из собак, каких вы видели, отличается от других, и каждый раз вы видите именно конкретную собаку. Все собаки различны и вы никогда бы не смогли увидеть конкретную собаку одним и тем же образом. Но весь ваш опыт относительно собак сходится к ментальному понятию «собака», которое неизменно. Платон был сбит с толку. Как возможно такое, что мы запоминаем и применяем понятия в этом мире бесконечно разнообразных форм и постоянно меняющихся ощущений?

Решением Платона была его известная Теория Форм. Он пришел к выводу, что наш высший разум должен быть связан с некоторым трансцендентным миром сверхреальности, где в безвременном совершенстве существуют постоянные, неизменные идеи (Формы с заглавной Ф). Наши души приходят из этого мистического места в момент рождения, заключил он, где они узнают о Формах в их первоначальном виде. После рождения у нас сохраняются скрытые знания о них. Обучение и понимание происходит потому что формы реального мира напоминают нам о Формах, которым они соответствуют. Вы можете узнать про окружности и собак, потому что они соответственно вызывают воспоминания Окружности и Собаки.

Все это достаточно далеко от современных взглядов. Но если вы уберете высокопарную метафизику, вы увидите, что он в действительности говорил об инвариантности. Его система объяснения была далека от цели, но его интуиция, что это один из важнейших вопросов о природе человека, которыми мы можем задаться, попала в яблочко.

\* \* \*

**Ч**тоб у вас не сложилось впечатление, что инвариантность относится только к зрению, давайте взглянем на некоторые примеры с другими чувствами. Рассмотрим тактильные ощущения. Когда вы тянетесь к бардачку в машине, чтоб найти солнечные очки, ваши пальцы должны только слегка задеть их, чтоб вы их нашли. Не важно, какая часть вашей руки соприкоснется; это может быть большой палец, любая часть любого пальца или ладонь. Соприкоснуться можно с любой частью очков, с линзой, дужкой, петлей или частью оправы. Всего лишь движения любой части вашей руки над любой частью очков достаточно, чтоб ваш мозг идентифицировал их. В каждом случае поток пространственных и временных паттернов, поступающих с ваших тактильных рецепторов, совершенно различный – различные участки кожи, различные части объекта – но вы хватаете ваши очки без раздумий.

Или рассмотрим сенсомоторную задачу – вставить ключ в замок зажигания. Положение, в котором вы сидите, положение тела, рук, каждый раз слегка отличаются. Для вас это выглядит как повторяющееся изо дня в день действие, но это потому что у вас инвариантное представление в мозгу. Если вы попытаетесь сделать робота, который смог бы попасть в машину и вставить ключ, вы быстро бы увидели, что это практически невозможно, пока вы не будете уверены, что робот находится в одном и том же положении и держит ключ одинаково каждый раз. И даже если вы смогли бы управлять им, чтоб сделать это, робота пришлось бы перепрограммировать под различные машины. Роботы и компьютерные программы, подобно искусственной автоассоциативной памяти, ужасны при обращении с вариациями.

Другой интересный пример – это ваша подпись. Где-то в вашем моторном кортексе, в передних долях, у вас есть инвариантное представление вашего автографа. Каждый раз, когда вы подписываетесь, вы используете одну и ту же последовательность ударов, изгибов и ритмов. Это верно и когда вы подписываетесь ежеминутно хорошей ручкой на весу, и когда неловко подписываетесь карандашом, зажатым между кончиками пальцев. Каждый раз получается что-то разное, конечно, особенно в неудобных условиях, которые я только что указал. Тем не менее, несмотря на масштаб, пишущую принадлежность или комбинацию частей тела, вы всегда запускаете одну и ту же абстрактную «моторную» программу.

Из примера с подписью видно, что инвариантное представление в моторном кортексе в некотором смысле это зеркальное отражение инвариантного представления в сенсорном кортексе. С сенсорной стороны, большое множество входных паттернов может активировать один и тот же ансамбль, который представляет один и тот же абстрактный паттерн (лицо вашей подруги, ваши очки). С моторной стороны, один и тот же ансамбль, представляющий некоторую абстрактную моторную команду (ловля мяча, подпись) способен проявит себя в широком множестве мышечных команд и удовлетворить широкому множеству других ограничений. Эта симметрия между восприятием и действием и есть то, что мы должны ожидать, если, как предположил Монткастл, во всех областях кортекса работает единый базовый алгоритм.

В качестве последнего примера давайте вернемся к сенсорному кортексу и снова взглянем на музыку (мне нравится использовать музыкальную память в качестве примера, потому что легко увидеть все проблемы, которые должен решить неокортекс). Инвариантное представление музыки иллюстрируется вашей способностью распознавать мелодию в любой тональности. Тональность, в которой сыграна мелодия, ссылается на музыкальную шкалу, на которой построена мелодия. Одна и та же мелодия, сыгранная в различных тональностях, построена из различных нот. Выбрав тональность для исполнения, вы задаете все оставшиеся ноты в мелодии. Мелодия может быть воспроизведена в различных тональностях. Это значит, что каждое исполнение «одной и той же» мелодии в новой тональности – фактически полностью отличающаяся последовательность нот! Каждое исполнение стимулирует различные участки вашей улитки, вызывая совершенно различные пространственно-временные паттерны, поступающие в слуховой кортекс… и тем не менее вы воспринимаете одну и ту же мелодию в каждом случае. Если у вас нет абсолютного слуха, вы не сможете даже различить, что одна и та же сыграна в другой тональности, если только не будете сравнивать.

Представьте песню «Somewhere over the Rainbow». Возможно вы впервые услышали ее в исполнении Джуди Гарленд в фильме «Волшебник страны Оз», но если у вас нет абсолютного слуха, то возможно вы не сможете даже вспомнить тональность, в которой она ее исполняла (Ля минор). Если я сяду за фортепиано и начну играть песню в такой тональности, в которой вы ее никогда не слышали – скажем, в Ре миноре – она будет звучать как точно та же самая песня. Вы не заметите, что все ноты отличаются от нот в той версии, с которой вы знакомы. Это означает, что ваша музыкальная память должна быть в форме, игнорирующей высоту ноты. Память должна хранить существенные отношения между нотами, но не сами ноты. В данном случае существенными отношениями являются относительная высота ноты, или «интервал». «Somewhere over the Rainbow» начинается с октавы вверх, затем полутон вниз, затем большая терция вниз и т.д. Интервальная структура мелодии одна и та же при исполнения в любой тональности. Ваша способность легко узнавать песню в любой тональности показывает, что ваш мозг хранит ее в форме, инвариантной к высоте нот.

Аналогично, память о лице вашей подруги также должно храниться в форме, не зависящей от особенностей точки зрения. Что делает ее лицо узнаваемым, так это его относительные размеры, относительные оттенки цвета, относительные пропорции, а не то, как оно выглядело в последний вторник в обед. Есть некоторые «пространственные интервалы» между чертами ее лица, также как и «высотные интервалы» между нотами в песне. Ее лицо шире относительно ее глаз. Ее нос короче относительно ширины ее глаз. Цвет ее волос и цвет ее глаз связаны подобным отношением, которое остается постоянным даже если при различном освещении их абсолютные цвета значительно изменяются. Когда вы запоминаете ее лицо, вы запоминаете эти относительные атрибуты.

Я верю, что подобное абстрагирование формы возникает во всем кортексе, в любой его области. Это общее свойство неокортекса. Память хранится в форме, которая фиксирует суть отношений, а не конкретные детали. Когда вы видите, чувствуете или слышите что-либо, кортекс получает детальную, высоко специфическую информацию и конвертирует ее в инвариантную форму. Именно инвариантная форма хранится в памяти, и любая новая информация сравнивается именно в инвариантной форме. Хранение, вспоминание и узнавание происходят в инвариантной форме. В вычислительной технике нет эквивалентной концепции.

\* \* \*

**Э**то поднимает интересную проблему. В следующей главе я докажу, что важной функцией кортекса является использование памяти для предсказания. Но исходя из того, что кортекс хранит воспоминания в инвариантной форме, каким образом он делает конкретные предсказания? Вот некоторые примеры, иллюстрирующие проблему и ее решение.

Вообразите, что сейчас 1980 год и вы в пограничном городке в Западной Америке. Ваша возлюбленная едет поездом из Восточной Америки, чтоб встретиться с вами в новом пограничном городке. Конечно же, вы хотите встретить ее на станции, когда она приедет. За несколько недель до приезда вы узнаете, когда поезд выходит и приходит. Расписания нет, и пока вы можете сказать, что поезд никогда не прибывает и не убывает в одно и тоже время дня. С такой позиции вы никогда не сможете предсказать, когда приедет ее поезд. Но затем вы замечаете, что есть некая структура в том, как поезд прибывает и убывает. Поезд с востока прибывает через четыре часа после того, как уедет на восток. Этот четырехчасовой перерыв остается постоянным изо дня в день, хотя конкретное время изменяется. В день ее прибытия вы отслеживаете поезд, уходящий на восток, и засекаете время. Через четыре часа вы возвращаетесь на станцию и встречаете ее поезд прямо в момент прибытия. Эта притча иллюстрирует и проблему распознавания лиц неокортексом, и то, как он ее решает.

Мир, ощущаемый вашими органами чувств никогда не повторяется; также как время отправки и прибытия поезда, он все время разный. Способ, которым вы понимаете мир, заключается в том, чтоб найти инвариантную структуру в постоянно изменяющемся потоке информации. Однако, этой инвариантной структуры самой по себе не достаточно для использования в качестве основы для предсказаний. Знание только того, что поезд приходит через четыре часа после отправки, не позволяет вам появиться на платформе вовремя для встречи возлюбленной. Чтоб сделать конкретное предсказание, мозг должен совместить знание инвариантной структуры с самыми последними деталями. Предсказание времени прибытия требует выявления четырехчасовой структуры в расписании поезда и комбинации ее с детальным знанием того, во сколько ушел последний поезд на восток.

Когда вы слушаете знакомую мелодию, исполняемую на фортепиано, ваш мозг предсказывает следующую ноту до того, как она будет сыграна. Но песня, как мы видели, хранится в памяти в форме, инвариантной к высоте нот. Ваша память говорит вам, какой интервал будет следующим, но не говорит ничего о конкретной ноте. Для предсказания точной ноты требуется комбинация следующего интервала с последней конкретной нотой. Если последний интервал это большая терция и последней нотой, которую вы слышали, было До, то вы можете предсказать, что следующей нотой будет Ми. Вы слышите в вашей голове Ми, а не «большую терцию». И если вы не ошибаетесь с идентификацией песни и пианист не спотыкается, ваши предсказания будут верными.

Когда вы видите лицо вашей подруги, ваш кортекс заполняется и предсказывает мириады деталей ее уникального изображения в данный момент. Он сверяется, что ее глаза правильные, что ее нос, губы и волосы точно такие, какие должны быть. Ваш кортекс делает эти предсказания с большой специфичностью. Он может предсказать низкоуровневые детали ее лица даже если вы никогда не видели ее в такой конкретной ориентации или в таком окружении. Если вы точно знаете, как располагаются глаза и нос вашей подруги, и вы знаете структуру ее лица, то вы можете точно предсказать, как должны располагаться ее губы. Если вы знаете, что ее кожа имеет оранжевый оттенок в свете заката, то вы знаете, какой оттенок должны приобрести ее волосы. И снова ваш мозг делает это путем комбинации памяти о инвариантной структуре ее лица с особенностями текущего опыта.

Пример с расписанием поезда всего лишь аналогия того, что происходит в вашем кортексе, но мелодия и лицо – нет. Комбинирование инвариантного представления и текущей информации для детального предсказания – это именно то, что происходит. Это повсеместный процесс, происходящий в каждой области вашего кортекса. Именно так вы делаете конкретные предсказания относительно комнаты, в которой вы сейчас сидите. Именно так вы способны предсказать не только слова, которые вам говорят другие, но также и интонацию, с которой они это скажут, каким тоном они это скажут и из какой части комнаты вы должны услышать голос. Именно так вы точно знаете, когда ваша нога коснется пола, и что она почувствует, когда вы будете подниматься по ступенькам. Именно так вы способны написать свое имя ногой или поймать брошенный мяч.

Три свойства кортикальной памяти, рассмотренные в этой главе (хранение последовательностей, авто-ассоциативность и инвариантное представление), являются необходимыми ингредиентами для предсказания будущего на основе памяти о прошлом. В следующей главе я покажу, что предсказание – это суть интеллекта.

# 5. Новая структура интеллекта

**Однажды** в апреле 1986 года я размышлял над тем, что значит «понимать» что-либо. Несколько месяцев я бился над фундаментальным вопросом, Чем же занимается мозг, если он не продуцирует поведение? Что делает мозг, когда он пассивно слушает речь? Что делает ваш мозг прямо сейчас, когда он читает? Информация поступает в мозг, но не выходит из него. Что с ним происходит? Ваше поведение в данный момент возможно элементарное – вроде дыхания или движения глаз – но, как вы уже знаете, ваш мозг делает гораздо больше, чем это, поскольку вы читаете и понимаете. Понимание должно быть результатом нервной активности. Но какой? Что нейроны делают, годна они понимают?

Когда я оглядел мой офис в тот день, я увидел знакомые стулья, плакаты, окна, растения, карандаши и т.д. Вокруг меня были сотни вещей. Мои глаза видели их, когда я мельком оглядывался, но просто созерцание их не вызывало у меня никаких действий. Никакого поведения не осуществлялось и не требовалось, но как-то я «понимал» комнату и ее содержимое. Я делал то, чего не могла сделать Китайская Комната Серла, и мне не требовалось возвращать что-либо через прорезь. Я понимал, но никакие действия не могли доказать этого. Что означало «Понимать»?

Как раз когда я размышлял над этой дилеммой, у меня появилось озарение, один из тех эмоционально сильных моментов, когда то, что было путаницей, неожиданно становилось ясным и понятным. Все что я сделал – это спросил, что произошло бы, если бы в комнате появился новый объект, которого я никогда раньше не видел – скажем, синяя кофейная чашка.

Ответ казался простым. Я должен был заметить новый объект, как несоответствующий. Он должен был поймать мое внимание, как нечто новое. Мне не надо было бы спрашивать себя сознательно, является ли кофейная чашка новой. Это просто должно было возникнуть, как несоответствие. Кажущийся тривиальным, лежащий в основе ответ является мощной концепцией. Чтобы заметить, что что-то отличается, некоторые нейроны в моем мозгу, которые не были активны до этого, должны были стать активными. Как эти нейроны узнали, что кофейная чашка является новой, а сотни других предметов – нет? Ответ на этот вопрос до сих пор для меня является неожиданным. Наш мозг хранит воспоминания, чтобы постоянно делать предсказания обо всем, что мы видим, чувствуем, слышим. Когда я оглядываю комнату, мой мозг использует память для формирования предсказаний о том, что он ожидает почувствовать, прежде чем я почувствую это. Громадное количество предсказаний возникает за пределами понимания. Это как если бы отдельные части моего мозга говорили: «Посреди стола есть компьютер? Да. Он черный? Да. В правом углу стола есть лампа? Да. Лежит ли словарь там, где я его оставил? Да. Прямоугольно ли окно, вертикальны ли стены? Да. С того ли направления светит солнце для данного времени дня? Да». Но когда поступает некоторый визуальный паттерн, которого я не помню в этом контексте, предсказание нарушается. И ошибка привлекает мое внимание.

Конечно, мозг не разговаривает сам с собой, когда делает предсказания, и он делает предсказания не последовательным образом. Он также не просто делает предсказания о различных объектах вроде кофейной чашки. Ваш мозг постоянно совершает предсказания о структуре того самого мира, в котором мы живем, и делает он это параллельным образом. Это будет верным и когда мы быстро обнаруживаем необычную текстуру, уродливый нос или необычное движение. Сразу не скажешь, насколько обширны эти в основном подсознательные предсказания, что, возможно, является причиной того, почему мы так долго упускали их важность. Они происходят настолько автоматически, настолько легко, что мы ошиблись в оценке того, что происходит в наших головах. Я надеюсь убедить вас в могуществе этой идеи. Предсказания настолько обширны, что наше «восприятие» - то есть, каким мы видим мир – происходит не только от наших органов чувств. То, что мы воспринимаем – это комбинация того, что мы чувствуем, и предсказаний, основанных на памяти нашего мозга.

\* \* \*

**Н**есколько минут спустя я провел мысленный эксперимент, чтоб помочь выразить то, что я понял в тот момент. Я называю его экспериментом с измененной дверью. Вот он.

Когда вы приходите каждый день домой, вам обычно нужно несколько секунд, чтоб войти через парадную дверь или ту дверь, какую вы используете. Вы подходите, поворачиваете ручку, входите и захлопываете ее за собой. Это крепко устоявшаяся привычка, то, что вы делает все время и уделяете этому мало внимания. Предположим, пока вас не было, я прокрался в ваш дом и заменил что-то в вашей двери. Это может быть что угодно. Я мог бы передвинуть ручку на дюйм, заменить вращающуюся ручку на шпингалет, или заменить бронзовую на хромированную. Я мог бы изменить вес вашей двери, заменив сплошную древесину на полую, или наоборот. Я мог бы сделать петли скрипучими и тугими, или сделать их легко скользящими. Я мог бы сделать шире или уже дверь и дверной проем. Я мог бы изменить ее цвет, добавить дверное кольцо туда, где должен быть дверной глазок, или добавить окошко. Я могу вообразить тысячи изменений, которые мог бы сделать с вашей дверью без вашего ведома. Когда вы вернетесь домой в тот день и попытаетесь открыть дверь, вы быстро обнаружите что что-то не так. Может пройти несколько секунд, пока вы поймете, что именно не так, но сам факт изменения вы заметите очень быстро. Когда ваша рука потянется к дверной ручке, вы обнаружите, что она не на своем месте. Или когда вы увидите новое дверное окошко, что-то покажется странным. Или если вес двери изменился, вы толкнете ее с неправильным усилием и будете удивлены. Суть в том, что вы заметите любое из тысячи изменений за очень короткий промежуток времени.

Как вы это делаете? Как вы замечаете эти изменения? Специалисты в ИИ или компьютерщики подходят к этой проблеме так, что создается список свойств двери и заносится в базу данных, поля которой могут содержать отдельные записи для каждой конкретной двери. Когда вы подходите к двери, компьютер опросил бы всю базу данных, глядя на ширину, цвет, размер, положение ручки, вес, звук и т.д. Хотя поверхностно это звучит аналогично тому, как я описывал выше то, как мой мозг проверяем мириады предсказаний, когда я оглядываю мой офис, на самом деле есть далеко идущие отличия. ИИ-стратегия нереальна. Во-первых, невозможно заранее указать все атрибуты, которые могут быть у двери. Список потенциально бесконечен. Во-вторых, нам понадобились бы подобные списки для любых объектов, которые мы встречаем ежесекундно в нашей жизни. В-третьих, ничто из того, что нам известно о мозге и нейронах, не позволяет предположить, что они работают именно так. И наконец, нейроны просто слишком медленны, чтоб реализовать базы данных в компьютерном стиле. У вас бы заняло двадцать минут вместо пары секунд, чтоб заметить изменения в двери.

Есть только один способ интерпретировать вашу реакцию на измененную дверь: ваш мозг делает низкоуровневые предсказания по поводу того, что он ожидает увидеть, услышать, почувствовать в каждый момент, и делает он это параллельно. Все области вашего неокортекса одновременно пытаются предсказать, какими должны быть их следующие ощущения. Визуальные области делают предсказания по поводу краев, форм, объектов, расположения и движений. Слуховые области делают предсказания по поводу высоты звуков, направлений на источник и звуковых паттернов. Соматосенсорные области делают предсказания о прикосновениях, текстурах, контурах и температуре.

«Предсказание» обозначает, что нейроны, задействованные в восприятии вами двери становятся активными до того, как они получат сенсорную информацию. Когда поступает сенсорная информация, она сравнивается с тем, что ожидалось. Когда вы подходите к двери, ваш кортекс формирует массу предсказаний на основе предыдущего опыта. Когда вы подошли к ней, он предсказывает, что почувствуют ваши пальцы, когда вы прикоснетесь к двери, и в каком положении будут ваши суставы, когда вы действительно дотронетесь до двери. Когда вы начинаете открывать дверь, ваш кортекс предсказывает, какое сопротивление двери возникнет и как она будет звучать. Когда все ваши предсказания совпадут, вы войдете в дверь не осознавая, что эти предсказания были проверены. Но если ваши ожидания нарушены, ошибка заставит вас обратить внимание. Верные предсказания приведут к пониманию, что дверь нормальна, неверные предсказания приведут вас в замешательство и вынудят вас обратить внимание. Дверная защелка не там, где предполагается. Дверь слишком легкая. Дверь не по центру. Текстура ручки не та. Мы делаем непрерывные низкоуровневые предсказания параллельно для всех органов чувств.

Но это не все. Я занимаю более строгую позицию. Предсказания не просто одна из вещей, которые делает ваш мозг. Это *первичная функция* неокортекса и фундамент интеллекта. Кортекс – это орган предсказаний. Если мы хотим понять, что такое интеллект, что такое творчество, как работает мозг и как построить интеллектуальные машины, мы должны понять природу эти предсказаний и как кортекс их делает. Даже поведение лучше понимать как продукт предсказаний.

\* \* \*

**Я** не знаю, кто первым предположил, что предсказания это ключ к пониманию интеллекта. В науке и промышленности никто не изобретал что-то полностью новое. Наоборот, люди видели, как существующие идеи укладываются в новые рамки. Компоненты новой идеи обычно витают в среде научных обсуждений, прежде чем ее откроют. Что является новым, так это упаковка этих компонентов в единое целое. Аналогично, идея о том, что первичная функция неокортекса делать предсказания, не является полностью новой. Она витала в различных формах некоторое время. Но еще не предполагалось, что ее законное положение в центре теории о мозге и в определении интеллекта.

По иронии судьбы, некоторые из пионеров искусственного интеллекта уделяли внимание компьютерным моделям мира с использованием предсказаний. В 1956 году, например, Д.М.Маккэй утверждал, что интеллектуальные машины должны иметь «механизм внутреннего ответа» предназначенный для «сопоставления входной информации». Он не использовал слова «память» и «предсказание», но он размышлял в том же направлении.

С середины 90-х такие термины, как *умозаключение*, *порождающие модели* и *предсказание* прокрались в научную номенклатуру. Все они ссылались на подобные идеи. Например, в 2001 году в своей книге «*I of the vortex*» Рудольфо Ллинас из Нью-Йоркского Медицинского Университета писал: «Способность предсказывать исход будущих событий – критическая для успешного продвижения – вероятнее всего является предельной и наиболее общей из всех функций мозга». Такие ученые, как Дэвид Мамфорд из Университета Брауна, Раеш Рао из Вашингтонского Университета, Стивен Гроссберг из Бостонского Университета и множество других писали и теоретизировали о роли обратных связей и предсказании. Есть целый подраздел математики, посвященный байесовским сетям. Названные в честь Томаса Байеса, английского министра, родившегося в 1702 году, пионера статистики, байесовские сети используют теорию вероятностей для предсказания.

Чего не хватало, так это собрать эти несопоставимые куски в согласованную теоретическую основу. Я утверждаю, что этого не было сделано ранее, и это является целью данной книги.

\* \* \*

**П**режде чем мы приступим к детальному изучению того, как кортекс делает предсказания, давайте рассмотрим несколько дополнительных примеров. Чем больше вы будете думать над этой идеей, тем больше вы осознаете, что предсказания сильно распространены и являются базисом для понимания мира.

Этим утром я готовил оладьи. На одном из этапов этого процесса я потянулся не глядя к дверце шкафчика. Я интуитивно знал, не видя, что я должен почувствовать – в данном случае ручку дверцы шкафчика – и когда я должен ее почувствовать. Я поворачивал крышку молочной канистры, ожидая, что она должна повернуться и открыться. Я повернулся к плите, ожидая небольшого нажатия и поворота с определенным усилием. Я ожидал услышать звук загоревшегося пламени через несколько секунд. Каждую минуту на кухне я делал дюжины или сотни движений и каждое содержало множество предсказаний. Я знал это, потому что если какое-либо из этих движение привело бы к другому результату, я бы обратил внимание.

Каждый раз, когда вы ставите ногу при ходьбе, ваш мозг предсказывает, когда ваша нога должна прекратить движение и каким эластичным должен быть материал, на который вы наступите. Если когда-нибудь вам попадется лестничный пролет с отсутствующей ступенькой, вы знаете, как быстро обнаружите, что что-то не так. Вы опускаете ногу и в тот момент, когда она «проходит сквозь» ожидаемую ступеньку, вы знаете, что у вас проблема. Нога ничего не чувствует, но ваш мозг делает предсказание и предсказание не сбывается. Робот с компьютерным управлением конечно же упадет, не осознавая, что что-то неправильно, тогда как вы узнаете об этом, как только ваша нога продвинется хоть на долю дюйма за тот предел, на котором мозг ожидал ее остановки.

Когда вы слушаете знакомую мелодию, вы слышите следующую ноту прежде, чем она возникнет. Когда вы слушаете любимый сборник, вы слышите начало каждой следующей песни за несколько секунд до того, как она начнется. Почему так происходит? Нейроны вашего мозга, которые должны возбудиться, когда вы услышите следующую ноту, возбуждаются до того, как вы действительно ее слышите, таким образом вы «слышите» песню в своей голове. Нейроны возбуждаются в ответ на память. Эта память может быть удивительно долгой. Не удивительно услышать музыкальный сборник впервые за несколько лет и все еще автоматически слышать следующую песню после окончания предыдущей. Создается забавное ощущение небольшой неопределенности, когда вы слушаете любимый компакт-диск в случайном порядке; вы знаете, что ваше предсказание следующей песни будет ошибочным.

Когда вы слушаете говорящего человека, вы часто знаете, что он собирается сказать, прежде чем он закончит говорить – или, по крайней мере, вы думаете, что знаете! Иногда мы даже не слышим, что он действительно говорит, а вместо этого слышим то, что ожидаем услышать (это часто происходило со мной, когда я был ребенком, так что мать дважды водила меня к врачу для проверки слуха). Вы это частично ощущаете, потому что люди склонны использовать общие фразы или выражения в своем общении. Если я скажу «How now brown…,» ваш мозг активирует нейроны, представляющие слово *cow* прежде, чем я его скажу (хотя, если английский не является вашим родным языком, вы не поймете, о чем я говорю). Конечно, вы не можете постоянно знать, что собираются сказать другие. Предсказания не всегда точны. Наоборот, наш разум работает, делая вероятностные предсказания, касающиеся того, что вот-вот должно произойти. Иногда мы точно знаем, что должно произойти, а иногда наши ожидания распределены по нескольким возможным исходам. Если мы будем ужинать за столом и я скажу, «пожалуйста, передайте мне…», для вашего мозга не будет неожиданностью, если я скажу «соль», «перец» или «горчицу». В некотором смысле предсказывает все эти возможные исходы одновременно. Однако, если я скажу, «пожалуйста, передайте мне тротуар», вы поймете, что что-то не так.

Возвращаясь к музыке, мы также можем найти здесь вероятностные предсказания. Если вы слышите песню, которую никогда раньше не слышали, у вас все равно могут быть достаточно сильные ожидания. В западной музыке я ожидаю услышать регулярный размер, повторяющийся ритм, я ожидаю услышать музыкальные фрагменты с одинаковым количеством долей, я ожидаю, что песня закончится тоникой. Вы можете не знать, что обозначают эти термины, но – полагая, что вы слышали подобные мелодии – ваш мозг автоматически предсказывает сильные доли, повторяющиеся ритмы, окончания фрагментов и конец песни. Если новая песня нарушает эти принципы, вы тут же чувствуете, что что-то не так. Задумайтесь об этом на несколько секунд. Вы слышите песню, которую никогда раньше не слышали, ваш мозг ощущает паттерны, которые он никогда раньше не ощущал, но тем не менее вы делаете предсказание и можете сказать, что что-то не так. Основанием для этих в основном бессознательных предсказаний являются воспоминания, хранящиеся в вашем кортексе. Ваш мозг не может точно сказать, что должно произойти после, но тем не менее он предсказывает, какой нотный паттерн скорее всего должен быть, а какой – нет.

У нас у всех должен быть опыт, когда мы неожиданно замечали, что источник постоянного фонового шума, такой как отдаленный отбойный молоток или приглушенная фоновая музыка, прекращается – хотя мы не замечали этого звука, пока он звучал. Ваши звуковые области предсказывают его продолжение, момент за моментом, и пока звук не меняется, вы не обращаете внимания. Прекратившись, он нарушает ваше предсказание и привлекает внимание. Вот исторический пример. Сразу после того, как в Нью-Йорке перестали ходить наземные поезда, люди вызывали полицию среди ночи, заявляя, что их что-то разбудило. Они имели тенденцию вызывать полицию как раз в то время, в которое раньше за пределами их квартир ходили поезда.

Иногда говорят, что видеть – значит верить. Иногда мы видим то, что ожидаем увидеть, так часто, как видим то, что реально видим. Один из наиболее восхитительных примеров этого является то, что исследователи называют замещением. Возможно вы ранее замечали, что у вас в каждом глазу есть небольшая слепая область, где ваш оптический нерв покидает сетчатку через отверстие, называемое оптическим диском. У вас нет фоторецепторов в этой области, так что вы постоянно слепы в соответствующей области визуального поля. Есть две причины, по которым вы обычно не замечаете этого, одна обычная, другая – весьма занимательная. Обычная причина это то, что слепые области не перекрываются, так что каждый глаз компенсирует другой.

Но интересно, что вы все равно не замечаете слепой области, когда открыт только один глаз. Ваша визуальная система «восполняет» отсутствующую информацию. Когда вы закрываете один глаз и смотрите на роскошно вытканный турецкий ковер или на волнистые контуры древесных волокон в полировке стола, вы не видите дыр. Целые узелки в ковре, целые темные прожилки древесных волокон постоянно выпадают из поля зрения вашей сетчатки, когда ваша слепая область попадает на них, но вы видите цельные участки текстуры и цвета. Ваш визуальный кортекс дорисовывает по памяти подобные паттерны и создает непрерывный поток предсказаний, которые восполняют любую отсутствующую информацию.

Замещение возникает во всех частях визуального изображения, не только в слепой области. Например, я показываю вам картину побережья с выброшенными на берег поленьями, лежащими на камнях. Границы между камнями и поленьями четкие и очевидные. Однако, если приблизить изображение, вы увидите, что камни и поленья имеют похожую текстуру и цвет в тех местах, где соприкасаются. В увеличенном виде края поленьев вообще неотличимы от камней. Если мы смотрим на картину целиком, края поленьев четкие, но в действительности мы домысливаем края по остальной части изображения. Когда мы смотрим на мир, мы воспринимаем четкие линии и границы, разделяющие объекты, но сырые данные, поступающие в глаз, часто зашумлены и неоднозначны. Наш кортекс заполняет отсутствующие или загрязненные участки тем, что мысленно там должно быть. Мы воспринимаем изображения недвусмысленно.

Визуальное предсказание также лежит в основе того, как двигаются ваши глаза. В главе 3, я упомянул саккады. Около трех раз в секунду ваши глаза фиксируются в одной точке, затем неожиданно перепрыгивают на другую. Обычно вы не осознаете эти движения, и обычно сознательно их не контролируете. И каждый раз, когда ваши глаза фиксируются в новой точке, паттерн, поступающий из глаз в ваш мозг, полностью меняется по сравнению с предыдущей фиксацией. Таким образом три раза в секунду ваш мозг видит что-то полностью отличное. Саккады не совсем случайны. Когда вы смотрите на лицо, ваши глаза обычно фиксируются в первую очередь на одном глазу, затем на другом, перескакивая туда-сюда, изредка фиксируясь на носу, на рте, на ушах и на других чертах лица. Вы воспринимаете просто «лицо», но ваши глаза видят глаз, глаз, нос, рот, глаз и т.д. я отдаю себе отчет в том, что это для вас является неосознаваемым. Все, что вы осознаете – это непрерывное изображение мира, но сырые данные, поступающие в вашу голову, такие же отрывистые, как с видеокамеры в неумелых руках.

Теперь представьте, что вы встречаете кого-то с дополнительным носом на том месте, где должен быть глаз. Ваши глаза фиксируют сначала один глаз, затем совершают саккаду на второй глаз, но вместо того, чтоб увидеть глаз вы видите нос. Вы определенно знали бы, что что-то не так. Чтоб так произошло, у вашего мозга должны быть ожидания или предсказания того, что он должен увидеть. Когда вы предсказываете глаз, но видите нос, предсказание нарушается. Так несколько раз в секунду попутно с каждой саккадой ваш мозг делает предсказания о том, что он должен увидеть. Когда это предсказание нарушается, немедленно привлекается ваше внимание. Вот почему так трудно не смотреть на людей с физическими дефектами. Если б вы увидели кого-то с двумя носами, было бы трудно удержаться от того, чтоб не уставиться на него, так ведь? Конечно, если вы живете с этим человеком, то через некоторое время вы привыкните к двум носам и больше не будете замечать ничего необычного.

Теперь подумайте о самом себе. Какие предсказания вы делаете? Когда вы переворачиваете страницы этой книги, вы ожидаете, что страница согнется до определенной степени и перевернется предсказуемым образом, который отличается от того, как переворачивается обложка. Если вы сидите, вы предсказываете, что ощущение давления на ваш тело не будет изменяться; но если намокшее сиденье начало бы прогибаться, или подверглось другим неожиданным изменениям, вы бы перестали уделять внимание книге и попытались бы сообразить, что происходит. Если вы потратите некоторое время на наблюдения за собой, вы можете начать понимать, что ваше восприятие мира, ваше понимание мира сильно связано с предсказанием. Ваш мозг построил модель мира и постоянно сравнивает модель с реальностью. Вы знаете, где вы и что делаете благодаря обоснованности этой модели.

Предсказание не ограничено паттернами низкоуровневой сенсорной информации вроде зрения и слуха. До сих пор я ограничивался обсуждением таких примеров, потому что это простейший путь познакомить вас с системой взглядов для понимания интеллекта. Однако, согласно принципу Монткастла, что верно для низкоуровневых сенсорных областей, должно быть верным для всех кортикальных областей. Человеческий мозг интеллектуальнее, чем у других животных, потому что он может делать предсказания о более абстрактных видах паттернов и более длинных временных последовательностях паттернов. Чтобы предсказать, что скажет моя жена, когда увидит меня, я должен знать, что она говорила раньше, что сегодня пятница, что до пятничного вечера надо вынести мусор, что я не сделал этого в течение недели, и что у жены определенное выражение лица. Когда она откроет рот, у меня будут очень сильные предсказания по поводу того, что она скажет. В данном случае я точно не знаю, какие слова она скажет, но я знаю, что она напомнит мне о том, что надо вынести мусор. Важное замечание, что высокий интеллект не является каким-то видом процесса, отличающихся от процессов перцептивного мышления. Он работает на той же самой неокортикальной памяти и алгоритме предсказания.

Заметьте, что наши тесты на интеллект по существу тесты на предсказание. С детского сада до колледжа IQ-тесты основаны на формировании предсказаний. Задана последовательность чисел, какое число должно быть следующим? Задано три различных проекции сложного объекта, какая из следующих также является его проекцией? Слово A относится к слову B, как слово C относится к какому слову?

Наука сама по себе это упражнения в предсказаниях. Мы продвигаем наше знание о мире через производство гипотез и их проверку. Эта книга по существу является предсказанием о том, что такое интеллект и как работает мозг. Даже дизайн продукции в корне процесс предсказания. При разработке дизайна одежды или мобильных телефонов дизайнеры и инженеры пытаются предсказать, как поведут себя конкуренты, чего хотят потребители, сколько будет стоить новый дизайн и какой стиль будет востребован.

Интеллект измеряется способностью вспоминать и предсказывать паттерны в мире, включая язык, математику, физические свойства объектов и социальные ситуации. Ваш мозг получает паттерны из внешнего мира, сохраняет из в виде воспоминаний и делает предсказания путем комбинации того, что он видел ранее и что происходит сейчас.

\* \* \*

**В** этом месте вы могли бы подумать: «Я согласен, что мой мозг занимается предсказаниями и я могу быть интеллектуальным, просто лежа в темноте. Как вы указали, мне не нужно что-либо делать, чтоб понимать или быть интеллектуальным. Но не являются ли те ситуации исключением? Вы действительно аргументируете, что интеллектуальное понимание и поведение полностью разрозненны? В конечном счете не поведение ли делает нас интеллектуальными, а не предсказание? В конце концов поведение играет конечную определяющую роль в выживании».

Это, конечно, справедливый вопрос, в конечном счете именно поведение имеет наибольшее значение в выживании животного. Предсказание и поведение не являются полностью разрозненными, но их отношения очень тонкие. Во-первых, неокортекс появился на эволюционной сцене уже после того, как животные обрели сложное поведение. Следовательно, значение кортекса для выживания должно в первую очередь рассматриваться в терминах постепенного усовершенствования, которое он мог дать для животных с уже существующим поведением. Сначала появилось поведение, затем интеллект. Во-вторых, большая часть из того, что мы ощущаем, в большей степени зависит от того, что мы делаем и как мы перемещаемся в мире. Следовательно предсказание и поведение близко соотносятся. Давайте взглянем на то, что из этого следует.

Млекопитающие эволюционировали с большим неокортексом, потому что он давал им некоторое преимущество в выживании, и эти преимущества укоренились в поведении. Но в начале кортекс служил для более эффективного использования существующего поведения, а не для того, чтоб создавать совершенно новое поведение. Чтобы прояснить это, на необходимо взглянуть на то, как эволюционировал мозг.

Простые нервные системы возникли вскоре после того, как многоклеточные создания начали распространяться по Земле, сотни миллионов лет назад, но история настоящего интеллекта началась совсем недавно с предками наших рептилий. Рептилии успешно завоевывали сушу. Они распространились по всем континентам и модифицировались в многочисленные виды. У них были острые чувства и хорошо развитый мозг, который наделил их сложным поведением. Их прямые потомки, живущие сегодня рептилии, до сих пор обладают им. У аллигатора, например, сложное восприятие, так же как у вас и у меня. У него хорошо развитые глаза, уши, нос, рот и кожа. Он проявляет сложное поведение, включающее способности к плаванию, бегу, маскировке, охоте, устройству засад, гнездованию и поиску партнеров.

В чем же различие между мозгом рептилии и человека? Это различие и большое и маленькое. Я сказал «маленькое», потому что в грубом приближении все, что существует в мозге рептилии, существует и в мозге человека. Я сказал «большое», потому что в мозге человека есть нечто действительно важное, чего нет у рептилий – большой кортекс. Иногда вы слышали ссылку на «старый» или «примитивный» мозг. У каждого человека есть наиболее древние структуры в мозге, такие же как у рептилий. Они регулируют кровяное давление, голод, половое влечение, эмоции и множество других аспектов движения. Когда вы стоите, сохраняете равновесие и ходите, например, вы в основном полагаетесь на старый мозг. Если вы слышите пугающий звук, паникуете и начинаете бежать – это в основном ваш старый мозг. Вы не нуждаетесь в чем то большем, чем мозг рептилии, чтоб совершать множество интересных и полезных вещей. Так что же делает неокортекс, если в нем нет острой необходимости для того, чтоб видеть, слышать и двигаться?

Млекопитающие более интеллектуальны, чем рептилии, по причине обладания неокортексом. (Само это слово происходит от латинских слов, обозначающих «новая кора», или «новая корка», потому что кортекс буквально покрывает старый мозг). Неокортекс впервые появился десятки миллионов лет назад, и он есть только у млекопитающих. Что делает человека умнее других млекопитающих, так это в основном большая поверхность неокортекса, которая значительно увеличилась всего несколько миллионов лет назад. Вспомните, что кортекс построен из одинаковых повторяющихся элементов. Кортикальный слой у человека такой же толщины и имеет примерно такую же структуру, как и у млекопитающих родственников. Когда эволюция делает что-то большое очень быстро, как это произошло с человеческим кортексом, она делает это копированием существующих структур. Мы стали умнее путем добавления значительного количества элементов, работающих на общем кортикальном алгоритме. Существует общее заблуждение, что человеческий мозг плод миллиардов лет эволюции. Это может быть верным, если говорить о нервной системе в целом. Однако, человеческий неокортекс сам по себе относительно новая структура и существует недостаточно долго, чтоб подвергнуться существенным эволюционным улучшениям.

Отсюда начинается ядро моей аргументации того, как понимать неокортекс, и почему память и предсказание являются ключами к загадке интеллекта. Мы начнем с мозга рептилий, не обладающего кортексом. Эволюция обнаружила, что если приделать систему памяти (неокортекс) к сенсорным путям примитивного мозга, животные получают способность к предсказанию будущего. Вообразите, что старый мозг рептилий так и занимается своими делами, но теперь сенсорные паттерны одновременно подаются в неокортекс. Неокортекс хранит сенсорную информацию в своей памяти. В будущем, когда животное попадает в ту же самую или похожую ситуацию, память распознает поступающую информацию как похожую и вспоминает, что происходило в прошлом. Воспоминания сравниваются с потоком сенсорной информации. Вместе они «замещают» текущий сенсорный поток и предсказывают, что будет дальше. Сравнивая актуальный сенсорный поток с воспоминаниями, животное не только понимает, где оно, но может также предвидеть будущее.

Теперь вообразите, что кортекс помнит не только то, что животное видело, но также помнит поведение старого мозга в подобной ситуации. Мы даже не предполагаем, что кортекс знает о различиях между ощущением и поведением; для кортекса все это паттерны. Когда животное обнаруживает себя в такой же или подобной ситуации, он не только предвидит будущее, но также вспоминает, какое поведение привело к такому видению будущего. Таким образом, память и предсказание позволяют животному использовать его существующее поведение (старый мозг) более интеллектуально.

Например, вообразите, что вы крыса, изучающая лабиринт в первый раз. Побуждаемый неудовлетворенностью или голодом, вы используете навыки, унаследованные от старого мозга, чтобы изучать новую обстановку – слушаете, смотрите, нюхаете и пробираетесь вдоль стен. Вся сенсорная информация используется вашим мозгом, но также передается в ваш неокортекс, где сохраняется. В какой-то момент вы обнаруживаете себя в том же самом лабиринте. Ваш неокортекс распознает текущий сенсорный поток, как тот, который он видел ранее и вспомнит сохраненные паттерны, представляющие то, что происходило в прошлом. По сути, это позволяет вам увидеть кратчайший путь в будущем. Если б вы были говорящей крысой, вы могли бы сказать, «О, я узнала этот лабиринт, и я помню этот угол». Как только неокортекс вспомнит, что происходило в прошлом, вы вообразите нахождение сыра, который вы видели в прошлый раз, когда вы были в этом лабиринте, и как до него добраться. «Если я здесь поверну направо, я знаю, что произойдет потом. В конце этого коридора есть кусочек сыра. Я вижу его в своем воображении». Когда вы несетесь через лабиринт, вы полагаетесь на старые, примитивные структуры, чтоб выполнять движения, подобные поднятию лап и шевеления усами. С вашим (относительно) большим неокортексом, вы можете вспомнить места, в которых вы были, снова узнать их в будущем и предсказать, что произойдет потом. Ящерица без неокортекса обладает более скудными возможностями вспомнить прошлое, и будет обыскивать лабиринт заново каждый раз. Вы (крыса) понимаете мир и ближайшее будущее, потому что обладаете кортикальной памятью. Вы видите живые картины получения вознаграждения и опасности, которые последуют после принятия каждого решения, и таким образом вы движетесь по миру более эффективно. Вы буквально видите будущее.

Но заметьте, вы не проявляете никакого особенного или фундаментально нового поведения. Вы не строите себе дельтаплан и не летите к сыру в конце коридора. Ваш неокортекс формирует предсказания о сенсорном потоке паттернов, что позволяет вам увидеть будущее, но ваша палитра доступных видов поведения практически не затрагивается. Ваша способность к бегу, лазанью и исследованию в основном все еще как у ящерицы.

Когда кортекс по мере эволюции становится больше, он начинает помнить о мире все больше и больше. Он может сформировать больше воспоминаний и сделать больше предсказаний. Сложность этих воспоминаний и предсказаний также увеличивается. Но что-то другое замечательное происходит, что ведет к исключительно человеческим способностям интеллектуального поведения.

Человеческое поведение выходит за рамки старого базового репертуара движений уровня крысы. Неокортикальная эволюция вывела нас на новый уровень. Только люди создали письменный и устный язык. Только люди готовят еду, шьют одежду, летают на самолетах и строят небоскребы. Наши двигательные способности и способности к планированию значительно превысили способности наших ближайших родственников среди животных. Как кортекс, который был создан для предсказаний, может генерировать невероятно сложное поведение только у человека? И как это совершенное поведение эволюционировало так внезапно? Есть два ответа на этот вопрос. Один заключается в том, что неокортикальный алгоритм настолько мощный и гибкий, что в результате небольшой перекоммутации, уникальной для человека, он смог создавать новое, усложненное поведение. Другой ответ заключается в том, что поведение и предсказание две стороны одной медали. Хотя кортекс может воображать будущее, он может делать точные сенсорные предсказания только в том случае, если знает, какое поведение будет сгенерировано.

В простейшем примере с крысой, ищущей сыр, крыса помнила лабиринт и использовала память для предсказания того, что она увидит сыр за углом. Хотя крыса могла бы повернуть налево или направо, только одновременные память о сыре и правильное поведение, «повернуть направо на развилке», могли бы заставить крысу сделать предсказания о сыре истинным. Хотя это тривиальный пример, он дает понимание того, насколько глубока связь сенсорных предсказаний и поведения. Любое поведение влияет на то, что мы видим, слышим и чувствуем. Большинство из того, что мы ощущаем в данный момент, сильно зависит от наших действий. Поместите вашу руку перед лицом. Чтоб предсказать, что вы увидите руку, ваш кортекс должен знать, что он дал руке команду двигаться. Если кортекс увидит вашу руку движущейся без соответствующих моторных команд, вы будете удивлены. Простейший путь интерпретировать это – предположить, что ваш мозг сначала двигает руку и только потом предсказывает, что он должен увидеть. Я считаю, что это неверно. Наоборот, я думаю, что кортекс сначала предсказывает увидеть руку, и это предсказание именно то, что вызывает моторные команды, делающие предсказание истинным. Сначала вы думаете, и это заставляет вас действовать, чтоб сделать ваши мысли правдой.

Сейчас мы хотим взглянуть на изменения, которые привели к тому, что человек обладает существенно расширенным репертуаром поведения. Существуют ли физические различия между кортексом обезьяны и человека, которые могли бы объяснить, почему только у человека есть речь и другое сложное поведение? Человеческий мозг примерно в три раза больше чем мозг шимпанзе. Но есть нечто большее, чем «больше - значит лучше». Ключ к пониманию к скачку человеческого поведения находится в соединениях между областями кортекса и частями старого мозга. Проще говоря, наш мозг и мозг шимпанзе отличаются соединениями.

Давайте взглянем поближе. Каждый знает, что мозг разделяется на левое и правое полушарие. Но есть также и другое разделение, которое менее известно, и именно там мы должны искать отличия человека. В любом мозге, особенно в большом, есть разделение кортекса на переднюю и заднюю половину. Ученые используют термин *антериорная* для передней и *постериорная* для задней. Разделением передней и задней половины является большая борозда, называемая «central sulcus». Задняя часть кортекса содержит сектора, куда приходит визуальная, слуховая и осязательная информация. Именно там в основном возникает сенсорное восприятие. Передняя часть содержит области кортекса, которые задействованы в высокоуровневом планировании и мышлении. Она также содержит моторный кортекс, отдел мозга, наиболее ответственный за движение мускулов и, следовательно, формирование поведения.

Когда неокортекс приматов стал со временем больше, антериорная половина стала непропорционально большой, особенно у человека. По сравнению с другими приматами и ранними гоминидами, у нас громадный лоб, предназначенный для того, чтоб вмещать наш очень большой антериорный кортекс. Но только этого увеличения недостаточно для объяснения развития наших моторных способностей по сравнению с другими существами. Наша способность производить исключительно сложные движения происходит из того факта, что наш моторный кортекс имеет гораздо больше соединений с мышцами нашего тела. У других млекопитающих передний кортекс играет менее ведущую роль в моторном поведении. Большинство животных полагаются в основном на старые части мозга для генерации их поведения. В отличие от этого, кортекс человека узурпировал большую часть моторного контроля у других частей мозга. Если вы повредите моторный кортекс крысы, крыса не получит какого либо заметного ущерба. Если повредить моторный кортекс человека, он станет парализованным.

Люди часто спрашивают меня о дельфинах. Большой ли у них мозг? Ответ - да; у дельфина большой неокортекс. Кортекс дельфина имеет более простую структуру, чем человеческий неокортекс (три слоя вместо наших шести), но по другим параметрам он больше. Похоже, что дельфин может помнить и понимать много вещей. Он может индивидуально распознавать других дельфинов. Возможно у него блестящая память о его собственной жизни, в автобиографическом смысле. Возможно он знает каждый уголок и каждую щелку в океане, где он когда-либо был. Но хотя он демонстрирует довольно сложное поведение, дельфины не близки к нам. Так что мы можем предполагать, что их кортекс имеет менее существенное влияние на их поведение. Суть в том, что кортекс эволюционировал в основном для обеспечения памяти о мире. Животные с большим кортексом могут воспринимать мир почти во многом как вы и я. Но люди единственные, у кого преобладающую, передовую роль в поведении играет кортекс. Вот почему у нас сложный язык и замысловатые приспособления, тогда как у других млекопитающих нет. Вот почему мы можем писать новеллы, бродить в интернете, посылать зонды на Марс и строить прогулочные корабли.

Теперь мы можем увидеть картину целиком. Сначала природа создала животных, таких как рептилии с усложненным восприятием и усложненным, но относительно ригидным поведением. Затем она обнаружила, что путем добавления системы памяти и направления в нее сенсорного потока животное может помнить прошлый опыт. Когда животное обнаруживает себя в той же самой или подобной ситуации, должны всплыть воспоминания, ведущие к предсказанию того, что скорей всего должно произойти потом. Таким образом, интеллект и понимание начались с системы памяти, которая направляла предсказания в сенсорный поток. Эти предсказания являются сутью понимания. Знать что-то обозначает уметь сделать предсказание об этом.

Кортекс эволюционировал в двух направлениях. Во первых, он стал больше и сложнее по типам воспоминаний, которые он может хранить; он стал способен помнить больше вещей и делать предсказания, основанные на более сложных отношениях. Во-вторых, он начал взаимодействовать с моторной системой старого мозга. Для предсказания того, что вы увидите, услышите и почувствуете потом, ему нужно знать, какие действия будут предприняты. В случае людей кортекс перехватил большую часть нашего моторного поведения. Вместо того, чтобы просто делать предсказания на основе поведения старого мозга, человеческий неокортекс управляет поведением, чтобы удовлетворить свои предсказания.

Человеческий кортекс особенно большой и, следовательно, имеет огромную емкость. Он постоянно предсказывает, что вы увидите, услышите и почувствуете в основном таким образом, что вы это не осознаете. Эти предсказания являются нашими мыслями, и, в комбинации с сенсорной информацией, они являются нашим восприятием. Я называю эту точку зрения на мозг моделью «*память-предсказание»*.

Если бы Китайская Комната Серла содержала подобную систему памяти, которая могла бы делать предсказания о том, какие китайские символы должны появиться следующими и что потом должно происходить в рассказе, мы могли бы с уверенностью сказать, что комната понимает китайский язык и понимает рассказ. Теперь мы можем видеть, в чем ошибался Алан Тьюринг. Предсказание, а не поведение является доказательством интеллекта.

Сейчас мы готовы вдаться в детали этой новой модели «память-предсказание». Для того, чтоб делать предсказания будущих событий, ваш неокортекс должен хранить последовательности паттернов. Чтобы вызвать соответствующие воспоминания, необходимо затребовать паттерны по их сходству с прошлыми паттернами (автоассоциативная память). И, наконец, воспоминания должны быть сохранены в инвариантной форме, так чтобы знания о прошлых событиях могли бы быть применены к новой ситуации, которая похожа, но не идентична прошлой. То, как кортекс разрешает эти задачи, а также более полное исследование его иерархии, является темой следующей главы.

# 6. Как работает кортекс

**Попытки** понять, как работает мозг, подобны решению гигантской головоломки из кусочков картинки. Вы можете подходить к ее решению двумя способами. Используя подход «сверху вниз», вы начинаете с того, что берете цельную картинку, и, пользуясь ей, решаете, какие кусочки игнорировать, а какие искать. Другой подход, «снизу вверх», когда вы фокусируетесь непосредственно на самих кусочках. Вы изучаете их на наличие особенностей и ищите наиболее подходящие среди других кусочков. Если у вас нет цельной картинки-решения, метод «снизу вверх» является иногда единственным путем.

Головоломка «познай мозг» особенно устрашающая. В отсутствии хорошего теоретического обоснования для понимания интеллекта ученые стараются придерживаться подхода «снизу вверх». Но задача очень трудна, если вообще не неосуществима, если головоломка такая сложная, как мозг. Чтобы ощутить трудность, вообразите головоломку с несколькими тысячами кусочков. Большинство кусочков могут быть интерпретированы различными путями, как если бы у каждого была подходящая картинка на обоих сторонах, но только одна из них правильная. Все кусочки почти одинаковы по форме, так что вы не смогли бы определенно сказать, подходят два кусочка или нет. Большинство из них не будет использовано в конечном решении, но вы не знаете, какие и сколько. Каждый месяц новые кусочки приходят по почте. Некоторые из этих новых кусочков заменяли бы более старые, как если бы разработчик головоломки сказал, «Я знаю, что вы работали с этими старыми кусочками несколько лет, но оказалось, что они не годятся. Извините. Вместо них используйте эти новые кусочки до будущих извещений». К сожалению, у вас нет идей, на что будет похож конечный результат; еще хуже, если у вас были идеи, но они оказались неправильными.

Эта аналогия с головоломкой является великолепным описанием сложности, с которой мы сталкиваемся при создании новой теории кортекса и интеллекта. Кусочки головоломки – это биологические и поведенческие данные, которые ученые собрали за сотни лет. Каждый месяц публикуются новые документы, создающие дополнительные кусочки головоломки. Иногда данные одного ученого противоречат данным другого. Поскольку данные могут быть интерпретированы различными способами, практически во всем есть разногласия. Без теоретических оснований «сверху вниз» не будет консенсуса в том, что искать, что наиболее важно или как интерпретировать горы накопленной информации. Наше понимание мозга застряло на подходе «снизу вверх». Все, что нам нужно – это теоретические обоснования для подхода «сверху вниз».

Модель «память-предсказание» может выступить в этой роли. Она может показать нам, как начинать складывать вместе кусочки головоломки. Чтобы делать предсказания, вашему кортексу нужен способ помнить и хранить знания о последовательностях событий. Чтобы делать предсказания новых событий, кортекс должен сформировать инвариантные представления. Вашему мозгу нужно создавать и хранить модель мира такого, какой он есть, независимо от того, как вы видите его в различных обстоятельствах. Зная, что должен делать кортекс, ведет нас к пониманию его архитектуры, особенно его иерархического дизайна и шестислойной формы.

Когда мы изучим эти теоретические основы, представленные здесь впервые, я перейду на детальный уровень, который может быть многообещающим для некоторых читателей. Большинство концепций, с которыми вы сейчас встретитесь, непривычны даже для экспертов в нейронауках. Но я верю, с небольшим усилием каждый сможет понять фундамент этих теоретических основ. Главы 7 и 8 этой книги гораздо менее технические и более широко исследуют следствия теории.

Наше решение головоломки может теперь обернуться к поиску биологических деталей, которые подтверждают гипотезу «память-предсказание»; это подобно тому, что мы оставим в стороне большой процент кусочков головоломки, зная, что относительно небольшой процент оставшихся кусочков приоткроет решение. Как только мы обнаружим, что ищем, задача станет управляемой.

В то же время я хочу заметить, что эти теоретические основания еще не завершены. Есть множество вещей, которые мне еще непонятны. Но многое я уже сделал, основываясь на дедуктивном выводе, экспериментах, выполненных во множестве различных лабораторий, и знании анатомии. Последние 5 – 10 лет исследователи из множества подобластей нейронаук исследовали идеи, подобные моим, хотя использовали другую терминологию и, насколько мне известно, не пытались сложить эти идеи в одну общую теорию. Они говорят об обработке «сверху вниз» и «снизу вверх», как паттерны распространяются через сенсорные области мозга и как важно инвариантное представление. Например, Габриэль Крейман и Кристоф Кох, нейрофизиологи из Кальтеха, совместно с нейрохирургом Ицхаком Фрейдом из UCLA, обнаружили клетки, которые возбуждаются, когда человек видит изображение Билла Клинтона. Одна из моих целей – объяснить, как образуются эти клетки Билла Клинтона. Конечно, все теории должны делать предсказания, которые могут быть протестированы в лаборатории. Я укажу несколько таких предсказаний в приложении. Сейчас мы знаем, что искать, и очень сложная система больше не выглядит такой сложной.

В следующих разделах этой главы мы глубже и глубже будем зондировать то, как работает модель «память-предсказание». Мы начнем с широкомасштабной структуры и широкомасштабных функций неокортекса, и будем двигаться к предстоящему пониманию более мелких кусочков и того, как они складываются в картинку.



**Рисунок 1. Первые четыре области визуального распознавания объектов.**

## 6.1. Инвариантное представление

Ранее я изобразил кортекс как слой клеток размером с обеденную салфетку, такой же толщины, как шесть визиток, где соединения между различными областями задают в целом иерархическую структуру. Сейчас я хочу нарисовать другую картину кортекса, которая высветит ее иерархические соединения. Вообразите, что мы разрезали обеденную салфетку на функциональные области – секции кортекса, которые специализируются на определенных задачах – и сложили эти области одна на другую подобно блинчикам. Если вы разрежете эту стопку и посмотрите со стороны, вы увидите рисунок1. Кортекс на самом деле не похож на это, как вы могли бы подумать, но картинка поможет вам увидеть, как проходит информация. Я показал четыре кортикальных области, в которые снизу поступает сенсорная информация и течет вверх от области к области. Заметьте, информация ходит в обоих направлениях.

Рисунок 1 представляет четыре визуальных области, задействованных в распознавании объектов – то, как вы видите и узнаете кошку, храм, вашу маму, Великую Китайскую Стену. Биологи обозначают их V1, V2, V4 и IT. Визуальная информация, представленная направленными вверх стрелками внизу рисунка 1, возникает в сетчатке обоих глаз и идет в V1. эта информация может рассматриваться как постоянно меняющиеся паттерны, распределенные приблизительно по миллиону аксонов, связанных вместе в оптический нерв.

Мы говорили ранее о пространственных и временных паттернах, но имеет смысл освежить вашу память, поскольку мы будем ссылаться на них очень часто. Вспомните, что ваш кортекс – это большой слой нервной ткани, который содержит функциональные области, специализированные на определенных задачах. Эти области соединяются большими связками аксонов или волокон, которые передают информацию от одного региона к другому, все одновременно. В любой момент времени некоторое множество волокон возбуждается электрическим импульсом, называемым потенциалом действия или спайком, тогда как другие остаются неактивными. Коллективная активность связки волокон и есть то, что обозначает *паттерн*. Паттерн, поступающий в V1, может быть пространственным, когда ваш взгляд задерживается на объекте, и временными, когда ваш взгляд движется по объекту.

Как отмечалось ранее, примерно три раза в секунду ваши глаза совершают быстрое движение, называемое саккадой, и остановку, называемую фиксацией. Если ученый подключит устройство, отслеживающее движение глаз, вы будете удивлены, какими отрывистыми являются саккады, хотя ваше визуальное ощущение непрерывно и стабильно. Рисунок 2 показывает, как у некоторого человека движутся глаза, когда он смотрит на лицо. Заметьте, что фиксации не произвольны. Теперь вообразите, что вы могли бы видеть паттерн активности, поступающие в V1 от глаз этого человека. Он меняется постоянно с каждой саккадой. Несколько раз в секунду кортекс видит совершенно новый паттерн.

Вы могли бы подумать, «хорошо, но это все еще то же самое лицо, просто смещающееся». В этом есть доля правды, но не так много, как вы думаете. Светочувствительные рецепторы в вашей сетчатке распределены неравномерно. Они плотно сконцентрированы в фовеальной области в центре, и постепенно редеют к периферии. В отличие от этого клетки кортекса распределены равномерно. В результате изображение с сетчатки, отображаемое в первичную визуальную область V1, сильно искажено. Когда ваши глаза фиксируются на носу, а не на глазу того же самого лица, картинка значительно отличается, как если бы ее рассматривали через искажающие линзы, которые постоянно дергаются туда-сюда. Но когда вы видите лицо, оно не кажется вам искаженным, и не кажется прыгающим. Большую часть времени вы даже не осознаете, что паттерны с сетчатки полностью изменяются. Вы видите «просто лицо». (Рисунок 2б показывает этот эффект на примере берегового ландшафта). Это подтверждение загадки инвариантного представления, о котором мы говорили в главе 4. То, что вы воспринимаете – это не то, что видит V1. Как же все таки ваш мозг узнает, что он видит одно и то же лицо, и почему вы не знаете, что поступающая информация изменяющаяся и искаженная?



**Рисунок 2а. Как глаза совершают саккады по человеческому лицу. Рисунок 2б. Искажение, вызванное неравномерным распределением рецепторов по сетчатке.**

Если мы поместим электроды в V1 и будем наблюдать, как отвечают отдельные клетки, мы обнаружим, что каждая конкретная клетка возбуждается только в ответ на визуальную информацию от крошечной части сетчатки. Этот эксперимент был проделан много раз и является опорным в исследовании зрения. Каждый нейрон в области V1 имеет так называемое рецептивное поле, которое сильно специфично для каждой мельчайшей части общего поля зрения – то есть, цельного мира перед вашими глазами. Представляется, что клетки в V1 совсем не знают о лицах, машинах, книгах или других значительных объектах, которые вы видите все время; они «знают» о крошечных, с игольное ушко, порциях визуального мира.

Каждая клетка в V1 также настроены на специфические виды поступающих паттернов. Например, конкретная клетка может активно пульсировать, когда она видит линию или край, наклоненный под углом в 30 градусов. Эти края сами по себе имеют небольшое значение. Они могли бы быть частью любого объекта – половицы, стволом отдаленного пальмового дерева, стороной буквы М или одной из почти бесконечного числа возможностей. При каждой новой фиксации, рецептивное поле клетки попадает на новую и совершенно отличную порцию визуального пространства. При некоторых фиксациях клетка будет сильно возбуждаться, на других будет возбуждаться слабо или вообще не будет. Таким образом, каждый раз, когда вы совершаете саккаду, множество клеток в V1 вероятнее всего изменяет свою активность.

Однако, нечто волшебное происходит, если вы помещаете электрод в верхнюю область, показанную на рисунке 1, область IT. Здесь мы обнаруживаем некоторые клетки, которые становятся и остаются активными, когда объект полностью появляется где-нибудь в поле зрения. Например, мы могли бы найти клетки, которые возбуждаются только тогда, когда видно лицо. Эти клетки остаются активными до тех пор, пока ваши глаза видят лицо где-нибудь в поле вашего зрения. Они не включаются и не выключаются при каждой саккаде, как это делают клетки в V1. Рецептивное поле этих клеток в IT покрывает большую часть визуального пространства и настроено на возбуждение, когда видно лицо.

Давайте откроем тайну. Походу охвата четырех кортикальных этапов от сетчатки до IT, клетки изменяются от быстро изменяющихся, пространственно специфичных, распознающих крошечные кусочки ячеек, до постоянно возбужденных, пространственно неспецифичных, распознающих объекты. Клетки в IT говорят нам, что мы видим лицо где-то в поле нашего зрения. Эти клетки, называемые обычно нейронами лица, будут возбуждаться независимо от того, наклонено ли лицо, повернуто ли, или частично загорожено. Это часть инвариантного представления для «лиц».

Написать эти слова кажется так просто. Четыре коротких этапа, и Вуаля, мы узнали лицо. Ни одна компьютерная программа или математическая формула не решает эту задачу с надежностью и общностью, близкой к человеческому мозгу. Но мы знаем, что мозг решает ее за несколько шагов, так что ответ не может быть сложным. Одна из основных целей этой главы объяснить, как получаются нейроны лица, нейроны Билла Клинтона или другие. Мы доберемся до этого, но мы должны охватить сначала много другого.

Взглянем на рисунок 1 по-другому. Вы видите, что информация также течет от высших областей к низшим через сеть обратных связей. Эти связки аксонов, которые идут от областей вроде IT к низшим областям вроде V4, V2 и V1. Более того, обратных связей много, если не больше, чем прямых.

Много лет ученые игнорировали обратные связи. Если ваше понимание мозга сфокусировано на том, как кортекс принимает информацию, обрабатывает ее и затем действует на ее основе, вам не нужны обратные связи. Все что вам нужно – это прямые соединения, ведущие от сенсорных областей кортекса к моторным. Но когда вы начинаете понимать, что функция кортекса – предсказание, то вам необходимо ввести в модель обратные связи; мозг должен посылать поступающую информацию обратно к областям, которые получили информацию первыми. Предсказание требует сравнения того, что происходит и того, что вы ожидаете. То, что действительно происходит идет вверх, то, что вы ожидаете идет вниз.

Те же самые прямые и обратные процессы возникают во всех областях кортекса, задействованных во всех органах чувств. На рисунке 3 рядом с визуальной стопкой блинчиков изображены похожие стопки для слуха и осязания. Там также изображена чуть более высшая кортикальная область, ассоциативная, которая получает и интегрирует информацию от нескольких различных органов чувств. Тогда как рисунок 1 основан на знании соединений между четырьмя известными областями кортекса, рисунок 3 чисто концептуальная диаграмма, не пытающаяся охватить действительные кортикальные области. В реальном мозгу человека масса кортикальных областей соединены различными способами. Фактически, большая часть человеческого кортекса состоит из ассоциативных областей. Анимированная характеристика, показанная здесь и на следующих рисунках, предназначена для того, чтобы помочь вам понять, что происходит, не вводя сильно в заблуждение.



**Рисунок 3. Формирование инвариантного представления для слуха, зрения и осязания.**

Трансформация – от быстро изменяющихся к медленно меняющимся и от пространственно специфичных к пространственно инвариантным – очень хорошо изучена для зрения. И хотя это не так очевидно и требует доказательств, многие нейрофизиологи верят, что то же самое происходит во всех сенсорных областях кортекса, не только в визуальных.

Возьмем слух. Когда кто-то разговаривает с вами, изменения в звуковом давлении происходят очень быстро; паттерны, поступающие в первичную слуховую область, называемую A1, изменяются очень быстро. Но если мы могли бы поместить электроды чуть выше по слуховому потоку, мы нашли бы инвариантные клетки, которые отвечают на слова или даже на фразы. Ваш слуховой кортекс мог бы иметь группу клеток, которые возбуждаются, когда вы слышите «спасибо» и другую группу клеток, возбуждающуюся на фразу «доброе утро». Такие клетки должны оставались бы активными, в течение всего высказывания, полагая, что вы распознали фразу.

Паттерны, получаемые первой слуховой областью могут изменяться очень широко. Слово может быть произнесено с различным акцентом, на различной высоте или с различной скоростью. Но чем выше по кортикальной иерархии, тем менее значимыми становятся низкоуровневые особенности; слово есть слово, несмотря на акустические детали. То же самое верно и для музыки. Вы можете услышать «Three Blind Mice», сыгранное на пианино, на кларнете или спетое ребенком, и ваш A1 будет получать совершенно различные паттерны в каждом случае. Но электрод, помещенный в высшие слуховые области, должен обнаружить клетки, которые монотонно возбуждаются каждый раз, когда играют «Three Blind Mice», не зависимо от инструмента, темпа или других деталей. Такой конкретный эксперимент не был проведен, конечно, потому что он требует слишком больших вмешательств для человека, но если вы согласны, что должен существовать общий кортикальный алгоритм, вы можете быть уверены, что такие клетки существуют. Мы видим, что в слуховом кортексе тот же самый вид обратных связей, предсказания и инвариантного вспоминания, что и в визуальном.

Наконец, осязание должно вести себя точно также. Опять же, конкретные эксперименты не были проведены, хотя полным ходом идут исследования на обезьянах с помощью аппаратуры, отображающей мозг с высоким разрешением. Поскольку сейчас я сижу и пишу, у меня в руке авторучка. Я трогаю колпачок авторучки, и мои пальцы поглаживают его металлический держатель. Паттерны, поступающие в мой соматосенсорный кортекс от сенсорных рецепторов моей кожи, постоянно изменяются, пока мои пальцы двигаются, но у меня постоянное ощущение авторучки. В один момент я могу согнуть металлический держатель пальцами, в другой момент я это сделаю другим набором пальцев или вообще губами. Много информации, поступающей из различных в соматосенсорный кортекс. Однако, наш электрод снова должен найти клетки в областях, удаленных на несколько шагов от первичной, которые инвариантно отвечают на «авторучку». Они должны оставаться активными, пока я поглаживаю авторучку, и им должно быть все равно, от каких именно пальцев или частей моего тела я дотрагиваюсь до нее.

Подумайте над этим. Слухом или осязанием вы не можете опознать объект с одномоментного сенсорного потока. Паттерны, поступающие от ушей или рецепторов кожи содержат недостаточно информации в каждый конкретный момент времени, чтоб сказать вам, что вы слышите или чувствуете. Когда вы воспринимаете серию слуховых паттернов, такую как мелодия, произнесенное слово или хлопающую дверь, или когда вы тактильно ощущаете объект, такой как авторучка, единственный способ сделать это – использовать поток информации во времени. Вы не можете узнать мелодию, услышав одну ноту, вы не можете узнать ощущение авторучки одним прикосновением. Таким образом, нейронная активность, соответствующая ментальному восприятию объекта, такого как произнесенное слово, должно длиться по времени дольше, чем отдельный паттерн. Это просто другой способ прийти к тому же самому выводу, что чем выше область кортекса в иерархии, тем меньше изменений по времени вы должны видеть.

Зрение также базируется на потоках информации во времени и работает тем же самым общим образом, как слух или осязание, но поскольку мы способны узнавать индивидуальный объект за одну фиксацию, оно портит общую картину. Несомненно, эта способность распознавать пространственные паттерны за короткое время фиксации многие годы сбивала с пути исследователей, работавших над машинным зрением. Они в основном игнорировали критическое значение времени. Хотя можно в лабораторных условиях заставить человека узнавать объекты без движения глаз, это не является нормой. Нормальное зрение, такое как чтение этой книги, требует постоянного движения глаз.

## 6.2. Интеграция чувств

Что же насчет ассоциативных областей? До сих пор мы видели, как информационные потоки идут вверх и вниз по конкретной сенсорной области кортекса. Потоки, идущие вниз, замещают поступающую информацию и делают предсказание о том, что мы ощутим далее. Те же самые процессы возникают между различными чувствами – то есть, между зрением, слухом и осязанием. Например, что-то, что я слышу, может привести к предсказанию того, что я должен увидеть или почувствовать. Сейчас я пишу в моей спальне. У нашей кошки Кео есть ошейник, который позвякивает, когда она ходит. Я слышу ее приближающееся позвякивание из коридора. По этой звуковой информации я узнаю мою кошку, поворачиваю голову к коридору и входящей Кео. Я ожидаю увидеть ее на основании ее звука. Если Кео не войдет или появится другое животное, я буду удивлен. В этом примере, звуковая информация сначала привела к слуховому узнаванию Кео. Информация поднялась по слуховой иерархии к ассоциативной области и соединила зрение и слух. Затем образ спустился обратно по слуховой и зрительной иерархии, ведя и к слуховому, и к зрительному предсказанию. Рисунок 4 иллюстрирует это.

Такого рода мультисенсорные предсказания возникают все время. Я сгибаю держатель моей ручки, я чувствую, как держатель соскальзывает с пальцев и я ожидаю услышать щелкающий звук, когда держатель ударится о корпус ручки. Если я не слышу щелчка, следующего за отпусканием держателя, я буду удивлен. Мой мозг точно предсказывает, когда я услышу звук и на что он должен быть похож. Чтоб происходило такое предсказание, информация поднимается через соматосенсорный кортекс и спускается обратно и в соматосенсорный, и в слуховой кортекс, ведя к предсказанию звука и ощущения щелчка.

Другой пример: несколько дней я ездил на работу на велосипеде. В то утро я шел в гараж, брал велосипед и выкатывал его на подъездную дорожку. В процессе этого, я получал множество визуальных, тактильных и слуховых ощущений. Велосипед ударялся о дверной косяк, цепь трещала, педали ударяли по моим ногам, и колеса вращались, когда терлись об пол. В процессе выноса велосипеда из гаража мой мозг встречался с плотиной зрительных, слуховых и тактильных ощущений. Каждый сенсорный поток делает предсказания для других весьма скоординированным образом. То, что я вижу, ведет к точному предсказанию того, что я почувствую и услышу, другие чувства примерно также. Вид велосипеда, ударяющегося об косяк заставляет меня ожидать услышать определенный звук и почувствовать подскакивание велосипеда. Ощущение педалей, ударяющихся о мои ноги, заставляет меня взглянуть вниз и увидеть педали там, где я их почувствовал. Предсказания настолько точны, что я должен сразу заметить, если одно из этих ощущений будет немного не скоординированным или необычным. Информация одновременно течет вверх и вниз по сенсорной иерархии, чтобы создать единый сенсорный опыт, включающий предсказание по всем органам чувств.



**Рисунок 4. Информация течет вверх и вниз по сенсорной иерархии, чтобы сформировать предсказания и создать единый сенсорный опыт.**

Проведите такой эксперимент. Прекратите читать и сделайте что-нибудь, любое движение, которое включает движение вашего тела и манипулирование объектами. Например, подойдите к раковине и откройте кран. Теперь, когда вы сделали это, попробуйте заметить каждый звук, прикосновение и изменение в визуальной информации. Вы должны сконцентрироваться. Каждое действие глубоко связано со зрением, слухом и тактильными ощущениями. Поднимите или поверните кран, и ваш мозг будет ожидать почувствовать давление на вашу кожу и сопротивление ваших мышц. Вы ожидаете увидеть и почувствовать движение рукоятки крана и услышать звук воды в кране. Когда вода ударится о раковину, вы ожидаете услышать другой звук и увидеть и почувствовать брызги.

Каждый шаг создает звук, который вы предвидите осознанно или нет. Даже простейшее действие - держание книги – ведет к многочисленным сенсорным предсказаниям. Вообразите, что вы почувствовали и услышали закрытую книгу, но визуально она остается открытой. Вы должны быть шокированы и сбиты с толку. Как мы видели в эксперименте с измененной дверью в главе 5, вы постоянно делаете предсказания о мире, которые скоординированы по всем вашим чувствам. Когда я концентрируюсь на всех мелких ощущениях, я поражаюсь, как сильно интегрированы наши сенсорные предсказания. Хотя эти предсказания могут показаться простыми или тривиальными, обратите внимание, какие они всепроникающие и как только они могут происходить с такой координацией паттернов, следующих вверх и вниз по кортикальной иерархии.

Когда вы поймете, как взаимосвязаны чувства, вы придете к выводу, что неокортекс, все сенсорные и ассоциативные области работают как одно целое. Да, у нас есть визуальный кортекс, но он просто один из компонентов одной общей сенсорной системы – изображения, звуки, прикосновения и другие чувства, скомбинированные, текут вверх и вниз по единой иерархии со многими ветвями.

Следующий факт: все предсказания обретаются на опыте. Мы ожидаем в настоящем и в будущем, что держатель авторучки будет издавать щелкающий звук, потому что он так делал в прошлом. Велосипеды, ударяющиеся в гаражах, выглядят, чувствуются и звучат для нас предсказуемым образом. Вы не родились с этими знаниями; вы обрели их благодаря невероятной способности вашего кортекса помнить паттерны. Если для поступающего в ваш мозг паттерна есть соответствующий паттерн, ваш мозг использует его, чтобы предсказать будущие события.

Хотя рисунки 3 и 4 не отображают моторный кортекс, вы можете вообразить его как еще одну стопку блинчиков, также как и сенсорную стопку, подсоединенную к сенсорной системе через ассоциативные области (хотя с более тесными соединениями с соматосенсорным кортексом для выполнения движений тела). В этом смысле моторный кортекс ведет себя почти так же, как и сенсорные области. Информация из любой сенсорной области может подниматься к ассоциативным областям, что может вызвать паттерн, попадающий в моторный кортекс и приводящий к поведению. Точно так же, как визуальная информация может вызвать паттерны, идущие к слуховым и сенсорным областям, она может вызвать и паттерны, идущие к моторному кортексу. В первом случае мы интерпретируем эти потоки вверх-вниз как предсказание. В случае с моторным кортексом – как моторные команды. Как указал Монткастл, моторный кортекс выглядит точно так же, как и сенсорный. Следовательно, как кортекс обрабатывает возвращающиеся сенсорные предсказания, похожим образом обрабатывает и моторные команды.

Вскоре мы увидим, что в кортексе нет чисто моторных или чисто сенсорных областей. Сенсорные паттерны одновременно текут и там и там – и затем возвращаются по всем областям иерархии, ведя к предсказанию или моторному поведению. Хотя у моторного кортекса есть некоторые специальные атрибуты, о нем можно думать всего лишь как о части одной большой системы «память-предсказание». Он практически похож на другие органы чувств. Зрение, слух, осязание и поведение глубоко переплетены.

## 6.3. Новая точка зрения на V1

Следующий шаг в понимании архитектуры кортекса требует взглянуть на кортикальные области по-другому. Мы знаем, что высшие области кортикальной иерархии формируют инвариантное представление. Но почему эта важная функция должна возникать только наверху? Держа на задворках мысли замечание Монткастла о симметрии, я начал изучать различные способы, которыми могли бы соединяться кортикальные области.

Рисунок 1 изображает четыре классических области визуального пути, V1, V2, V4 и IT, где V1 внизу, на нем V2, V4 и на самом верху IT. Каждый из них условно рассматривается и изображен как единая, непрерывная область. Таким образом все клетки V1 предположительно делают одно и то же, хотя в различных частях визуального поля. Все клетки V2 решают задачу подобного типа. Все клетки V4 специализированы подобным же образом.

С традиционной точки зрения, когда изображение лица поступает в область V1, клетки в нем создают грубый набросок лица в терминах простых линейных сегментов и других элементарных деталей. Этот набросок поступает в V2. Затем V2 делает свое дело с изображением, производя более сложный анализ черт лица, и передает результат в V4, и т.д. Инвариантность и распознавание объекта достигается только тогда, когда информация достигает верхней точки, IT.

К несчастью, с такой точкой зрения на V1, V2 и V4 есть некоторые проблемы. Почему инвариантное представление должно возникать только в IT? Если все кортикальные области выполняют одну и ту же функцию, почему IT Должна быть особенной?

Во-вторых, лицо может появиться на левой стороне вашего V1 или на правой, и вы должны узнать его. Но эксперименты ясно показывают, что несмежные колонки V1 не имеют прямого соединения; левая сторона V1 не может знать, что видит правая. Отступите и подумайте над этим. Различные части V1 явно занимаются похожими вещами, так как все они участвуют в распознавании лица, но в то же время они физически независимы. Подобласти или кластеры V1 физически разъединены, но делают одно и то же.

В конечном счете эксперименты показывают, что все высшие области кортекса получают информацию, сходящуюся от двух или более сенсорных областей ниже по иерархии (рисунок 3). В настоящем мозге десятки областей могут сходиться к ассоциативной области. Но в традиционных интерпретациях нижние сенсорные области, наподобие V1, V2 и V4 имеют различные виды соединений. Каждая рассматривается, как если бы у нее был только один вход – только одна стрелка, идущая снизу – без явного схождения информации от других регионов. V2 получает информацию только от V1 и только. Почему некоторые кортикальные области получают сходящуюся информацию, а другие – нет? Это также несовместимо с идеей Монткастла о едином кортикальном алгоритме.

По этим и другим причинам я пришел к уверенности, что V1, V2 и V4 не должны рассматриваться как единые кортикальные области. Наоборот, каждая является набором множества мелких подобластей. Давайте вернемся к аналогии с обеденной салфеткой – плоской версией кортекса. Давайте воспользуемся авторучкой для разметки всех функциональных областей кортекса на нашей кортикальной салфетке. Наибольшей областью безоговорочно является V1, первичная визуальная область. Следующей была бы V2. Они огромны по сравнению с большинством других областей. Я полагаю, что V1 в действительности должна рассматриваться как множество очень маленьких областей. Вместо одной большой области на салфетке мы нарисовали бы множество маленьких областей, которые все вместе занимали бы область, предназначенную для V1. Другими словами, V1 состоит из нескольких отдельных маленьких кортикальных областей, которые не соединяются со своими соседями напрямую, а только через выше или ниже по иерархии. V1 имела бы наибольшее количество подобластей из всех визуальных областей. V2 также состояла бы из меньшего количества подобластей чуть большего размера. Это же было бы верным и для V4. Но со временем, когда вы доберетесь до области IT, это будет действительно единая область, вот почему у клеток в IT «птицеглазое» видение целого визуального мира.

В этом есть привлекательная симметрия. Давайте взглянем на рисунок 5, на котором показана та же самая иерархия, что и на рисунке 3, за исключением того, что там сенсорная иерархия изображена, как я описал выше. Заметьте, что теперь кортекс везде выглядит одинаковым образом. Возьмите любую область, и вы найдете множество нижних областей, обеспечивающих схождение информации. Принимающие области посылают проекции обратно к входным областям, говоря им, какие паттерны они должны ожидать увидеть далее. Высшие ассоциативные области объединяют информацию от нескольких чувств, таких как зрение или осязание. Нижние области, подобные подобластям V2 объединяют информацию от отдельных подобластей в V1. Области не знают – и, конечно же, не могут знать – что обозначает любой из их входов. Подобластям V2 не нужно знать, что они обрабатывают информацию от нескольких частей V1. Ассоциативным областям не нужно знать, что они обрабатывают информацию от зрения и слуха. Наоборот, цель любой кортикальной области – найти, как соотносятся ее входы, запомнить последовательности корреляций между ними и использовать эту память для предсказания того, как входы поведут себя в будущем. Кортекс есть кортекс. Везде происходит один и тот же процесс: общий кортикальный алгоритм.



**Рисунок 5. Альтернативный взгляд на кортикальную иерархию.**

Это новое иерархическое описание помогает нам понять процесс создания инвариантного представления. Давайте поближе взглянем на то, как это работает в зрительном канале. На первом уровне обработки левая сторона зрительного поля отличается от правой стороны таким же образом, как зрение отличается от слуха. Левый V1 и правый V1 формируют один и тот же вид представления только потому, что на них отображаются подобные паттерны в течение жизни. Как слух и зрение, они могут рассматриваться как отдельные сенсорные потоки, которые объединяются выше.

Подобный образом маленькие области в V2 и V4 являются ассоциативными областями зрения. (подобласти могут перекрываться, но это фундаментально не изменило бы способ работы этих областей). Интерпретация визуального кортекса подобным образом не противоречит и не изменяет что-либо, что нам известно о его анатомии. Информация течет вверх и вниз по всем ветвям иерархического дерева памяти. Паттерны в левом поле зрения могут привести к предсказанию в правом поле зрения тем же путем, как звон колокольчиков моей кошки ведет к визуальному предсказанию того, что она входит в мою спальню.

Наиболее важным результатом этой новой картины кортикальной иерархии является то, что теперь мы можем сказать, что каждая область кортекса формирует инвариантное представление. По-старому у нас не было завершенной картины инвариантного представления – такого, как лица – до тех пор, пока информация не достигала верхнего слоя, IT, который видит цельную картину мира. Теперь мы можем сказать, что инвариантное представление вездесуще. Инвариантное представление формируется в каждой кортикальной области. Инвариантность не является чем то магически проявляющимся, когда мы достигаем верхних областей кортекса, таких как IT. Каждая область формирует инвариантное представление из информации от областей ниже по иерархии. Таким образом, подобласти V4, V2 и V1 создают инвариантное представление на основе того, что поступает в них. Они могут видеть только крошечную часть мира, и словарь сенсорных объектов, с которыми они оперируют, является более простым, но они выполняют ту же самую функцию, что и IT. Также, ассоциативные области выше IT формируют инвариантное представление паттернов от нескольких органов чувств. Таким образом, все области кортекса формируют инвариантное представление мира в изображении нижестоящих областей. В этом есть определенная красота.

Наша головоломка стронулась с места. Мы больше не задаемся вопросом, как формируется инвариантное представление за четыре шага от верха до низа. Вместо этого мы задаемся вопросом, как инвариантное представление формируется в каждой кортикальной области. Это создает ощущение совершенства, если мы всерьез принимаем существование общего кортикального алгоритма. Если одна область хранит последовательность паттернов, то каждая область должна хранить последовательности. Если одна область создает инвариантное представление, но все области создают инвариантное представление. Переосознание кортикальной иерархии подобно тому, как изображено на рисунке 5, делает возможным такую интерпретацию.

## 6.4. Модель мира

Почему неокортекс построен иерархически?

Вы можете думать о мире, перемещаться в нем, делать предсказания будущего, потому что ваш кортекс строит модель мира. Одна из важнейших концепций этой книги – это то, что иерархическая структура кортекса хранит модель иерархической структуры мира. Вложенная структура реального мира отражается вложенной структурой кортекса.

Что я имел в виду под вложенной или иерархической структурой? Подумайте о музыке. Ноты комбинируются, формируя интервалы. Интервалы комбинируются, формируя мелодические фразы. Фразы комбинируются, формируя мелодии или песни. Песни комбинируются в сборники. Подумайте о письменном языке. Буквы комбинируются, формируя слоги. Слоги комбинируются, формируя слова. Слова комбинируются, формируя словосочетания и предложения. Глядя по-другому, подумайте о вашем окружении. Оно возможно состоит из дорог, домов. В домах есть комнаты. В каждой комнате есть стены, потолок, пол, дверь и одно или несколько окон. Каждая из этих частей состоит из еще более мелких объектов. Окна сделаны из стекла, рам, защелок и жалюзи. Защелки сделаны из еще более мелких частей, вроде шурупов.

На минуту взгляните на то, что вас окружает. Паттерны с сетчатки поступают в ваш первичный визуальный кортекс и комбинируются, чтоб сформировать линейные сегменты. Линейные сегменты комбинируются, чтоб сформировать формы. Эти сложные формы комбинируются, чтоб сформировать объекты, такие как нос. Нос комбинируется с глазами и ртом, чтоб сформировать лицо. Лицо комбинируется с другими частями тела, чтоб сформировать человека, который сидит в комнате напротив вас.

Все объекты в мире состоят из субобъектов, которые появляются вместе согласованно; это то самое определение объекта. Когда мы даем чему-то название, мы делаем это, потому что набор свойств согласованно движется вместе. Лицо является лицом именно потому, что два глаза, нос и рот всегда появляются вместе. Глаз является глазом именно потому, что зрачок, радужка, веки и остальное всегда появляются вместе. То же самое может быть сказано о стульях, машинах, деревьях, парках и ландшафте. И, наконец, песня есть песня, потому что серии интервалов всегда появляются вместе в последовательности.

В этом отношении мир похож на песню. Каждый объект в мире состоит из набора мелких объектов, и большинство объектов являются частями больших объектов. Это то, что я называю вложенной структурой. Поняв ее, вы увидите вложенную структуру во всем. В полной аналогии с этим, ваша память о вещах и способ, которым мозг представляет их, сохранены в иерархической структуре кортекса. Ваша память о вашем доме состоит не из одной области кортекса. Она хранится в иерархии кортикальных областей, которые отражают иерархическую структуру дома. Широкомасштабные отношения сохраняются в верхней части иерархии, а мелкомасштабные – в нижней.

Дизайн кортекса и метод, с помощью которого обучается, естественным путем раскрывают иерархические отношения в мире. Вы не родились со знанием языка, домов или музыки. У кортекса хитрый алгоритм обучения, который естественным путем обнаруживает любую существующую иерархическую структуру и захватывает ее. Когда структура отсутствует, мы впадаем в замешательство, даже в хаос.

В один момент времени вы можете ощущать только с подмножеством мира. Вы можете быть только в одной комнате дом, смотреть только в одном направлении. По причине иерархии кортекса, вы способны знать, что вы в доме, в комнате, смотрите в окно, даже хотя в этот момент ваши глаза фиксируются на защелке окна. Высшие области кортекса обрабатывают представление о вашем доме, тогда как нижние области представляют комнату, еще более нижние смотрят на окно. Аналогично, иерархия позволяет вам знать, что вы слушаете и песню и сборник, хотя в любой момент времени вы слышите только одну ноту, которая сама по себе говорит, какая нота будет следующей. Она позволяет вам знать, что вы с вашей лучшей подругой, хотя ваши глаза фиксируются на ее руках. Высшие области кортекса отслеживают большую картинку, тогда как низшие области активно работают с быстро изменяющимися маленькими деталями.

Поскольку мы можем трогать, слышать и видеть только очень маленькую часть мира в любой момент времени, информация, поступающая в мозг, является последовательностями паттернов. Кортексу необходимо изучать эти последовательности, которые возникают снова и снова. В некоторых случаях, таких как мелодия, последовательности паттернов приходят в жестком порядке, порядке интервалов. Большинство из нас знакомо с таким видом последовательностей. Но я собираюсь использовать слово *последовательность* в более общем смысле, ближе к значению математического термина *множество*. Последовательность это множество паттернов, которые обычно следуют один за другим, но не всегда в фиксированном порядке. Что важно, так это то, что паттерны в последовательности следуют один за другим во времени, даже если не в фиксированном порядке.

Некоторые примерны должны прояснить это. Когда я гляжу на лицо, последовательность входных паттернов, которую я вижу, не фиксирована, но определяется моими саккадами. В один момент времени я могу фиксировать в порядке «глаз глаз нос рот», а в другой момент – в порядке «рот глаз нос глаз». Компоненты лица – это последовательность. Они статистически соотносятся и обычно возникают во времени вместе, хотя порядок может меняться. Если вы воспринимаете «лицо», когда фиксируетесь на «нос», то вероятнее всего следующим паттерном будет «глаз» или «рот», но не «авторучка» или «автомобиль».

Каждая область кортекса видит потоки таких паттернов. Если паттерны соотносятся таким образом, что область может научиться предсказывать, какой паттерн будет следующим, кортикальная область формирует постоянное представление, или память для последовательности. Изучение последовательностей это наиболее базовый ингредиент для инвариантного представления объектов реального мира.

Объекты реального мира могут быть конкретными, вроде ящерицы, лица или двери, или они могут быть абстрактными, как слово или теория. Мозг обрабатывает абстрактные и конкретные объекты одинаковым образом. И те и другие просто последовательности паттернов, возникающих вместе во времени предсказуемым образом. Фактически, определенные входные паттерны повторяются, что позволяет кортикальной области знать, что эти образы вызваны реальными объектами мира.

Предсказуемость это определение реальности. Если область кортекса обнаруживает, что она может надежно и предсказуемо двигаться по этим паттернам, используя серии физических движений (таких как саккады глаз или скольжение пальца) и может аккуратно предсказывать их, когда они развертываются во времени (как звуки, составляющие песню или произнесенные слова), мозг интерпретирует их, как имеющие причинно-следственные отношения. Разногласия в многочисленных входных паттернах, возникающие в одном и том же отношении снова и снова без очевидной причинной связи, исчезающе малы. Предсказуемые последовательности паттернов должны быть частью большего реально существующего объекта. Таким образом, надежная предсказуемость является надежным способом узнать, что различные события мира физически связаны вместе. У каждого лица есть глаза, уши, рот и нос. Если мозг видит глаза, затем делает саккаду и видит другой глаз, затем еще одну саккаду и видит рот, он определенно чувствует, что видит лицо.

Если бы кортикальные области могли говорить, они могли бы сказать: «Я ощущаю множество различны паттернов. Иногда я могу предсказать, какой паттерн будет следующим. Но эти паттерны определенно соотносятся один с другим. Они всегда возникают вместе, и я могу надежно перепрыгивать между ними. Таким образом, когда я вижу любое из этих событий, я ссылаюсь на них общим названием. Именно это групповое имя, а не индивидуальные паттерны я передаю в высшие области кортекса».

Таким образом, можно было бы сказать, что мозг хранит последовательности последовательностей. Каждая область кортекса изучает последовательности, вырабатывает то, что я назвал «именами» для последовательностей, которые она знает, и передает эти имена в следующие области выше по кортикальной иерархии.

## 6.5. Последовательности последовательностей

Когда информация движется вверх от первичных сенсорных областей к высшим уровням, мы видим, что они все меньше и меньше изменяются во времени. В первичной визуальной области, типа V1, множество активных клеток быстро изменяется, поскольку новые паттерны падают на сетчатку несколько раз в секунду. В визуальной области IT клетки, возбуждающиеся на паттерны, более стабильны. Что там происходит? Каждая область кортекса имеет репертуар последовательностей, которые ей известны, аналогичные репертуару песен. Области хранят эти песнеподобные последовательности обо всем: звук прибоя, разбивающегося о пляж, лицо вашей матери, путь от дома до ближайшего магазина, как произносить слово «попкорн», как тасовать колоду карт.

У нас есть названия для песен, и подобным образом каждая кортикальная область имеет названия для каждой из последовательностей, которые она знает. Это «имя» - группа клеток, совместное возбуждение которых представляет набор объектов в последовательности. (Не задумывайтесь сейчас о том, как выбираются эти группы клеток; мы придем к этому позже). Эти клетки остаются активными, пока идет последовательность, и именно ее «имя» передается в следующую область по иерархии. Пока поступающие паттерны являются частью предсказуемой последовательности, область выдает постоянно «имя» для следующей области выше по иерархии.

Это как если бы области сказали: «Вот имя последовательности, которую я слышу, вижу или ощущаю. Вам не нужно знать о конкретных нотах, краях или текстурах. Я вам дам знать, если произойдет что-то новое или непредсказуемое». Более конкретно мы можем вообразить область IT на верхушке визуальной иерархии, передающую эстафету ассоциативным областям выше нее: «Я вижу лицо. Да, с каждой саккадой глаза фиксируются на различных частях лица; я вижу различные части лица там, где они должны быть. Но это все еще то же самое лицо. Я дам вам знать, когда увижу что-то другое». Подобным образом предсказуемые последовательности событий становятся идентифицированными «именем» - постоянным паттерном возбужденных клеток. Это происходит снова и снова по мере продвижения по иерархической пирамиде. Одна область могла бы распознать последовательность звуков, которые составляют фонемы, и передает паттерны, представляющие фонемы, вышестоящей области. Вышестоящая область распознает последовательность фонем, чтоб составлять слова. Следующая вышестоящая область распознает последовательности слов, чтоб составить фразы, и т.д. Примите во внимание, что «последовательности» в нижних областях кортекса могут быть очень простыми, такими как визуальное пространственное движение края чего-либо.

Свертыванием предсказуемой последовательности в «именованный объект» в каждой кортикальной области по иерархии мы достигаем все большей и большей стабильности, чем выше мы находимся. Это создает инвариантное представление.

Обратный эффект возникает, когда паттерн движется обратно вниз по иерархии: стабильный паттерн «раскрывается» в последовательность. Предположим, когда вы учились в седьмом классе, вы заучивали Геттисбергское послание, и сейчас вы хотите повторить его. В высших речевых областях кортекса хранится паттерн, представляющий известную речь Линкольна. Во-первых, этот паттерн разворачивается в воспоминание о последовательности фраз и предложений. В следующей области ниже по иерархии каждая фраза разворачивается в воспоминание о последовательности слов. В этом месте разворачивающийся паттерн раздваивается и идет вниз и по слуховой части кортекса, и по моторной. Следуя по моторному пути, каждое слово разворачивается в воспоминание о последовательности фонем. И, наконец, в нижней области каждая фонема развертывается в последовательность мышечных команд для произнесения звуков. Чем ниже по иерархии вы смотрите, тем быстрее сменяются паттерны. Единый, неизменный паттерн на вершине моторной иерархии в конечном счете приводит к сложной и длинной последовательности звуков речи.

Инвариантность также работает в нашу пользу, когда информация идет вниз по иерархии. Если вы хотите напечатать Геттисбергское послание вместо того, чтоб произнести, все начинается с того же самого паттерна на вершине иерархии. В следующей области ниже по иерархии паттерн разворачивается в фразы. Еще ниже фразы разворачиваются в слова. До сих пор нет разницы между тем, произносится Геттисбергское послание или печатается. Но в следующей области ниже по иерархии моторный кортекс выбирает другой путь. Слова разворачиваются в буквы, буквы разворачиваются в мышечные команды для того, чтоб ваши пальцы печатали. "Four score and seven years ago our fathers brought forth…" – воспоминание об этих словах обрабатывается как инвариантное представление; не имеет значения, произносите ли вы, печатаете или пишете их от руки. Заметьте, вам не нужно запоминать речь дважды, один раз для произнесения, один – для написания. Единое воспоминание об этой речи может принимать различные формы поведения. В любой области инвариантный паттерн может разветвиться и следовать различными путями.

Как дополнительная эффективность, представление о простых объектах на нижних уровнях иерархии может быть многократно использовано для различных высокоуровневых последовательностей. К примеру, нет необходимости изучать одну последовательность слов для Геттисбергского послания и совершенно другую – для речи "I Have a Dream" Мартина Лютера Кинга, даже если только некоторые слова в двух высказываниях совпадают. Иерархия вложенных последовательностей позволяет разделять и многократно использовать низкоуровневые объекты – слова, фонемы и буквы – всего лишь отдельные примеры. Это замечательный эффективный способ хранить информацию о мире и его структуре, и он сильно отличается от того, как работает компьютер.

Та же самая развертываемая последовательность возникает как в сенсорных, так и в моторных областях. Этот процесс позволяет воспринимать и понимать объекты с различных точек зрения. Если вы идете к холодильнику, чтоб взять мороженого, ваш визуальный кортекс активен на множестве уровней. На высшем уровне вы постоянно воспринимаете «холодильник». В областях ниже эти визуальные ожидания разбиваются на серию более локализованных визуальных паттернов. Взгляд на холодильник состоит из фиксаций на дверной ручке, на дозаторе мороженого, на магнитах двери, на детских рисунках и так далее. За несколько миллисекунд, когда вы производите саккаду с одной черты холодильника на другую, предсказание о результате каждой саккады идет вниз по визуальной иерархии. Пока эти предсказания подтверждаются от саккады к саккаде, ваши высшие визуальные области остаются удовлетворенными тем, что вы фактически смотрите на холодильник. Заметьте, что в этом случае, в отличие от фиксированного порядка слов в Геттисбергском послании, последовательность, которую вы видите, когда смотрите на холодильник, не фиксирована; поток информации и паттерны воспоминаний зависят от ваших действий. Таким образом, в случае, подобном этому, развертывание паттерна не является жесткой последовательностью, но результат тот же самый; медленно меняющийся, высокоуровневый паттерн развертывается в более быстро меняющиеся, низкоуровневые паттерны.

Способ, которым вы запоминаете последовательности и представляете их именами, когда информация течет вверх и вниз по кортикальной иерархии, может напомнить вам иерархию военных команд. Генералы на армейской верхушке говорят: «Выдвинуть войска во Флориду на зиму». Простая высокоуровневая команда разворачивается в более детальные последовательности команд, когда она просачивается вниз по уровням иерархии. Нижестоящее командование понимает, что команда требует последовательности таких шагов, как подготовка к покиданию старой позиции, транспортировки во Флориду и подготовки к прибытию на новое место. Каждый из этих шагов разбивается на еще более подробные, чтоб быть выполненными подчиненными. Внизу иерархии тысячи отдельных структур выполняют десятки тысяч действий, приводящих в результате к перемещению войск. Отчеты о том, что происходит, генерируются на каждом уровне. Когда они просачиваются вверх по иерархии, они снова и снова суммируются, до тех пор, пока генерал на верхнем уровне иерархи не получит краткий отчет: «Перемещение во Флориду произошло успешно». Генерал не вникает в детали.

Из этого правила есть исключение. Если что-то идет не так и не может быть выполнено на подчиненных уровнях по цепочке команд, то отчет идет вверх по иерархии до тех пор, пока кто-то не будет знать, что делать далее. Офицер, который знает, как разрешить ситуацию, не видит исключения. Что было непредвиденной проблемой для подчиненных, всего лишь следующая задача в его списке. Офицер затем дает новую команду подчиненным. Неокортекс ведет себя подобным образом. Как мы увидим вскоре, когда возникающие события (другими словами - паттерны) не ожидаемы, информация о них поступает наверх по кортикальной иерархии до тех пор, пока какая-нибудь область не сможет обработать ее. Если нижние области кортекса не смогут предсказать, какой паттерн они видят, они рассматривают это как ошибку и передают ошибку вверх по иерархии. Это повторяется до тех пор, пока какая-то область не предскажет паттерн.

\* \* \*

**В** силу своей конструкции, каждая кортикальная область пытается хранить и вспоминать последовательности. Но это все еще простое описание мозга. Необходимо добавить немного усложнения в эту модель.

Информация, идущая снизу вверх к областям кортекса – это входные паттерны, распределенные по тысячам и миллионам аксонов. Эти аксоны приходят из различных областей и содержат все виды паттернов. Количество паттернов, которое возможно даже на тысяче аксонов, больше, чем количество молекул во вселенной. За время жизни область кортекса видят только крошечную долю этих возможных паттернов.

Здесь возникает вопрос: Когда отдельная область хранит последовательность, то последовательность чего именно она хранит? Ответ в том, что область сначала классифицирует информацию, как одну из ограниченного количества возможностей, и только потом ищет последовательность. Вообразите, что вы – отдельная кортикальная область. Ваша задача – сортировать кусочки цветной бумаги. Вам предоставили десять корзин, каждая из которых помечена образцом цвета. Одна корзина для зеленых, другая для желтых, третья для красных и так далее. Вам дают кусочки цветной бумаги, один за другим, и велят сортировать их по цветам. Каждая полученная вами бумажка слегка отличается. Поскольку в мире бесконечное количество цветов, вы никогда не получите двух бумажек в точности одного и того же цвета. Иногда легко сказать, в какую корзину должна быть помещена цветная бумажка, но иногда это сложно. Бумажка, которая наполовину красная, а наполовину оранжевая, может быть помещена в любую корзину, но вы должны поместить только в одну из низ, либо в красную, либо в оранжевую, даже если придется выбирать произвольно. (Цель данного упражнения – показать, что мозг должен классифицировать паттерны. Области кортекса делают это, но нет ничего подобного корзинам, в которые клались бы паттерны.)

Теперь вам дали дополнительное задание найти последовательность. Вы замечаете, что часто возникает последовательность «красный красный зеленый пурпурный оранжевый зеленый». Вы называете это «ККЗПОЗ» - последовательностью. Заметьте, что ни одну последовательность невозможно распознать, если вы не классифицируете каждый кусочек бумаги. Без предварительной классификации каждого кусочка бумаги к одной из десяти категорий вы не сможете сказать, что две последовательности являются одинаковыми.

И так, вы поняли задачу и выполняете ее. Вы собираетесь просматривать все входные паттерны – кусочки цветной бумаги, поступающие из низших кортикальных областей – классифицировать их и искать последовательности. Оба шага, классификация и формирование последовательности необходимы для создания инвариантного представления, и каждая область кортекса делает это.

Процесс формирования последовательности окупается, когда информация неоднозначна, как в случае с кусочком бумаги, который попадает иногда между красным и оранжевым. Вы должны выбрать корзину для бумажки, даже если вы не уверены, более красная она или более оранжевая. Если вы знаете наиболее подходящую последовательность для этой серии входных паттернов, вы используете это знание для того, чтоб решить, как классифицировать неоднозначную информацию. Если вы уверены, что вы в «ККЗПОЗ» последовательности, потому что вы только что получили две красных, зеленую и пурпурную, вы ожидаете, что следующая бумажка будет оранжевой. Но следующий кусок бумаги оказывается не оранжевым. Наоборот, он совершенно не там между красным и синим. Он может быть даже слегка красным, чем оранжевым. Но вы знакомы с ожиданием «ККЗПОЗ»-последовательности, и следовательно, вы помещаете бумажку в оранжевую корзину. Вы используете контекст знания последовательности для разрешения неоднозначности.

Мы видим, что этот феномен происходит постоянно в каждодневном опыте. Когда человек говорит, его конкретные слова очень часто не могут быть поняты без контекста. Но когда вы слышите неоднозначное слово в предложении, вы не заморачиваетесь на неоднозначности слова. Вы понимаете его. Аналогично рукописные слова часто неясны без контекста, но часто читаемы внутри целого предложения. Большую часть времени вы не осведомлены, что вы ощущаете неоднозначность или неполноту информации. Вы слышите то, что ожидаете услышать, и видите то, что ожидаете увидеть – по крайней мере, когда то, что вы видите и слышите, удовлетворяет вашему прошлому опыту.

Заметьте, память о последовательностях позволяет вам не только разрешать неоднозначности в полученной информации, но также предсказывать, какая информация должна поступить далее. Пока ваш кортекс сортирует цветную бумагу, вы можете сказать человеку, подающему вам бумагу: «Эй, если ты сомневаешься, какую бумажку дать, то, по моим воспоминаниям, это должна быть оранжевая». Распознавая последовательности паттернов, кортикальные области будут предсказывать будущую информацию и говорить нижестоящим областям, чего они должны ожидать.

Область кортекса не только изучает последовательности, она также обучается модифицировать их классификацию. Скажем, вы начали с набора корзин, помеченных «зеленые», «желтые», «красные», «пурпурные» и «оранжевые». Вы готовы распознавать последовательность «ККЗПОЗ» также, как и другие комбинации этих цветов. Но что если цвет будет сильно отличаться? Что если каждый раз, когда вы видите последовательность «ККЗПОЗ», пурпурный нередко сильно отличается? Новый цвет больше похож на индиго. Так что вы заменяете пурпурную корзину на «индиго». Теперь корзина лучше соответствует тому, что вы видите; вы сократили неоднозначность. Кортекс пластичен.

В кортикальных областях классификации снизу вверх и последовательности сверху вниз постоянно взаимодействуют, изменяясь в течение жизни. В этом суть обучения. Фактически, все области кортекса пластичны, таким образом они могут модифицироваться с накоплением опыта. Формирование новых классификаций и новых последовательностей – это именно то, как вы помните мир.

Напоследок, давайте взглянем на то, как классификация и предсказание взаимодействуют в области кортекса выше по иерархии. Другое назначение вашего кортекса – передать в вышестоящую область имя последовательности, которую вы видите, таким образом, вы передаете кусочек бумаги с надписью «ККЗПОЗ». Эти буквы сами по себе мало что значат для вышестоящей области; имя – это просто паттерн, который комбинируется с другой информацией, классифицируется и затем передается в еще более вышестоящие области. Подобно вам, он отслеживает последовательности, которые он видит. В некоторый момент он может сказать вам: «Эй, если ты сомневаешься, что передать мне, то, по моим воспоминаниям, я предсказываю, что это должна быть последовательность ‘ЖЖКЗЖ’». Это по сути для вас инструкция о том, что искать в вашем собственном входном потоке. Вы лучше будете интерпретировать то, какую последовательность видите.

Поскольку большинство людей слышали термин *классификация паттернов*, используемый в исследованиях по ИИ и по машинному зрению, давайте взглянем на то, как этот процесс в его обычном понимании отличается от того, что делает кортекс. В попытках заставить машины распознавать объекты исследователи обычно создают шаблоны – скажем, изображение чашки, или некоторый прототип чашки – и затем дают машине инструкции сопоставлять поступающую информацию с прототипом. Если обнаруживается близкое совпадение, компьютер скажет, что он обнаружил чашку. Но в нашем мозгу нет шаблонов подобных этому, и паттерны, которые получает каждая область кортекса, не похожи на картинки. У вас нет воспоминаний о мгновенных снимках с вашей сетчатки, или мгновенных снимков паттернов от улитки или от кожи. Иерархическое представление, являющаяся памятью об объектах, распределено по кортикальной иерархии; она не располагается в одной точке. Также, поскольку каждая область формирует инвариантные воспоминания, то, что обычно область кортекса изучает в последовательностях в инвариантной форме, само по себе и является последовательностью инвариантных воспоминаний. Вы не найдете изображения чашки или какого-либо другого объекта, хранящегося в вашем мозге.

В отличие от памяти видеокамеры, ваш мозг помнит мир таким, какой он есть, а не таким, как он видится. Когда вы думаете о мире, вы вспоминаете последовательности паттернов, которые соответствуют тому, чем являются и как себя ведут объекты мира, а не тому, как они представляются через те или иные органы чувств в любой момент времени. Последовательности, с помощью которых вы получаете опыт об объектах мира, отражают инвариантную структуру самого мира. Порядок, в котором вы воспринимаете части мира, определяется структурой мира. Например, вы можете попасть в самолет, идя прямо по переходу-«рукаву», но не через пункт регистрации на рейс. Последовательности, с помощью которых вы воспринимаете мир, *являются* реальной структурой мира, и они именно то, что помнит кортекс.

Не забывайте, однако, что инвариантное представление в любой области кортекса может быть развернуто в детальное предсказание того, как ситуация появится в ваших органах чувств, путем распространения паттерна вниз по иерархии. Аналогично инвариантное представление в моторном кортексе может быть развернуто в моторные команды, специфические в данной ситуации, путем распространения паттерна вниз по моторной иерархии.

## 6.6. На что похожа область кортекса

Сейчас я собираюсь обратить ваше внимание на отдельную область кортекса, одну из тех, что изображены на рисунке 5. На рисунке 6 такая область показана более детально. Моя цель показать вам, как клетки в области кортекса могут запоминать и вспоминать последовательности паттернов, что является наиболее существенным элементом для формирования инвариантного представления и предсказаний. Мы начнем с описания, как выглядит область кортекса, и как она образуется. Кортикальные области значительно отличаются по размерам, наиболее крупные – в первичных сенсорных областях. V1, например, размером с паспорт в смысле площади, которая она занимает на кортексе. Но, как я указал ранее, в действительности она состоит из множества мелких областей размером с букву на этой странице. Сейчас давайте предположим, что типичная кортикальная область размером с небольшую монету.



**Рисунок 6. Слои и колонки в кортикальной области.**

Вспомните о шести визитках, о которых я упоминал в главе 3, где каждая визитка представляет отдельный слой кортикальной ткани. Почему говорят о слоях? Если вы возьмете нашу кортикальную область монетного размера и поместите ее под микроскоп, вы увидите, что плотность и форма клеток изменяется по мере того, как вы движетесь сверху вниз. Эти отличия задают слои. Верхний слой, называемый первым, наиболее отличный от остальных. В нем очень мало клеток, и он состоит преимущественно из переплетения аксонов, идущих параллельно кортикальной поверхности. Слои 2 и 3 выглядят почти одинаково. В них содержится множество плотно упакованных пирамидальных клеток. В слое 4 клетки звездообразной формы. В слое пять как обычные пирамидальные клетки, так и сверхбольшие пирамиды. Нижний слой, слой 6 также содержит несколько уникальных типов нейронов.

Визуально мы обнаруживаем горизонтальные слои, но очень часто ученые говорят о колонках клеток, которые идут перпендикулярно слоям. Вы можете думать о колонках как о вертикальных «модулях» из клеток, работающих совместно. (Термин *колонка* вызывает множество дебатов в содружестве нейроученых. Их размеры, функции и значимость спорны. Для наших целей, однако, вы можете думать в общих терминах о колончатой архитектуре, с существованием которой согласны все.) Слои внутри каждой колонки соединяются аксонами, которые идут вверх и вниз и соединяются синапсами по пути. Колонки не выделяются подобно колоннам с четкими границами – в кортексе нет ничего простого – но об их существовании можно догадываться исходя из нескольких фактов.

Одна причина – это то, что вертикально выровненные клетки в каждой колонке имеют тенденцию активизироваться в ответ на один и тот же стимул. Если мы взглянем поближе на колонки в V1, мы обнаружим, что некоторые отвечают на линейные сегменты, которые наклонены в одном направлении (/), а другие отвечают на линейные сегменты, наклоненные в другом направлении (\). Клетки в каждой колонке сильно связаны, именно поэтому колонка целиком отвечает на один и тот же стимул. В особенности, активные клетки в слое 4 заставляют становиться активными клетки в слоях 2 и 3, что затем приводит к активизации клеток в слоях 5 и 6. Активность распространяется вверх и вниз по клеткам колонки.

Другая причина, почему мы говорим о колонках, лежит в том, как формируется кортекс. У эмбриона единственная клетка-предшественник мигрирует из внутренней полости мозга туда, где формируется кортекс. Каждая из этих клеток делится и создает около сотни нейронов, называемых микроколонкой, которые по вертикали соединяются так, как я только что описал. Термин *колонка* часто неточно используется только для описания различных феноменов; он может ссылаться на вертикальные соединения или на специфические группы клеток от одного и того же предка. Используя последнее определение, мы можем сказать, что в человеческом кортексе приблизительно несколько сотен миллионов микроколонок.

Чтобы увидеть эту колончатую структуру, вообразите отдельную микроколонку толщиной в человеческий волос. Возьмите тысячи волос и отрежьте от них очень короткий сегмент – скажем, высотой с букву i без точки. Выровняйте эти волоски или колонки и склейте их в виде очень плотной кисточки. Затем создайте слой из длинных, очень тонких волос – представляющих аксоны из слоя 1 – и приклейте их горизонтально поверх слоя из коротких волосинок. Это похожий на кисточку слой является очень упрощенной моделью маленькой кортикальной области. Информация течет в основном в направлении этих волос: горизонтально в слое 1 и вертикально в слоях со второго по шестой.

Есть еще одна деталь о колонках, которую вам необходимо знать, и затем мы приступим к тому, для чего это все. При близком рассмотрении мы видим, что как минимум 90 процентов синапсов на клетках в каждой колонке приходит извне этой колонки. Некоторые соединения приходят от соседних колонок. Другие приходят с другой половины мозга. Как же мы можем говорить о значимости колонок, если так много кортикальных соединений распространяются вбок на больших расстояниях?

Ответ в модели «память-предсказание». В 1979, когда Вернон Монткастл указал, что существует единый кортикальный алгоритм, он также предположил, что кортикальные колонки являются базовым вычислительным модулем кортекса. Однако, он не знал, какую функцию выполняют колонки. Я верю, что колонки являются базовым модулем предсказания. Чтобы колонка могла предсказать, когда она должна активизироваться, она должна знать, что происходит в других местах – отсюда синаптические соединения с различных направлений.

Скоро мы вдадимся в детали, но вот обзор того, почему нам надо знать этот вид соединений в мозге. Для предсказания следующей ноты в песне вам нужно знать название песни, в каком месте песни вы находитесь, сколько времени прошло с момента последней ноты и какая была последняя нота. Большое число синаптических соединений, соединяющих клетки в колонках с другими частями мозга обеспечивает каждую колонку контекстом, который ей нужен для того, чтоб предсказать ее активность во множестве различных ситуаций.

\* \* \*

**С**ледующее, что нам необходимо рассмотреть – как эти маленькие кортикальные области (и их колонки) посылают и получают информацию вверх и вниз по кортикальной иерархии. Сначала взглянем на восходящий поток, который имеет относительно прямой маршрут, отображенный на рисунке 7. Вообразите, что мы смотрим на кортикальную область с ее тысячами колонок. Рассмотрим в большом масштабе только одну. Сходящаяся информация от нижестоящих областей всегда приходит в слой 4 – основной входной слой. По пути входные пути формируют соединения в слое 6 (мы увидим позже, почему это существенно). Клетки слоя 4 затем посылают проекции вверх к клеткам в слоях 2 и 3 внутри колонки. Когда колонка отправляет информацию вверх, большинство клеток слоев 2 и 3 посылают аксоны ко входному слою колонок следующей вышестоящей области. Таким образом информация течет от области к области вверх по иерархии.

Информация, идущая вниз по кортикальной иерархии имеет менее прямой маршрут, что изображено на рисунке 8. Клетки слоя 6 являются отправителями нисходящих соединений колонки и проецируются на слой 1 в иерархически нижестоящих областях. В слое 1, аксон распространяется на большие дистанции в низших кортикальных областях. Таким образом информация, текущая вниз по иерархии от одной колонки имеет возможность активизировать множество колонок в нижестоящей области. В слое 1 очень мало клеток, но клетки слоев 2, 3 и 5 имеют дендриты в слое 1, таким образом клетки могут быть возбуждены обратными связями, идущими через слой 1. Аксоны, идущие от клеток в слоях 2 и 3 формируют синапсы в слое 5, когда они покидают кортекс, и, предположительно, возбуждают клетки в слоях 5 и 6. Таким образом мы можем сказать, что когда информация течет вниз по иерархии, ее путь более извилист. Она может разветвиться по множеству различных направлений через распределение в слое 1. Обратная информационная связь начинается с клетки в слое 6 в области, выше по иерархии; она распространяется по слою 1 в нижестоящих областях. Некоторые клетки в слоях 2, 3 и 5 в нижестоящих областях возбуждаются, и некоторые из них возбуждают клетки слоя 6, которые проецируют на слой 1 в еще более нижестоящих по иерархии, и так далее. (Если вы изучите рисунок 8, будет гораздо понятнее.)



**Рисунок 7. Восходящие информационные потоки в кортикальную область**

Вот предварительные сведения о том, почему информация распределяется по слою 1. Чтобы преобразовать инвариантное представление в конкретное предсказание требуется способность решать от момента к моменту, каким путем посылать сигнал, когда он распространяется вниз по иерархии. Слой 1 обеспечивает способ преобразования инвариантного представления в более детальное и конкретное представление. Как я говорил выше, вы можете вспомнить Геттисбергское послание либо в устной, либо в письменной форме. Общее представление идет по одному из двух путей, один для устной, другой для письменной речи. Аналогично, когда я слышу следующую ноту мелодии, мой мозг должен взять интервал, например квинту, и преобразовать ее к конкретной ноте, такой как До или Соль. Горизонтальные потоки активности по слою 1 обеспечивают механизм для этого. Чтобы высокоуровневое инвариантное предсказание распространилось вниз по иерархии и стало конкретным предсказанием, мы должны иметь механизм, который позволяет потокам паттернов ветвиться на каждом уровне. Слой 1 отвечает всем требованиям. Мы могли бы предсказать его необходимость, даже если бы мы не знали о его существовании.



**Рисунок 8. Нисходящие информационные потоки в кортикальной области.**

Последний анатомический штрих: когда аксоны покидают слой 6 чтоб уйти в другие места, они упаковываются в оболочку из белой жировой субстанции, называемой миелином. Эта так называемое белое вещество похоже на изоляцию электрических проводов в вашем доме. Оно помогает предотвратить искажение сигналов и повысить скорость их прохождения, увеличивая ее до двухсот миль в час. Когда аксоны покидают белое вещество, они входят в новую кортикальную колонку в слой 6.

\* \* \*

**В** конечном счете есть еще один метод непрямой коммуникации кортикальных областей.

Прежде чем я опишу его в деталях, я хочу напомнить вам об автоассоциативной памяти, которая обсуждалась в главе 2. Как вы помните, автоассоциативная память может быть использована для хранения последовательностей паттернов. Когда выход группы искусственных нейронов используется как обратная связь на входы всех нейронов, и к обратной связи добавляется задержка, то паттерны обучаются следовать один за другим в последовательности. Я верю, что кортекс использует тот же самый базовый механизм для хранения последовательности, хотя с небольшими дополнительными ухищрениями. Вместо того, чтоб формировать автоассоциативную память из нейронов, он формирует автоассоциативную память из кортикальных колонок. Выход всех колонок направляется к слою 1. Таким образом, слой 1 содержит информацию о том, какие колонки были только что активны.

Давайте пройдемся по элементам, как показано на рисунке 9. Уже давно известно, что особенно большие клетки в слое 5 в моторном кортексе (область M1) направляет прямые соединения к мышцам и моторным областям в спинном мозге. Эти клетки буквально приводят в действие мышцы и заставляют вас двигаться. Когда вы либо говорите, печатаете или выполняете какие то сложные действия, эти клетки возбуждаются скоординированным образом, заставляя ваши мышцы сокращаться.



**Рисунок 9. Как текущее состояние и моторное поведение широко взаимодействуют через таламус.**

Совсем недавно исследователи открыли, что гигантские клетки в слое 5 могут играть роль в поведении в других частях кортекса, не только в моторных областях. Например, большие клетки в визуальном кортексе проецируются на часть мозга, управляющую движением глаз. Таким образом, визуальные области кортекса, такие как V2 и V4, не только обрабатывают визуальную информацию, но также влияют на движение глаз, и следовательно на то, что вы видите. Большие клетки слоя 5 наблюдаются по всему неокортексу, во всех областях, претендуя на более значительную роль во всех видах движений.

В дополнение к поведенческой роли, аксоны этих больших клеток слоя 5 раздваиваются. Одна ветвь идет к части мозга, называемой таламусом, показанной на рисунке 9 круглым объектом. Таламус человека имеет форму и размер двух яиц небольших птиц. Он располагается в самом центре мозга, на верхушке старого мозга и окружен белым веществом и кортексом. Таламус получает множество аксонов от всех частей кортекса и посылает аксоны обратно к тем же самым областям. Большинство деталей эти соединений известно, но сам по себе таламус сложная структура и его роль не вполне ясна. Но таламус необходим для нормальной жизни; повреждение таламуса ведет к постоянному вегетативному состоянию.

Есть несколько путей из таламуса в кортекс, но только один интересен нам сейчас. Этот путь начинается в больших клетках слоя 5, которые проецируются на считающийся неспецифическим класс таламических клеток. Неспецифические клетки проецируют аксоны обратно в слой 1 во многие различные области кортекса. Например, клетки 5-го слоя со всей площади областей V2 и V4 посылают аксоны в таламус, а таламус посылает информацию обратно в слой 1 на всю площадь областей V2 и V4. Другие части кортекса делают то же самое; клетки 5-го слоя с различных кортикальных областей проецируются на таламус, который посылает информацию обратно к 1-му слою этих же самых и ассоциированных областей. Я предполагаю, что эти контуры в точности похожи на обратные связи с задержками, которые позволяют модели автоассоциативной памяти запоминать последовательности.

Я сейчас должен упомянуть о двух путях поступления информации в слой 1. Вышестоящие области кортекса распространяют активность по слою 1 нижестоящих областей. Активные колонки в этих областях также распространяют активность через слой 1 в тех же самых областях через таламус. Мы можем думать об этих входах в слой 1 как о названии песни (вход из вышестоящих областей) и как о позиции в песне (задержанная активность от активных колонок в этой же области). Таким образом, слой 1 принимает большинство той информации, которая нам нужна для предсказания того, когда колонка должна быть активной – имя последовательности и позиция в последовательности. Используя эти два сигнала в слое 1, область кортекса может запоминать и вспоминать множество последовательностей паттернов.

## 6.7. Как работают области кортекса: детали

Держа в уме эти три вещи – схождение паттернов, идущих вверх по иерархии, расхождение паттернов идущих вниз по иерархии и обратная связь с задержкой через таламус – мы можем приступить к рассмотрению того, как области кортекса выполняют необходимые им функции. Вот что мы хотим знать:

1. Как область кортекса классифицирует поступающую информацию (аналогично корзинам)?
2. Как она запоминает последовательности паттернов (такие как интервалы мелодии или последовательности «глаз нос глаз» при рассматривании лица)?
3. Как она формирует постоянные паттерны или «названия» последовательностей?
4. Как она делает конкретные предсказания (встреча поезда в правильное время или предсказание конкретной ноты в мелодии)?

Давайте начнем с предположения, что колонки в области кортекса похожи на корзины, которые мы использовали при классификации цветных бумажек. Каждая колонка представляет метку корзины. Клетки в слое 4 в каждой колонке получают входные волокна от нескольких областей ниже и возбуждаются, если поступает правильная комбинация. Когда возбуждаются клетки в слое 4, это «вынесение решения» о том, что информация удовлетворяет метке. По аналогии с сортировкой бумажек, информация может быть неоднозначной, так что возможно несколько колонок могу соответствовать этой информации. Мы хотим, чтобы область кортекса принимала однозначное решение; бумажка либо красная, либо оранжевая, но не то и другое одновременно. Колонка с наиболее интенсивным входом должна предотвратить возбуждение других колонок.

В мозге есть клетки, которые делают именно это. Они интенсивно тормозят другие нейроны в соседних колонках, эффективно приводя только к одному победителю. Эти тормозные клетки воздействуют только на область, окружающую колонку. Таким образом, даже если имеется значительное торможение, все равно большинство колонок в области может быть активными одновременно. (В реальном мозге нет ничего, что бы представлялось единственным нейроном или колонкой). Для того, чтоб облегчить дальнейшее понимание, вы можете временно предположить, что область выбирает одну и только одну колонку-победителя. Но держите в уме, что одновременно будут активны множество колонок. Актуальный процесс, используемый областью кортекса для классификации информации и то, как она запоминает, слишком сложно и не до конца понятно. Я не буду пытаться протащить вас через выводы. Вместо этого я хочу предположить, что область кортекса классифицирует информацию как набор активных колонок. Затем мы сможем сфокусироваться на последовательностях и названиях последовательностей.

Как кортикальная область хранит последовательность классифицированных паттернов? Я уже предлагал ответ на этот вопрос, но я не вдавался в детали. Вообразите, что вы колонка из нейронов, и информация из нижележащих областей заставляет одну из клеток слоя 4 возбуждаться. Вы счастливы и клетка из вашего 4-го слоя заставляет также возбуждаться клетки в слоях 2 и 3, затем 5 и 6. Колонка целиком становится активной, когда возбуждена из нижестоящих областей. Ваши клетки в слоях 2,3 и 5 имеют тысячи синапсов в слое 1. Если некоторые из этих синапсов активны, когда возбуждаются клетки в слоях 2, 3 и 5, эти синапсы усиливаются. Если это происходит достаточно часто, эти синапсы в слое 1 становятся настолько сильными, что заставляют возбуждаться клетки в слоях 2, 3 и 5 даже если клетки 4-го слоя не возбуждены – значит, некоторые колонки могут активизироваться, даже не получая информации от нижестоящих областей кортекса. В этом случае клетки в слоях 2, 3 и 5 обучаются предсказывать момент, когда они должны возбудиться от паттерна в слое 1. До обучения колонка может становиться активной только если возбуждается из 4-го слоя. После обучения колонка может активизироваться частично по памяти. Когда колонка активизируется от синапсов в слое 1, ей легче возбудиться от нижестоящих областей. Это предсказание. Если бы колонка могла говорить, она сказала бы: - «Когда я активизировалась в прошлом, определенное множество синапсов в слое 1 были активны. Так что когда я снова увижу это определенное множество, я начну возбуждаться заранее».

Вспомните, что половина информации поступает в слой 1 от клеток 5-го слоя в соседних колонках и областях кортекса. Эта информация представляет то, что происходило моментом ранее. Она представляет колонки, которые были активны до того, как стала активна ваша колонка. Она представляет предыдущий интервал мелодии, или последнее, что я видел, или последнее, что я чувствовал или предыдущую фонему в речи, которую я слушаю. Если порядок, в котором возникают эти паттерны, постоянен, то колонки запоминают порядок. Они будут возбуждаться одна за другой в правильной последовательности.

Другая половина информации приходит в слой 1 от клеток слоя 6 в вышестоящих областях. Эта информация более стационарна. Она представляет название последовательности, которую вы в данный момент ощущаете. Если колонки – музыкальные интервалы, то это название мелодии. Если колонки – фонемы, то это слово, которое вы слышите. Если колонки – слова, то сигнал сверху – это предложение. Таким образом информация в слое 1 представляет и название последовательности и последний ее элемент. В этом случае определенные колонки могут входит в состав различных последовательностей без всякой путаницы. Колонки учатся возбуждаться в правильном контексте и в правильном порядке.

Прежде чем двинуться дальше, я должен указать, что не только синапсы в слое 1 участвуют в запоминании того, когда колонка должна стать активной. Как я упоминал ранее, клетки получают и посылают информацию во множество окружающих колонок. Вспомните, что более 90 процентов всех синапсов приходят от клеток из других колонок, и большинство этих синапсов располагаются не в слое 1. Например, клетки в слоях 2, 3 и 5 имеют тысячи синапсов в слое 1, но также тысячи синапсов в их собственном слое. Общая идея в том, что клетки ищут любую информацию, которая поможет им предсказать, когда они должны быть активизированными снизу. Обычно активность в близкорасположенных колонках коррелирует, таким образом мы видим множество прямых соединений к близлежащим колонкам. Например, если линия движется через визуальное поле, она будет активизировать соответствующие колонки. Часто, однако, информация, необходимая для предсказания активности колонки, более глобальна, в этом случае играет роль слой 1. Если б вы были нейроном в колонке, вы бы не знали, что обозначает любой из этих синапсов, все, что вы знали бы – это то, что они помогают вам предсказать, когда вы должны активизироваться.

\* \* \*

**С**ейчас давайте рассмотрим вывод того, как область кортекса формирует название последовательности. Снова вообразите, что вы область кортекса. Ваши активные колонки изменяются с поступлением новой информации. Вы успешно выучили порядок, в котором колонки становятся активными, а значит, некоторые из клеток становятся активными до прибытия информации из нижестоящих областей. Какую информацию вы посылаете в область кортекса выше по иерархии? Мы видели ранее, что клетки в ваших слоях 2 и 3 посылают аксоны в следующую область выше по иерархии. Активность этих клеток является входной информацией для вышестоящих регионов. Но тут есть одна проблема. Для того, чтоб работала иерархия, вы должны транслировать постоянный паттерн в течение запомненной последовательности; вы должны передавать название последовательности, но не ее детали. До того, как вы запомните последовательность, вы можете передавать детали, но после того, как вы ее запомнили и способны успешно предсказывать, какие колонки будут активными, вы должны передавать только постоянный паттерн. Однако, я еще не показал вам, как это сделать. Как подразумевается сейчас, вы будете передавать в ответ на любой изменившийся паттерн, независимо от того, можете ли вы предсказать его. Когда любая колонка становится активной, клетки в ее слоях 2 и 3 будут посылать сигнал вверх по иерархии. У кортекса должен быть способ удерживать постоянный паттерн, передаваемый в вышестоящие области в течение запомненной последовательности. У нас должен быть способ отключить выход от клеток слоев 2 и 3, когда колонка предсказывает свою активность, или наоборот, сделать эти клетки активными, когда колонка не может предсказать свою активность. Это единственный способ сформировать постоянный паттерн.



**Рисунок 10. Формирование постоянного названия для запомненной последовательности.**

О кортексе недостаточно известно, чтобы точно утверждать, как он это делает. Я могу вообразить несколько способов. Я опишу наиболее предпочитаемый мной сейчас, но имейте в виду, что более важна концепция, чем конкретный метод. Создание постоянного «названия» паттерна – требование этой теории. Все, что я могу сейчас показать это то, что существует правдоподобный механизм для процесса именования.

Снова вообразите, что вы колонка, как показано на рисунке 10. Мы хотим понять, как вы обучаетесь представлять в вышестоящую область постоянный паттерн, когда вы можете предсказать свою активность, и изменяющийся паттерн – когда не можете. Давайте начнем с предположения, что в слоях 2 и 3 есть несколько классов нейронов. (В дополнение к нескольким типам тормозящих нейронов многие анатомы различают типы нейронов, которые они называют слоями 3а и 3б, так что это предположение не беспочвенно).

Давайте также предположим, что один класс нейронов, называемый нейронами слоя 2, учатся задерживать активность, пока длится заученная последовательность. Группа таких нейронов представляет название последовательности. Они предоставляют постоянный паттерн вышестоящей кортикальной области, пока наша область может предсказывать, какие колонки станут активными. Если наша область кортекса запомнила последовательность из трех различных паттернов, то нейроны в слое 2 всех колонок, представляющих эти паттерны должны оставаться активными, пока внутри последовательности. Они являются названием последовательности.

Теперь давайте предположим, что есть еще один класс нейронов, нейроны слоя 3б, которые не возбуждаются, если наша колонка успешно предсказывает поступающий паттерн, но возбуждаются, когда она не предсказала свою активность. Нейроны слоя 3б представляют неожиданный паттерн. Они возбуждаются, когда колонка становится активной неожиданно. Они возбуждаются каждый раз, когда колонка становится активной до того, как обучится. Но по мере обучения предсказанию активности нейроны слоя 3 становятся все спокойнее. Слои 2 и 3б совместно удовлетворяют нашим требованиям. До обучения и те и другие активизируются и дезактивируются в такт с колонкой, но после обучения нейроны слоя 2 остаются постоянно активными, а нейроны слоя 3 становятся молчащими.

Как эти нейроны обучаются этому? Во-первых, давайте рассмотрим, как заставить замолчать нейроны слоя 3б, когда колонка успешно предсказывает свою активность. Скажем, есть другой нейрон, расположенный выше в слое 3б, а слое 3а. У этого нейрона есть также дендриты в слое 1. Его единственная цель – предотвратить возбуждение нейрона в слое 3б, когда он видит соответствующий паттерн в слое 1. Когда нейроны слоя 3а видят заученный паттерн в слое 1, они быстро активизируют тормозящие нейроны, которые предотвращают возбуждение нейронов слоя 3б. Все это могло бы остановить возбуждение нейронов слоя 3б, когда колонка корректно предсказывает активность.

Теперь давайте рассмотрим более сложную задачу – удержание постоянной активности в течение известной последовательности паттернов. Это сложнее, потому что различные множества нейронов в слое 2 во множестве различных колонок должны оставаться активными все вместе, даже когда их индивидуальные колонки неактивны. Вот как я себе это представляю. Нейроны слоя 2 могли бы обучаться становиться активными только от иерархически вышестоящих областей кортекса. Они могли бы формировать синапсы преимущественно с аксонами из слоя 6 из вышестоящих областей. Нейроны слоя 2 могли бы таким образом представлять постоянное имя паттерна от вышестоящей области. Когда вышестоящая область кортекса посылает паттерн вниз к слою 1 нижестоящей области, множество нейронов в слое 2 в нижестоящей области должно стать активными, представляя все колонки, которые являются членами последовательности. Поскольку эти нейроны слоя 2 также проецируются обратно в вышестоящую область, они должны формировать полустабильную группу нейронов. (Это отличается от того, если бы эти нейроны оставались активными постоянно. Они возможно возбуждаются синхронно в некотором ритме). Это как если бы вышестоящая область посылала бы название мелодии в слой 1 ниже. Это событие заставляет множество нейронов слоя 2 возбуждаться, те, у которых колонка должна быть активной, когда слышится мелодия.

В сумме эти механизмы позволяют кортексу запоминать последовательности, делать предсказания и формировать константные представления, или «названия» последовательностей. Это базовая операция для формирования инвариантного представления.

\* \* \*

**К**ак мы делаем предсказания о событиях, которые мы никогда раньше не видели? Как мы выбираем из множества интерпретаций входной информации? Как область кортекса делает конкретное предсказание из инвариантного воспоминания? Я приводил несколько примеров ранее, таких как предсказание точной следующей ноты в мелодии, когда ваша память помнит только интервалы между нотами, притчу о поезде и процесс вспоминания Геттисбергского послания. В этих случаях единственным путем решения проблемы является использование последней конкретной информации для преобразования инвариантного предсказания в конкретное. Перефразируя это в терминах кортекса, мы должны скомбинировать прямой поток информации (актуальную информацию) с обратным потоком информации (предсказание в инвариантной форме).

Вот простой пример, как, я думаю, это происходит. Скажем, вы – область кортекса и вам сказали ожидать музыкальный интервал величиной в квинту. Колонки области представляют всевозможные интервалы, такие как До-Ми, До-Соль, Ре-Ля и тому подобное. Вам необходимо решить, какая из ваших колонок должна активизироваться. Когда область выше говорит вам ожидать квинту, она заставляет нейроны слоя 2 возбуждаться во всех колонках, которые являются квинтами, такие как До-Соль, Ре-Ля и Ми-Си. Нейроны слоя 2 в колонках, представляющих другие интервалы, неактивны. Сейчас вы должны выбрать одну из колонок из множества возможных квинт. Информация, поступающая в вашу область, является конкретной нотой. Если последняя нота, которую вы слышали, была Ре, то все колонки, представляющие интервал, начинающийся на Ре, такие как Ре-Ми и Ре-Си, частично активизируются входной информацией. Таким образом, теперь в слое 2 у нас активны все колонки, являющиеся квинтами, и в слое 4 у нас идет частичное возбуждение во всех колонках, представляющих интервалы, начинающиеся на Ре. Пересечение этих двух множеств дает нам ответ, колонку, представляющую интервал Ре-Ля (см. рисунок 11).

Как кортекс находит это пересечение? Вспомните, что ранее я упоминал факт, что аксоны от нейронов из слоев 2 и 3 в основном формируют синапсы в слое 5, когда они покидают кортекс, и аналогично, аксоны, идущие в слой 4 из нижележащих областей кортекса формируют синапсы в слое 6. Пересечение этих двух синапсов (сверху вниз и снизу вверх) дает нам все, что необходимо. Нейроны слоя 6, получающие активность из обоих каналов, возбуждаются. Нейроны слоя 6 представляют то, что по мнению области кортекса происходит, конкретное предсказание. Если бы нейроны слоя 6 могли говорить, они могли бы сказать: «Я часть колонки, представляющей нечто. В моем конкретном случае моя колонка представляет музыкальный интервал Ре-Ля. Другие колонки обозначают другое. Я говорю от лица всей кортикальной области. Когда я становлюсь активным, это обозначает, что мы верим, что музыкальный интервал Ре-Ля либо уже возник, либо возникнет. Я мог бы стать активным, потому что информация снизу вверх от уха заставляет нейроны слоя 4 в моей колонке возбуждать всю колонку. Или моя активность могла бы обозначать, что мы узнали мелодию и предсказываем следующую конкретный интервал. В любом случае моя обязанность сказать нижестоящим областям кортекса то, что по нашему мнению происходит. Я представляю нашу интерпретацию мира, не смотря на то, является ли это истиной или просто воображаемо.»



**Рисунок 11. Как область кортекса делает конкретное предсказание из инвариантных воспоминаний.**

Позвольте мне описать это, используя другую мысленную картинку. Вообразите два кусочка бумаги с множеством маленьких дырочек. Эти дырочки в бумаге представляют колонки, у которых нейроны в слоях 2 и 3 активны, наше инвариантное предсказание. Дырочки на другой бумаге представляют колонки с частичным возбуждением от нижестоящих областей. Если вы положите одну бумажку поверх другой, некоторые из этих дырочек совпадут, другие нет. Совпавшие дырочки представляют колонки, которые должны быть активными.

Этот механизм не только делает конкретные предсказания, он также разрешает неоднозначности в сенсорной информации. Очень часто информация, поступающая в область кортекса, является неоднозначной, как мы видели в случае с цветными бумажками, или когда вы слышите полуискаженное слово. Этот механизм сопоставления потоков информации снизу вверх и сверху вниз позволяет вам выбрать из двух или более интерпретаций. Выбрав, вы сообщаете вашу интерпретацию нижестоящей области.

В каждый момент вашего бодрствования, каждая область неокортекса сравнивает множество ожидаемых колонок, возбуждаемых сверху, с множеством наблюдаемых колонок, возбужденных снизу. Пересечение этих двух множеств и есть то, что мы воспринимаем. Если б мы имеет идеальную информацию и идеальное предсказание, то множество воспринимающих колонок всегда содержалось бы в множестве предсказываемых. Часто такого согласия не наблюдается. Метод комбинирования частичного предсказания с частичной входной информацией решает проблему неоднозначности информации, он восполняет недостающие кусочки информации и выбирает между альтернативными точками зрения. Именно так мы комбинируем ожидаемый инвариантный к тону интервал с последней услышанной нотой для предсказания следующей конкретной ноты мелодии. Именно так мы решаем, является ли картинка вазой или парой лиц. Именно так мы направляем наш моторный поток либо для того, чтоб написать, либо чтобы сказать Геттисбергское послание.

В конце концов, в дополнение к проекции в нижестоящие области, нейроны слоя 6 могут посылать информацию обратно к нейронам слоя 4 своей собственной колонки. Когда они делают это, наше предсказание становится входной информацией. Именно это происходит, когда мы мечтаем или думаем. Это позволяет нам видеть следствия наших собственных предсказаний. Мы делаем это нередко в течение дня, когда планируем будущее, репетируем речь, и беспокоимся о предстоящих событиях. Архитектор кортикальных моделей Стивен Гроссберг называет это «folded feedback» («закольцованные обратные связи»). Я предпочитаю называть «воображением».

\* \* \*

**П**оследнее замечание прежде чем мы завершим этот раздел. Я несколько раз указывал, что чаще всего то, что мы видим, слышим или чувствуем зависит от наших собственных действий. То, что мы видим, зависит от того, куда совершили саккаду наши глаза и как мы повернули голову. То, что мы чувствуем, зависит от того, как мы двигаем наши пальцы и конечности. То, что мы слышим, иногда зависит от того, что мы говорим и делаем.

Таким образом, чтобы предсказать, что мы почувствуем далее, мы должны знать, какие действия мы предпримем. Моторное поведение и сенсорное восприятие сильно взаимозависимы. Как мы можем делать предсказания, если то, что мы почувствуем далее, в основном результат наших собственных действий? К счастью, есть неожиданное и элегантное решение этой проблемы, хотя некоторые детали непонятны.

Первое неожиданное открытие в том, что восприятие и поведение в основном одно и то же. Как я упоминал ранее, большинство, если не все области кортекса, даже визуальные области, участвуют в создании поведения. Нейроны слоя 5, которые проецируются в таламус и затем в слой 1 также, видимо, имеют моторную функцию, потому что они одновременно проецируются в моторные области старого мозга. Таким образом, знание о том, «что сейчас произойдет» - и сенсорное и моторное – доступно в слое 1.

Вторая неожиданная вещь и следствие первой то, что моторное поведение также должно быть представлено в иерархии инвариантного представления. Вы генерируете движение, необходимое для выполнения определенного действия, размышляя над его выполнением в инвариантной к деталям форме. Когда моторная команда идет вниз по иерархии, она транслируется в сложные и детальные последовательности, требуемые для выполнения той активности, которую вы ожидаете. Это происходит и в моторном кортексе, и в сенсорном кортексе, что стирает различие между ними. Если область IT визуального кортекса воспринимает «нос», простое действие переключения на представление для «глаза» генерирует саккаду, необходимую для того, чтоб сделать предсказание реальным. Конкретная саккада, необходимая для перевода взгляда с носа на глаз изменяется в зависимости от того, где расположено лицо. Близкорасположенное лицо требует большей саккады; лицо подальше требует саккаду поменьше. Наклоненное лицо требует совершения саккады под углом, отличающимся от того, когда лицо стоит ровно. Детали необходимой саккады определяются как предсказание увидеть движение «глаза» к V1. Саккада становится все более и более конкретной по мере продвижения вниз по иерархии, приводя в результате к саккаде, наводящей ваши фовеальные области точно на цель или близко к ней.

Давайте взглянем на другой пример. Если мне нужно физически переместиться из комнаты в кухню, все, что моему мозгу достаточно – это переключиться с инвариантного представления комнаты на инвариантное представление кухни. Это переключение вызывает сложную разворачивающуюся последовательность. Процесс генерации последовательности предсказаний того, что я увижу, почувствую и услышу пока иду из комнаты на кухню, также генерирует и последовательность моторных команд, которые заставляют меня идти из комнаты на кухню и перемещать взгляд так, как надо. Предсказание и моторное поведение работают рука об руку, когда паттерны идут вверх и вниз по кортикальной иерархии. Как бы странно это ни звучало, когда задействовано поведение, ваше предсказание не только предсказывает ощущения, оно определяет ощущения. Мысль о следующем паттерне в поведении вызывает каскад предсказаний того, что вы должны ощутить после. Когда каскад предсказаний разворачивается, он генерирует моторные команды, необходимые для успешного завершения предсказания. Думая, предсказывая и выполнение – это части одного и того же процесса разворачивания последовательностей вниз по кортикальной иерархии.

«Выполнение» через размышление, параллельное разворачивание перцептивного и моторного поведения – это суть того, что называется целенаправленным поведением. Целенаправленное поведение – это Святой Грааль робототехники. Он встроен в кортекс.

Конечно, мы можем отключить моторное поведение. Я могу размышлять о разглядывании чего-либо, на самом деле не видя этого, и я могу размышлять о походе на кухню, на самом деле не делая этого. Но размышление о выполнении чего-либо – это буквально начало того, как мы делаем что-то.

## 6.8. Вверх и вниз

Давайте отойдем слегка назад и поразмышляем о том, как информация движется вверх и вниз по кортикальной иерархии. Когда вы движетесь в мире, в низшие области кортекса попадают изменяющиеся паттерны. Каждая область пытается интерпретировать ее поток паттернов как часть известной последовательности паттернов. Колонки пытаются предсказать свою активность. Если у них получается, они передают стабильный паттерн, имя последовательности в вышестоящую область. И снова, как если бы область сказала: «Я слышу песню, вот ее название. В ее деталях я разберусь сама.»

Но что если придет неожиданный паттерн, неожиданная нота? Или что если мы увидим что-то, что не является частью лица? Неожиданный паттерн автоматически передается вышестоящей кортикальной области. Это происходит естественным путем, когда возбудятся нейроны из слоя 3б, которые не являются частью ожидаемой последовательности. Вышестоящая область может быть способна понять этот новый паттерн как часть ее собственной последовательности. Она могла бы сказать: «О, я вижу, что пришла новая нота. Может быть, это первая нота следующей песни в сборнике. Очень похоже, поэтому я предсказываю, что мы должны перейти к следующей песне. Нижестоящая область, вот название следующей песни, которая, как я думаю, вы должны сейчас услышать». Но если этого распознавания не происходит, неожиданный паттерн продолжает распространяться вверх по иерархии до тех пор, пока некоторая высшая область не сможет проинтерпретировать его как часть ее нормальной последовательности событий. Чем выше неожиданный паттерн поднимется, тем больше областей кортекса будет вовлечено в объяснение неожиданной информации. Наконец, когда область где-то вверху иерархии думает, что сможет понять неожиданное событие, она генерирует новое предсказание. Это новое предсказание распространяется вниз по иерархии до тех пор, пока получается. Если новое предсказание не верное, будет обнаружена ошибка и снова она будет карабкаться вверх по иерархии до тех пор, пока какая-либо область не сможет проинтерпретировать ее как часть ее активной в данный момент последовательности. Таким образом мы можем увидеть, как наблюдаемые паттерны идут вверх по иерархии, а предсказания – вниз по иерархии. В идеале, если бы мир был извествен и предсказуем, большинство потоков вверх-вниз происходили бы мгновенно и возникали бы в нижележащих областях кортекса. Мозг быстро пытается обнаружить, какая часть его модели мира соответствует неожиданной информации. Только тогда он поймет эту информацию и будет знать, чего ожидать дальше.

Если я иду обычным путем по комнате в моем доме, совсем немного ошибок будет распространяться вверх по кортикальной иерархии. Глубоко заученные последовательности моего дома могут быть обработаны в низших областях визуальной, соматосенсорной и моторной иерархии. Я знаю комнату настолько хорошо, что могу пройти по ней даже в темноте. Моя осведомленность об окружении эффективно освобождает большую часть моего кортекса для других задач, таких как размышление о мозге и написание книг. Однако, если бы я попал в незнакомую комнату, особенно в такую, которая отличается от любой из комнат, которые я видел ранее, мне нужно было бы не только смотреть и видеть, куда я иду, но неожиданные паттерны постоянно поднимались бы на самый верх кортикальной иерархии. Чем больше мой сенсорный опыт не соответствует заученным последовательностям, тем больше ошибок возникало бы. В такой нестандартной ситуации я не смог бы больше думать о мозге, потому что большая часть моего кортекса уделяла бы внимание проблемам навигации по комнате. Это распространенная ситуация для людей, которые выходят из самолета в незнакомой стране. Тогда как дороги могут казаться похожими на те, к которым вы привыкли, машины могут проноситься не по той стороне улицы, деньги будут странными, язык непонятным, и поиск туалета может занять всю вашу кортикальную мощь. Не пытайтесь произносить длинные речи, когда идете по чужой стране.

Ощущение неожиданного понимания, момент «Ага!» может быть объяснен в этой модели. Вообразите, что вы смотрите на неоднозначную картинку. Заполненная чернильными кляксами и разбросанными линиями, она не похожа ни на что. Она не имеет смысла. Когда кортекс не может найти какие-либо воспоминания, соответствующие поступающей информации, возникает замешательство. Ваши глаза сканируют все детали картинки. Новая информация полностью поступает на вершину вашей кортикальной иерархии. Высокоуровневый кортекс пробует множество различных гипотез, но, когда эти предсказания идут вниз по иерархии, каждая из них конфликтует с поступающей информацией и кортекс вынужден пробовать снова. В процессе этого замешательства ваш мозг полностью занят попытками понять картинку. В конце концов, вы делаете высокоуровневое предсказание, которое оказывается правильным. Когда это происходит, предсказание начинается на верхушке кортикальной иерархии и успешно распространяется все ниже и ниже. Менее чем за секунду, каждая область получает последовательность, которая удовлетворяет входным данным. Больше ни одна ошибка не поднимается наверх. Вы понимаете картинку, вы видите далматинца среди точек и каракуль (смотрите рисунок 12).



**Рисунок 12. Вы видите далматинца?**

## 6.9. Действительно ли обратные связи могут делать это?

За десятилетия стало известно, что соединения в кортикальной иерархии являются реципрокными. Если область А проецируется на область Б, то и Б проецируется на А. Часто бывает, что больше аксонов идет вниз, чем вверх. Но даже хотя это описание является широко принятым, превалирующей парадигмой является то, что играют второстепенную, или «модулирующую» роль в мозге. Идея, что сигнал в обратных связях мог бы немедленно и точно вызывать возбуждение различных множеств нейронов в слое 2, не является превалирующей точной зрения среди нейрофизиологов.

Почему бы так могло быть? Частично причина в том, как я упомянул раньше, что нет реальной необходимости привлекать обратные связи, если вы не принимаете центральную роль предсказания. Если вы считаете, что информация течет напрямую в моторную систему, то зачем вам нужны обратные связи? Другая причина игнорирования обратных связей в том, что сигнал в обратных связях распределен по большому числу областей слоя 1. Интуитивно мы должны были бы ожидать, что сигнал, рассеянный по большой площади мог бы иметь только второстепенный эффект на большинстве нейронов, и несомненно в мозгу есть несколько таких модулирующих сигналов, которые не воздействуют на конкретный нейрон, а изменяют глобальные атрибуты, такие как уровень внимания.

Последняя причина игнорирования обратных связей в том, сколько ученых верят в независимую работу нейронов. Обычный нейрон имеет тысячи или десятки тысяч синапсов. Некоторые расположены очень далеко от тела нейрона, другие почти рядом с ним. Синапсы ближе к телу клетки имеют сильное влияние на возбуждение нейрона. Примерно с десяток синапсов возле тела нейрона могут заставить его сгенерировать спайк или импульс электрического разряда. Это известный факт. Однако, подавляющее большинство синапсов располагаются вдали от тела нейрона. Они распространяются вдаль и вширь по древоподобной структуре дендритов нейрона. Поскольку эти синапсы далеки от тела нейрона, ученые склоняются к уверенности, что спайк, прибывающий на один из таких синапсов, имел бы слабый или почти незаметный эффект на то, сгенерирует нейрон спайк или нет. Эффект от отдаленных синапсов рассеивался бы за то время, пока он достигнет тела нейрона.

Как правило, информация, идущая вверх по кортикальной иерархии, передается по синапсам, расположенным близко к телу нейрона. Информация, идущая вверх по иерархии, имеет, следовательно, более определенный путь от области к области. Также, как правило, обратные связи идущие вниз по кортикальной иерархии, используют для этого синапсы, удаленные от тела нейрона. Нейроны в слоях 2, 3 и 5 посылают дендриты в слой 1 и формируют там множество синапсов. Слой 1 – это масса синапсов, но все они от далеких нейронов из слоев 2, 3 и 5. Более того, любой конкретный нейрон, скажем, в слое 2 будет формировать совсем мало синапсов с помощью конкретного волокна обратной связи. Таким образом, некоторые ученые могут возражать идее, что короткий паттерн в слое 1 мог бы точно вызывать возбуждение набора нейронов в слоях 2, 3 и 5. Но именно это соответствует имеющейся у меня теории.

Разрешение это дилеммы в том, что нейроны ведут себя не так, как в классической модели. Фактически, в последние годы становилось все больше ученых, предполагающих, что синапсы на удаленных тонких дендритах могут играть активную и особую роль в возбуждении нейронов. В этих моделях удаленные синапсы ведут себя не так, как синапсы на толстых дендритах возле тела нейрона. Например, если два синапса располагаются очень близко на тонком дендрите, они могут выступать как «детектор совпадений». То есть, если бы оба синапса получали входной импульс в течение короткого промежутка времени, они могли бы вызывать значительный эффект на нейрон, даже если они далеко от тела нейрона. Они могли бы заставить тело нейрона сгенерировать спайк. Как ведут себя дендриты нейрона до сих пор является загадкой, так что я не могу сказать что-нибудь еще о них. Что важно, так это то, что модель кортекса «память-предсказание» требует, чтобы синапсы, далекие от тела нейрона были способны обнаруживать специфические паттерны.

Глядя в прошлое, мне кажется почти безрассудным говорить, что тысячи синапсов нейрона играют только модулирующую роль. Многочисленные обратные связи и огромное число синапсов существуют с вполне конкретной целью. Используя эту догадку, мы можем сказать, что типичный нейрон имеет способность запоминать сотни точных совпадений на волокнах обратных связей, если они формируют синапсы на тонких дендритах. Это значит, что каждая колонка в неокортексе очень пластична в том смысле, какие паттерны обратных связей могут ее активизировать. Это значит, что каждое конкретное свойство может быть точно ассоциировано с тысячами различных объектов и последовательностей. Моя модель требует, чтобы обратные связи были быстродействующими и точными. Нейроны должны возбуждаться, если они чувствуют некоторое количество совпадений на их удаленных дендритах. Новые модели нейронов должны допускать такое.

## 6.10. Как обучается кортекс

У всех нейронов во всех слоях кортекса есть синапсы, и большинство этих синапсов может изменяться при обучении. Будет корректно сказать, что обучение и память возникают во всех слоях, вот всех колонках и во всех областях кортекса.

Ранее в книге я упоминал правило Хебба, названное в честь канадского нейропсихолога Дональда О. Хебба. Его формулировка проста: Когда два нейрона возбуждаются одновременно, синапсы между ними усиливаются. (Коротко и точно это передается фразой «Fire together, wire together»). Мы знаем, что Хебб был в основном прав. Конечно, в природе не все так просто, и в действительности детали реального мозга более сложны. Наша нервная система использует множество вариаций правила Хебба; например, некоторые синапсы изменяют свою силу в ответ на небольшие изменения во времени прихода сигнала, некоторые синаптические изменения короткоживущие, некоторые долгоживущие. Но Хебб создал только основу для изучения памяти, не законченную теорию, и эта основа невероятно полезная.

Принципы обучения по Хеббу могут объяснить большую часть того кортикального поведения, которое я упомянул в этой главе. Вспомните, еще в 1970 году устройства автоассоциативной памяти, использовавшие классические алгоритмы обучения по Хеббу, могли запоминать пространственные паттерны и последовательности паттернов. Основная проблема была в том, что те устройства не могли нормально работать с вариациями образов. Согласно теории, предложенной в этой книге, кортекс обошел это ограничение путем использования иерархической стопки из модулей автоассоциативной памяти, и частично, за счет использования усложненной колончатой архитектуры. В этой главе было почти все о иерархии и о том, как она работает, потому что иерархия это то, в чем заключается сила кортекса. Так что вместо того, чтоб идти через детали того, как каждый нейрон запоминает то или это, я хочу охватить более общие принципы обучения в иерархии.

Когда человек рождается, его кортекс практически ничего не знает. Он не знает ни про язык, ни про культуру, ни про дом, ни про город, ни про людей, вместе с которыми он вырастет, ничего. Вся эта информация, структура мира, должна быть выучена. Двумя базовыми компонентами обучения являются классификация паттернов и построение последовательностей. Эти две комплементарные компоненты памяти взаимодействуют друг с другом. Когда одна область запоминает последовательность, информация, которую она посылает в слой 4 в вышестоящую кортикальную область, изменяется. Таким образом, нейроны слоя 4 обучаются формировать новую классификацию, которая изменяет паттерны, проецируемые обратно в слой 1 нижестоящей области, на которую влияет последовательность.

Основа формирования последовательностей – это группирование паттернов, которые являются частью одного объекта. Один способ сделать это заключается в группировании паттернов, которые возникают рядом во времени. Если ребенок держит игрушку в его руке и медленно двигает ее, его мозг может вполне законно предположить, что изображение на сетчатке от момента к моменту относятся к одному и тому же объекту, и, следовательно, изменяющееся множество паттернов может быть сгруппировано. Иногда вам нужны дополнительные сведения, чтоб принять решение, какие паттерны должны принадлежать к одной группе. Чтобы запомнить, что яблоки и бананы являются фруктами, а морковь и сельдерей не являются, требуется учитель, который поможет сгруппировать эти вещи. В любом случае ваш мозг медленно строит последовательности паттернов, которые должны быть вместе. Но когда область кортекса строит последовательности, информация, поступающая к вышестоящим областям, меняется. Информация изменяется от представления индивидуальных паттернов к представлению групп паттернов. Информация, поступающая в область, меняется с нот на мелодии, с букв на слова, с носов на лица и так далее. Поскольку информация, идущая снизу вверх, становится более «объектно-ориентированной», вышестоящие области кортекса могут теперь запоминать последовательности этих высокоуровневых объектов. Там, где раньше область строила последовательность букв, теперь она строит последовательность слов. Неожиданным результатом такого процесса обучения является то, что в процессе повторяющегося обучения представление объектов продвигается вниз по кортикальной иерархии. В течение первых лет жизни ваши знания о мире формируются сначала в вышестоящих областях кортекса, но по мере обучения они реформируются все ниже и ниже по кортикальной иерархии. Мозг не двигает знания; он переобучается снова и снова. (Я не говорю, что все знания начинаются на верхушке кортекса. В действительности формирование знаний более сложно. Я считаю, что классификация паттернов в слое 4 начинается внизу и движется вверх. Но когда это происходит, мы начинаем формировать последовательности, которые движутся вниз. Я полагаю, что эта память о последовательности переформируется все ниже и ниже по кортексу.) По мере продвижения простых последовательностей вниз, верхние области получают возможность запоминать более сложные и утонченные паттерны.

Вы можете наблюдать создание и движение иерархической памяти вниз на примере обучения ребенка. Рассмотрим, как мы учимся читать. Первое, чему мы учимся – это узнавать отдельные печатные буквы. Это медленная и сложная задача, требующая сознательных усилий. Затем мы переходим к узнаванию простых слов. И снова это сложно и медленно поначалу даже для трехбуквенных слов. Ребенок может читать каждую букву в последовательности и произносить буквы одну за другой, но необходимо определенное количество практических занятий, прежде чем слова сами по себе будут распознаваться как слова. После обучения чтению простых слов мы приступаем к многосложным словам. Поначалу мы произносим каждый слог, соединяя их, как мы делали это с буквами, когда учили простые слова. Через годы практики человек может читать быстро. Мы переходим к тому, что мы в действительности не видим отдельных букв, а вместо этого распознаем целые слова и часто целые фразы одним взглядом. Мы не просто быстрее стали читать; мы действительно распознаем слова и фразы как одно целое. Когда мы читаем одновременно слова целиком, видим ли мы буквы? И да и нет. Очевидно, что сетчатка видит буквы, и область V1 соответственно. Но распознавание букв возникает довольно низко по кортикальной иерархии, скажем, в областях V2 или V4. Когда сигнал достигает IT, отдельные буквы уже не представлены. Что поначалу требует усилий целого визуального кортекса – распознавание отдельных букв – теперь возникает ближе к сенсорному входу. По мере продвижения простых объектов, вроде букв, вниз по иерархии, верхние области получают возможность запоминать сложные объекты, такие как слова или фразы.

Другой пример – обучение чтению музыки с листа. Поначалу вам приходится концентрироваться на каждой ноте. Практикуясь, вы начинаете узнавать общие нотные последовательности, затем целые фразы. После достаточно долгой практики вы как бы совсем не замечаете отдельных нот. Нотная запись только напоминает вам о глобальной структуре пьесы; детальные последовательности запоминаются на самых нижних уровнях иерархии. Этот тип обучения возникает и в моторной, и в сенсорной областях.

Необученный мозг медленнее распознает информацию и медленнее формирует моторные команды, потому что воспоминания, используемые в этих задачах, находятся в нем на верхних уровнях иерархии. Информация должна пройти весь путь вверх и вниз, возможно многочисленными путями, чтобы разрешить конфликты. Нейронным сигналам требуется время, чтобы пропутешествовать вверх и вниз по кортикальной иерархии. Необученный мозг еще не сформировал сложные последовательности на верхнем уровне, и, следовательно, не может распознавать и воспроизводить сложные паттерны. Необученный мозг не может понимать высокоуровневую структуру мира. По сравнению с взрослыми, речь ребенка проста, его музыка проста и его социальные взаимодействия просты.

Если вы изучаете определенный набор объектов снова и снова, ваш кортекс переформирует воспоминания для представления этих объектов в нижней части иерархии. Это освобождает верхнюю часть для изучения более тонких, более сложных отношений. Согласно теории именно так появляются эксперты.

В моей работе по разработке компьютеров, некоторые люди удивляются тому, как быстро я могу взглянуть на изделие и увидеть проблемы, присущие его конструкции. После 25 лет разработки компьютеров у меня сформировалась более-чем-средняя модель следствий, ассоциированных с мобильными вычислительными устройствами. Аналогично, опытные родители могут с легкостью узнать, отчего расстроился их ребенок, тогда как начинающие родители могут прилагать все усилия для разрешения ситуации. Опытный бизнес-менеджер может с легкостью увидеть недостатки и преимущества структуры организации, тогда как новичок просто не может понять этих вещей. У них одна и та же информация, но модель новичка недостаточно сложна. Во всех этих и тысячах других случаев мы начинаем обучение с базовых, простых структур. С течением времени мы продвигаем наши знания вниз по кортикальной иерархии и, следовательно, мы получаем возможность изучения высокоуровневых структур. Именно знание высокоуровневых структур делает нас опытными. У экспертов и гениев мозг видит структуру структур и паттерны паттернов, чего не могут другие. Вы можете стать экспертом практикуясь, но определенно существует генетическая компонента и для талантов и для гениев.

## 6.11. Гиппокамп: вершина всего этого

Под кортикальным слоем лежат и взаимодействуют с ним три больших мозговых структуры. Это базальные ганглии, мозжечок и гиппокамп. Все три существовали до неокортекса. В общем мы можем сказать, что базальные ганглии это примитивная моторная система, мозжечок - изучает точные временные соотношения между событиями, а гиппокамп – хранит память о конкретных местах и событиях. В некоторой степени неокортекс взял на себя их исходные функции. Например, человек, рожденный без большей части мозжечка, будет испытывать трудности в определении временных интервалов и должен будет применять больше сознательных усилий при движении, но во всем остальном он будет практически нормальным.

Мы знаем, что неокортекс отвечает за все сложные моторные последовательности и может напрямую контролировать конечности. Это не значит, что базальные ганглии не важны, просто неокортекс берет на себя большую часть моторного управления. Поэтому я описал общие функции неокортекса независимо от базальных ганглий и мозжечка. Некоторые ученые могут не согласиться с этим предположением, но именно его я использовал в этой книге и своей работе.

С гиппокамп, однако, другое дело. Это одна из наиболее глубоко изученных областей мозга, потому что в нем суть формирования воспоминаний. Если вы потеряете обе половинки гиппокампа (как и многие другие части нервной системы, он существует и в левой, и правой стороне мозга), вы потеряете способность формировать новые воспоминания. Без гиппокампа вы сможете продолжать говорить, ходить, видеть и слышать, и на коротких промежутках времени будете казаться нормальным. Но, фактически, вы глубоко неполноценны: вы не можете запомнить ничего нового. Вы можете помнить друзей, которых вы знали до потери гиппокампа, но вы не можете запомнить ни одного нового человека. Даже если вы встречаете доктора пять раз в день в течение года, каждый раз был бы для вас как первый раз. У вас не было бы воспоминаний о событиях, которые произошли после потери вами гиппокампа.

На протяжении многих лет не хотелось думать о гиппокампе, потому что он не имел для меня смысла. Понятно, что он необходим для обучения, но он не является конечным хранилищем большинства наших знаний. Им является неокортекс. Классическая точка зрения на гиппокамп заключается в том, что там формируются знания, а затем в течение нескольких дней недель или месяцев эти новые знания переносятся в неокортекс. Это мне ничего не говорило. Мы знаем, что зрение, слух, осязание – все наши сенсорные потоки – поступают прямо в сенсорные области кортекса без предварительного прохождения через гиппокамп. Мне казалось, что эта сенсорная информация должна автоматически формировать новые знания в кортексе. Почему же нам для обучения нужен гиппокамп? Как такая отдельная структура, как гиппокамп, может взаимодействовать и опережать обучение в кортексе, только в последствии передавая информацию обратно в кортекс?

Я решил отложить гиппокамп, надеясь, что наступит день, когда его роль для меня прояснится. Этот день настал в 2002 году, практически в то время, когда я только начал писать эту книгу. Один мой коллега из Института Нейронаук в Редвуде, Бруно Ольшозен, указал, что соединения между гиппокампом и неокортексом наводят на мысль, что гиппокамп является самой верхней областью кортекса, а не отдельной структурой. С этой точки зрения гиппокамп занимает верхушку неокортикальной пирамиды, самый верхний блок на рисунке 5. Неокортекс появился на эволюционной сцене в промежутке между гиппокампом и остальными структурами мозга. Несомненно, такая точка зрения на гиппокамп, как на вершину кортикальной иерархии, был уже известен; просто я об этом не знал. Я поговорил с несколькими экспертами по гиппокампу и попросил их объяснить, как эта структура в форме морского конька может предавать знания в кортекс? Никто не смог объяснить. И никто не упомянул о том, что гиппокамп является вершиной кортикальной пирамиды, возможно потому что гиппокамп не только сидит на верхушке кортикальной пирамиды, но также напрямую соединяется с многими другими частями мозга.

Но я тут же увидел новую перспективу, как решение моего замешательства.

Подумайте над информационными потоками от глаз, ушей и кожи в неокортекс. Каждая область неокортекса пытается понять, что обозначает информация. Каждая область пытается понять информацию в терминах известных ей последовательностей. Если она понимает информацию, она говорит: «я поняла, это просто часть объекта, на который я уже смотрю. Я не буду сообщать о деталях». Если область не понимает текущую информацию, она передает ее вверх по иерархии до тех пор, пока не поймет одна из вышестоящих областей. Однако паттерн, который действительно совершенно новый, будет подниматься все выше и выше по иерархии. Последовательно каждая вышестоящая область говорит: «я не знаю, что это, я не предвидела этого, почему бы тебе, вышестоящая область, не взглянуть на это?». Загвоздка в том, что когда вы доберетесь до верхушки кортикальной пирамиды, все, что останется – это информация, которая не может быть понята через предыдущий опыт. У вас осталась действительно новая и неожиданная часть информации.

В повседневности мы встречаем множество новых вещей, поступающих на вершину кортикальной иерархии – например, статья в газете, имя человека, которого вы встретили сегодня утром, автомобильная авария, которую вы видели по пути домой. Именно эти необъясненные и непредсказуемые остатки информации, новый материал поступает в гиппокамп и сохраняется там. Эта информация не будет храниться вечно. Либо она будет передана в кортекс, либо, соответственно, будет потеряна.

Я заметил, что по мере старения я начинаю испытывать проблемы с запоминанием новых вещей. Например, мои дети помнят детали большинства театральных спектаклей, которые они видели за последний год. Я не помню. Возможно, это потому что я видел так много спектаклей за свою жизнь, что я не вижу ничего действительно нового. Новые спектакли укладываются в мои воспоминания о предыдущих спектаклях, и информация просто не попадает в мой гиппокамп. Для моих детей каждый спектакль содержит гораздо больше новой информации и достигает гиппокампа. Если это так, мы можем сказать, что чем больше мы знаем, тем меньше мы запоминаем.

В отличие от неокортекса, у гиппокампа гетерогенная структура из нескольких специализированных областей. Она великолепно подходит для задачи быстрого сохранения поступающих паттернов. Гиппокамп находится в отличной позиции, на верхушке кортикальной пирамиды, чтобы запоминать то, что является новым. Он также находится в отличной позиции для того, чтоб вспоминать эту новую информацию, позволяя ей сохраняться в кортикальной иерархии, что в некоторой степени является медленным процессом. Вы можете мгновенно запомнить новое событие в гиппокампе, но вы запомните что-либо в кортексе только если вы ощущаете это снова и снова, либо в реальности, либо думая об этом.

## 6.12. Альтернативный путь по иерархии

В кортексе есть другой значительный путь для передачи информации от области к области вверх по иерархии. Этот альтернативный путь начинается в нейронах слоя 5, которые проецируются в таламус (отличие одних частей таламуса от других мы обсуждали ранее), и затем из таламуса в вышестоящую область кортекса. Если две области кортекса соединяются напрямую в иерархическом стиле, они также соединяются косвенно через таламус. Этот второй путь передает информацию только вверх по иерархии, но не в низ. Таким образом, при движении вверх по кортикальной иерархии есть прямой путь между двумя областями и косвенный путь через таламус.

Второй путь имеет два режима функционирования, определяемые нейронами таламуса. В одном режиме путь практически не работает, так что информация по нему не идет. В другом режиме информация идет точно между двумя областями. Двое ученых, Мюррей Шерман из Нью-Йоркского Государственного Университета и Рей Гиллери из Медицинского Университета Висконсина описали альтернативный путь и постулировали, что он может быть настолько же важным, как и прямой путь (возможно даже больше), что и является темой данной главы. У меня есть предположения о том, для чего нужен этот второй путь.

Прочтите это слово: *imagination*. Большинство людей может прочесть это слово с первого взгляда. Теперь взгляните на букву *i* в середине слова. Теперь взгляните на точку над *i*. Ваши глаза могут смотреть в одно и то же место, но в одном случае вы видите слово, в другом – букву, в последнем случае вы видите точку. Вглядитесь пристально в букву *i* и попытайтесь переключать ваше восприятие между словом, буквой и точкой. Если у вас не получается, попробуйте произносить «точка», «*i*» и «*imagination»*, пока вглядываетесь в точку. Во всех случаях одна и та же информация поступает в V1, но по мере того, как она достигает области IT, вы воспринимаете различные вещи, различные уровни детализации. Область IT знает, как распознать все три объекта. Она может распознать отдельную точку, букву *i*, и слово целиком. Но когда вы воспринимаете слово целиком, V4, V2 и V1 обрабатывают детали и все, о чем знает IT – это слово. Обычно вы не воспринимаете отдельные буквы при чтении; вы воспринимаете слова или фразы. Но вы можете воспринимать буквы, если захотите. Мы занимаемся таким сдвигом внимания все время, но обычно мы это не осознаем. Я могу слушать музыку и воспринимать только мелодию, но если я попытаюсь, я могу выделить певца или бас-гитару. В мои уши поступают одни и те же звуки, но я могу фокусировать мое восприятие. Каждый раз, когда вы чешете голову, движение вызывает внутри громкий звук, но обычно вы не осознаете его. Однако, если сфокусироваться на нем, вы можете отчетливо услышать звук. Это другой пример того, как сенсорная информация, которая обычно обрабатывается нижними уровнями кортикальной иерархии, может быть перенесена на высшие уровни если вы обращаете на нее внимание.

Я полагаю, что альтернативный путь через таламус – это механизм, посредством которого мы обращаем внимание на детали, которые обычно мы не замечаем. Он пропускает группирование последовательностей в слое 2, посылая сырые данные в вышестоящую область кортекса. Биологи показали, что альтернативный путь может включаться одним из двух способов. Один из них – это сигнал из вышестоящей области самого кортекса. Этот способ вы использовали, когда я попросил вас обращать внимание на детали, которые вы обычно не замечаете, такие как точка над *i* или звук почесывания головы. Второй способ активировать этот путь – это сильный неожиданный сигнал снизу. Если сигнал на альтернативный путь достаточно сильны, он посылает вышестоящей области сигнал пробуждения, которая снова может включить альтернативный путь. Например, если бы я показал вам лицо и спросил, что это было, вы бы ответили «Лицо». Если бы я показал вам то же самое лицо, но со странной отметиной на носу, вы сначала узнали бы лицо, но затем немедленно ваши нижние зрительные уровни заметили бы, что что-то не так. Эта ошибка форсирует открытие альтернативного пути. Детали теперь пойдут по альтернативному пути, пропуская группирование, которое обычно возникает, и ваше внимание привлечет отметина. Теперь вы видите отметину, а не просто лицо. Если б отметина было достаточно необычной, она могла бы привлечь все ваше внимание целиком. В этом случае необычные события быстро привлекают ваше внимание. Именно поэтому мы не можем не обращать внимание на уродство или другие необычные паттерны. Ваш мозг делает это автоматически. Однако, часто ошибка недостаточно сильная, чтоб открыть альтернативный канал. Вот почему мы иногда не замечаем, что слово написано с ошибкой.

## 6.13. Завершающие мысли

Чтобы найти и обосновать новую систему научных взглядов, необходимо искать простейшие концепции, способные объединить и объяснить большое количество несопоставимых фактов. Неизбежным следствием такого процесса является то, что уходят в крайность излишнего упрощения. Важные детали вероятно буду проигнорированы и факты будут проинтерпретированы неправильно. Если такая система взглядов становится устоявшейся, неизбежно будут найдены улучшения и исправления, показывая, как далеко зашли начальные предположения, или недостаточно далеко, или вообще были ошибочны.

В этой главе я ввел множество умозрительных идей о том, как работает неокортекс. Я ожидаю, что некоторые из этих идей будут опровергнуты, и, вероятно, каждая из них будет пересмотрена. Есть также множество деталей, о которых я даже не упомянул. Мозг очень сложен; специалисты по нейронаукам, читающие эту книгу, должны знать, что я привел грубую характеристику сложности реального мозга. Но я верю, что в целом озвучил новую систему взглядов. Все, на что я могу надеяться, это то, что основные идеи будут сохранены, и изменятся только детали вопреки новым данным и новому осмыслению.

В конце концов, вам может быть неприемлема идея, что простая, но большая система памяти может действительно в результате вести себя так, как человек. Можем ли вы и я быть просто иерархической системой памяти? Могут ли наши жизнь, верования и желания быть сохраненными в триллионах синапсов? В 1984 я начал профессионально писать программы. До этого я писал маленькие программки, но тогда я в первый раз программировал компьютер с графическим интерфейсом и впервые работал над большим и сложным приложением. Я писал приложение для операционной системы, созданной Grid Systems. С окнами, множественными шрифтами и меню, операционная система Grid была действительно продвинутой для того времени.

Однажды меня поразила практическая невозможность того, что я делаю. Как программист, я написал одну строчку кода. Я сгруппировал строчки кода в блоки, называемые подпрограммами. Подпрограммы были сгруппированы в модули. Модули были скомбинированы, чтоб стать приложениями. В программе для электронных таблиц, над которой я работал, было так много подпрограмм и модулей, что никто не мог понять это полностью. Это было сложно. Хотя каждая строчка кода делала совсем немного. Чтобы нарисовать один пиксел на дисплее, требовалось несколько строчек кода. Чтобы нарисовать целый экран с электронной таблицей, компьютеру требовалось выполнить миллионы инструкций, распределенных по сотням подпрограмм. Подпрограммы использовали другие подпрограммы рекурсивным способом. Это было настолько сложно, что было невозможно знать все, что должно было бы происходить в выполняющейся программе. Меня поразило, насколько невероятно то, что программа за время выполнения рисует картинку, которая нам кажется появившейся мгновенно. Ее внешним проявлением были таблицы чисел, меток, текстов и графов. Она вела себя как электронная таблица. Но я знал, что должно происходить внутри компьютера с процессором, выполняющим одну простую инструкцию в один момент времени. Было трудно поверить, что компьютер смог бы пройти его путь по лабиринту модулей и подпрограмм, и выполнить все эти инструкции настолько быстро. Если бы я не знал подробностей, я был бы уверен, что это не может работать. Я понимал, что если кто-то изобрел концепцию компьютера с графическим интерфейсом и электронными таблицами, и представил ее мне на бумаге, я бы отверг ее как непрактичную. Я сказал бы, что ему потребовалась бы вечность, чтоб что-то сделать. Это было унизительной мыслью, потому что это работало. Затем я понял, что мое интуитивное ощущение скорости микропроцессора и мое ощущение силы иерархического дизайна были неадекватными.

В этом был урок про неокортекс. Он не был сделан из сверхбыстрых компонент и правила, по которым он работает, не сложны. Однако он имеет иерархическую структуру, содержащую миллиарды нейронов и триллионы синапсов. Если для нас сложно вообразить, как такая логически простая, но численно громадная система памяти может создавать наше сознание, наш язык, нашу культуру, наше искусство, эту книгу, нашу науку и технологию, то это потому, предположил я, что наше интуитивное ощущение емкости неокортекса и силы его иерархической структуры являются неадекватными. Неокортекс работает. Это не волшебство. Мы можем понять его. И, подобно компьютеру, в конце концов мы сможем построить интеллектуальные машины, которые будут работать по тем же самым принципам.

# 7. Сознание и творчество

**Когда** я веду разговор о моей теории мозга, слушатели обычно быстро схватывают значимость предсказаний, как они соотносятся с человеческим поведением. Они задают множество сопутствующих вопросов. Откуда берется творчество? Что такое сознание? Что такое воображение? Как мы можем отделить реальность он неправильных представлений? Хотя эти темы преимущественно не были тем, что мотивировало меня на изучение мозга, они интересуют практически каждого. Я не претендую на звание эксперта по этим темам, но теория «память-предсказание» может дать некоторые ответы и полезные мысли. В этой главе я обращусь к некоторым из наиболее часто задаваемым вопросам.

## 7.1. Обладают ли животные интеллектом?

Обладает ли крыса интеллектом? Обладает ли им кошка? Когда интеллект появился на эволюционной шкале? Мне нравится этот вопрос, потому что ответ на него я нахожу неожиданным.

Все, что я написал до этого про неокортекс и его работу, основано на самой базовой предпосылке – что у мира есть структура и следовательно он предсказуем. В мире есть паттерны: на лице есть глаза, у глаз есть зрачки, огонь обжигает, гравитация заставляет объекты падать, дверь открывается и закрывается, и так далее. Мир не случаен, но и не однороден. Память, предсказание и поведение мало бы что значили, если б у мира не было структуры. Любое поведение, поведение ли человека, или улитки, или одноклеточного организма или дерева – это средство использования структуры мира для осуществления размножения.

Вообразите одноклеточное, живущее в водоеме. У клетки есть жгутик, позволяющий ей плавать. На поверхности клетки есть молекулы, реагирующие на присутствие питательных веществ. Поскольку не все области водоема имеют одинаковую концентрацию питательных веществ, есть градиент питательных веществ с одной стороны клетки на другую. Когда она плывет через пруд, клетка может обнаруживать изменение. Это простая структура мира одноклеточного. Клетка эксплуатирует свою химическую осведомленность, плывя в места с высокой концентрацией питательных веществ. Мы могли бы сказать, что этот простой организм делает предсказание. Он предсказывает, что плавая определенным образом, он найдет больше питательных веществ. Задействована ли в этом предсказании память? Да, задействована. Память заключается в ДНК организма. Одноклеточные не обучаются в течение жизни, как использовать этот градиент. Наоборот, обучение возникает в течение эволюции и сохраняется в ДНК. Если внезапно изменилась бы структура мира, одноклеточные не смогли бы научиться и адаптироваться. Они не смогли бы изменить ДНК или свое поведение. Для этих видов обучение может возникать только в эволюционном процессе на многих поколениях.

Есть ли интеллект у одноклеточного? Используя обыденное понимание человеком интеллекта, мы ответим «нет». Но животные находятся на дальнем конце континуума видов, использующих память и предсказание для того, чтоб размножаться более успешно, и по академическим соображениям ответ должен быть «да». Суть не в том, чтоб сказать, какие виды интеллектуальны, а какие – нет. Память и предсказание используются всеми живыми формами. Просто есть бесчисленное множество методов и ухищрений, как они делают это.

Растения также используют память и предсказание для эксплуатации структуры мира. Деревья занимаются предсказанием, когда посылают свои корни в почву и ветви и листья в небо. Деревья предсказывают, где они найдут воду и минералы, основываясь на опыте своих предков. Конечно, дерево не думает; его поведение автоматическое. Но они эксплуатируют структуру мира таким же способом, как и одноклеточные. Каждый вид растения имеет различный набор поведенческих действий, который использует немного другую часть структуры мира.

Следовательно, растения развили коммуникационные системы, базирующиеся в основном на медленном выделении химических сигналов. Если насекомое повреждает часть дерева, дерево посылает химические сигналы по его сосудистой системе к другим частям дерева, которые включают защитные системы, такие как выделение токсинов. С помощью такой коммуникационной системы дерево может демонстрировать немного более сложное поведение. Нейроны возможно появились как средство передачи информации более быстрым путем, чем сосудистые системы растений. Мы можем думать о нейроне просто как о клетке с ее собственным сосудистым придатком. С некоторых пор, вместо медленного продвижения химических сигналов вдоль этих придатков, нейроны начали использовать электрохимические импульсы, которые перемещаются гораздо быстрее. Вначале быстрая синаптическая передача и простые нервные системы возможно не были задействованы для обучения. Суть была просто в более быстрой передаче сигналов.

Но затем в марше по эволюционной лестнице случилось нечто действительно интересное. Соединения между нейронами стали изменчивыми. Нейроны могли посылать или не посылать сигнал, в зависимости от того, что только что произошло. Теперь поведение могло быть модифицировано в течение жизни организма. Нервные системы стали пластичными, и таким же стало поведение. Поскольку знания могли быть достаточно быстро сформированы, животное могло изучать структуру мира уже в течение жизни. Если мир внезапно изменялся – скажем, на сцене появлялся новый хищник – животному не надо было оставаться привязанным к своему генетически определенному поведению, которое больше могло уже быть неприемлемым. Пластичные нервные системы стали огромным эволюционным преимуществом и привели к прорыву новых видов, от рыб к пресмыкающимся и к людям.

Как мы увидели в главе 3, у всех млекопитающих есть старый мозг, на поверхности которого располагается неокортекс. Неокортекс – это просто то, во что эволюционировала новейшая нервная ткань. Но с его иерархической структурой, инвариантным представление и предсказанием по аналогии, кортекс позволил млекопитающим эксплуатировать гораздо больше структуры мира, чем могут животные без неокортекса. Наши предки, наделенные неокортексом, смогли вообразить, как сделать сеть и поймать рыбу. Рыба не способна запомнить, что сеть значит смерть, или вообразить, как построить инструмент для разрезания сетей. У всех млекопитающих, от крыс до человека, есть неокортекс. Они интеллектуальны, но в разной степени.

## 7.2. Чем отличается интеллект человека?

Теория «память-предсказание» предлагает два ответа на этот вопрос. Первый почти очевиден: наш неокортекс больше, чем, скажем, у обезьяны или у собаки. Увеличив кортикальную поверхность до размеров обеденной салфетки, наш мозг смог изучать более сложную модель мира и делать более сложные предсказания. Мы видим более глубокие аналогии, больше структур в структуре, чем другие млекопитающие. Если мы хотим найти супруга, мы не просто смотрим на такие атрибуты, как здоровье, мы разговариваем с его друзьями и родителями, мы наблюдаем, как они себя ведут и разговаривают, и судим, насколько они честны. Мы смотрим на эти вторичные и третичные атрибуты в попытке предсказать, как наш потенциальный супруг поведет себя в будущем. Трейдеры на фондовой бирже смотрят на структуру коммерческих паттернов. Математики ищут структуру в числах и уравнениях. Астрономы ищут структуру в движении планет и звезд. Наш увеличенный неокортекс позволяет нам видеть наш дом, как часть города, который является частью региона, который является частью планеты, которая является частью большой вселенной – структуру в структуре. Никакие другие млекопитающие не могут размышлять так глубоко. Я почти уверен, что у моей кошки нет концепции мира за пределами нашего дома.

Второе отличие интеллекта человека от интеллекта других животных в том, что у нас есть язык. Целые книги написаны о предположительно уникальных свойствах языка и о том, как он развивался. Однако язык очень хорошо укладывается в модель «память-предсказание» без какого либо специального «языкового соуса» или специальной языковой машины. Произнесенные и написанные слова – это просто паттерны мира, также как мелодии, машины и дома. Синтаксис и семантика языка не отличаются от иерархической структуры других повседневных объектов. И тем же самым способом, как мы ассоциируем звук поезда с визуальной картинкой поезда, мы ассоциируем произнесенные слова с нашими знаниями о их физических и семантических составляющих. С помощью языка человек может вызывать воспоминания и создавать новое сопоставление ментальных объектов других людей. Язык является чистой аналогией, и с его помощью мы можем заставлять других людей получать знания и изучать вещи, которые они никогда в действительности не видели. Развитие языка требует большого неокортекса, способного управляться с вложенными структурами синтаксиса и семантики. Он также требует более полно развитого моторного кортекса и мускулатуры, чтобы позволить нам создавать сложные, сильно артикулированные звуки или жесты. С помощью языка мы можем брать паттерны, изученные в течение жизни и передавать их нашим детям или нашему племени. Язык, письменный, устный или язык тела, становится средством, с помощью которого мы передаем от поколения к поколению то, что мы знаем о мире. Сегодня, печатные и электронные коммуникации позволяют нам разделять наши знания с миллионами людей в мире. Животные без языка почти не передают столько информации своим потомкам. Крыса может запомнить множество паттернов в течение своей жизни, но они не передают в деталях новую информацию – «Эй, сынок, вот как мой отец научил меня избегать электрических ударов».

Таким образом, интеллект мог бы быть отслежен в течение трех эпох, в каждой из которых использовались память или предсказание. Первая – это когда виды использовали ДНК в качестве носителя памяти. Отдельные особи не могли адаптироваться в течение жизни. Они могли только передавать своим потомкам через их гены память о мире, базирующуюся на ДНК.

Вторая эпоха началась, когда природа изобрела модифицируемые нервные системы, которые могли быстро формировать знания. Отдельные особи теперь могли изучать структуру их мира и адаптировать свое поведение в течение жизни. Но отдельные животные все еще не могли передавать эти знания своим потомкам каким либо путем, кроме прямого наблюдения. Создание и расширение неокортекса произошло во время второй эпохи, но не определяло ее.

Третья эпоха уникальна появлением человека. Она началась с изобретения языка и расширением нашего неокортекса. Мы, люди, можем изучить очень большую часть структуры мира в течение нашей жизни, и мы можем эффективно передать эти знания другим людям посредством языка. Вы и я участвуем в этом процессе прямо сейчас. Я потратил большую часть моей жизни на поиски структуры в мозгу и того, как эта структура приводит к мышлению и интеллекту. Все, что я изучил, распределено для вас по этой книге. Конечно, я не смог бы сделать этого, если бы не имел доступа к знаниям, собранным сотнями ученых, которые также узнали что-то от других, и так далее вглубь времен. Я смог усвоить и добавить к тому, что другие написали об их собственных мыслях и наблюдениях.

МЫ стали самыми приспосабливающимися существами на этой планете и единственными со способностью передавать наши знания о мире широко по нашей популяции. Человеческая популяция пережила взрывообразный рост, потому что мы изучали и эксплуатировали довольно большую часть структуры мира и передавали ее другим людям. Мы можем выжить везде, в дождливом лесу, в пустыне, в морозной тундре, или в крупных городах. Комбинация большого неокортекса и языка привела к постепенному успеху нашего вида.

## 7.3. Что такое творчество?

Меня часто спрашивают о творчестве, я подозреваю потому что большинство людей рассматривает творчество как нечто, чего не может машина, и следовательно это вызов самой идее построения интеллектуальных машин. Что такое творчество? Мы уже встречались с ответом несколько раз в этой книге. Творчество не является чем-то, что возникает в определенных областях кортекса. Оно не является подобным ни эмоциям, ни равновесию, за которыми закреплены определенные структуры и контуры за пределами кортекса. Наоборот, творчество это неотъемлемое свойство каждой кортикальной области. Это необходимый компонент предсказания.

Как это может быть истинным? Разве творчество не является некоторым экстраординарным качеством, требующим высокого интеллекта и таланта? Совершенно не обязательно. Творчество может быть определено просто как процесс предсказания по аналогии, нечто, что возникает повсюду в кортексе и чем вы постоянно занимаетесь, пока бодрствуете. Творчество возникает непрерывно. Оно простирается от простых повседневных актов восприятия, возникающих в сенсорных областях кортекса (когда мы слышим песню в новой тональности) до сложных, редких актов гениальности, возникающих в высших областях кортекса (создание симфонии в совершенно новом стиле). На фундаментальном уровне повседневные акты восприятия подобны редким полетам мысли. Просто повседневные акты настолько банальны, что мы не замечаем их.

Теперь у вас есть базовое понимание того, как мы создаем инвариантные знания, как мы используем инвариантные знания для предсказаний, и как мы делаем предсказания будущих событий, которые всегда в чем-то отличаются от того, что мы видели ранее. Вспомните также, что инвариантные знания это последовательности событий. Мы делаем предсказания путем комбинирования инвариантных знаний того, что должно произойти далее с деталями, относящимися к настоящему моменту времени (вспомните притчу о предсказании времени прибытия поезда). Предсказание это применение инвариантных знаний к новой ситуации. Следовательно, все кортикальные предсказания – это предсказания по аналогии. Мы предсказываем будущее по аналогии с прошлым.

Вообразите, что вам предстоит ужин в незнакомом ресторане и вы хотите помыть руки. Даже если вы никогда ранее не были в этом здании, ваш мозг предсказывает, что в ресторане где-то должен быть туалет с раковиной для мытья рук. Откуда он это знает? В других ресторанах, в которых вы бывали, всегда был туалет, и по аналогии в этом ресторане он также должен быть. Более того, вы знаете где и что искать. Вы предсказываете, что должна быть дверь или знак с соответствующими символами, ассоциирующимися с мужчинами и женщинами. Вы предсказываете, что он должен находиться где-то в задней части ресторана, либо за баром, либо возле гардероба, но не в пределах прямой видимости залов для приема пищи. Опять же, вы никогда не были в этом конкретном ресторане, но по аналоги с другими заведениями общепита вы способны найти то, что вам нужно. Вы не бродите в поисках произвольно. Вы ищите ожидаемые паттерны, которые позволят вам быстро найти туалет. Этот тип поведения является творческим актом; это предсказание будущего по аналоги с прошлым. Обычно мы не задумываемся о том, что это творческий процесс, но в большой степени это именно так.

Недавно я купил виброфон. У нас есть пианино, но до этого я никогда не играл на виброфоне. В тот день, когда мы принесли виброфон, я взял ноты для пианино, поставил их на подставку виброфона и начал играть простые мелодии. Моя способность сделать это не является чем-то выдающимся. Но на фундаментальном уровне, это был творческий акт. Подумайте над тем, что здесь было задействовано. Я взял инструмент, который очень сильно отличается от пианино. У виброфона золотистые металлические пластины; у пианино – черные и белые клавиши. Золотистые пластины большие и постепенно изменяются по размеру; клавиши пианино маленькие и двух различных размеров. Золотистые пластины размещены в два различных ряда; черные и белые клавиши – перемежаются. На одном инструменте я использую пальцы, а на другом – я размахиваю молоточками. Перед этим я стою, перед тем – сижу. Конкретные мышцы и движения, необходимые для игры на виброфоне, совершенно отличаются от тех, что нужны для игры на пианино.

Так как же я смог сыграть мелодию на незнакомом инструменте? Ответ в том, что мой кортекс видит аналогии между клавишами пианино и пластинами виброфона. Использование этого подобия позволило мне сыграть мелодию. Это практически не отличается от того, чтоб спеть песню в другой тональности. В обоих случаях, мы знаем, что нужно делать, по аналоги с прошлым опытом. Я понимаю, что для вас подобие этих двух инструментов может казаться очевидным, но только потому, что наш мозг автоматически видит аналогии. Попытайтесь запрограммировать компьютер, чтобы он нашел подобие между такими объектами, как пианино и виброфон, и вы увидите, насколько невероятно сложно это. Предсказание по аналогии – творчество – настолько распространено, что обычно мы не замечаем его.

Однако, мы считаем, что творчество – это когда система «память-предсказание» работает на высшем уровне абстракции, когда она делает необычные предсказания, используя необычные аналогии. Например, большинство людей согласились бы, что математик, доказывающий сложную теорему, занимается творчеством. Но давайте взглянем поближе, что вовлечено в его ментальные усилия. Наш математик пристально глядит на уравнения и говорит: «Как же мне взяться за эту проблему?». Если ответ не вполне очевиден, он может переформулировать уравнения. Записывая их в другой форме, он может взглянуть на ту же самую проблему с другой точки зрения. Он вглядывается все больше и больше. Неожиданно он видит, что часть уравнений ему знакома. Он думает: «О, я узнаю это. В этих уравнениях есть структура, которая похожа на структуру других уравнений, над которыми я работал несколько лет назад». Затем он делает предсказание по аналогии. «Возможно, я смогу решить эти новые уравнения, используя ту же самую методику, которую я успешно использовал на старых уравнениях». Он может решить проблему по аналогии с предыдущей проблемой. Это творческий акт.

У моего отца было странное заболевание крови, которое врачи никак не могли диагностировать. Как же они могли узнать, какое лечение предложить? Они рассмотрели данные, взятые из анализов крови моего отца за несколько месяцев, чтобы найти определенные паттерны. (Мой отец напечатал великолепный график, так что врачи могли ясно увидеть необходимые им данные). Хотя симптомы не соответствовали близко какому-либо конкретному заболеванию, некоторые сходства все таки были. Врачи пришли к решению, что лечение должно базироваться на смеси нескольких стратегий, которые подходили для других заболеваний крови. Использованное лечение предположительно было основано на аналогиях к заболеваниям, которые врачи лечили ранее. Узнавание этих паттернов требовало глубокого знания других необычных заболеваний.

Метафоры Шекспира – это образец творчества. "Love is a smoke made with the fume of sighs". "Adversity's sweet milk, philosophy". "There's daggers in men's smiles". Такие метафоры становятся очевидными, когда вы видите их, но их трудно придумать, что является одной из причин, почему Шекспира считают литературным гением. Чтобы создать такие метафоры, он должен был увидеть непрерывный ряд хитрых аналогий. Когда он пишет «There's daggers in men's smiles» («В человеческих улыбках есть кинжалы») он говорит не только о кинжалах или улыбках. Кинжал – это аналогия к дурным намерениям, а человеческие улыбки – это аналогия к обману. Две хитрых аналогии всего лишь в пяти словах! По крайней мере все зависит от их интерпретации. У поэтов есть способность к коррелированию внешне не похожих слов или понятий таким образом, чтоб представить эти слова в новом свете. Они создают неожиданные аналогии как средство для изучения высокоуровневых структур.

Фактически произведения искусства ценятся высоко, потому что они нарушают наши предсказания. Когда вы смотрите фильм, который разрушает известные представления о характерах, сюжетной линии, или кинематографии (включая спецэффекты), он вам нравится, потому что не похож на другие. Живопись, музыка, поэзия, проза – все творческие формы искусства – стараются разрушить обычаи и нарушить ожидания потребителя. Именно во внутреннем противоречии заключается великолепие произведения искусства. Мы хотим, чтоб произведение было понятно, но в тоже время уникально и неожиданно. Слишком понятные произведения неинтересны или кичливы; слишком уникальные – раздражающие и их сложно оценить. Лучшие произведения нарушают некоторые ожидаемые паттерны, одновременно приучая нас к новым. Рассмотрим хорошее произведение классической музыки. Наилучшее произведение привлекает на простом уровне – хороший ритм, простая мелодия и музыкальные обороты. Любой может понять и оценить его. Однако, оно также немного необычно и неожиданно. Но чем больше вы его слушаете, тем больше вы видите паттерны в неожиданных местах, таких как повторение необычных гармоний или изменение тональности. Это же верно для хороших литературных произведений или для хороших фильмов. Чем больше вы читаете или смотрите их, тем больше творческих деталей или сложностей структуры вы наблюдаете.

Возможно у вас был опыт наблюдения таких вещей, когда трудно понять, что же вам знакомо: «Хмм, где то я уже видел этот паттерн раньше, где-то еще…». Вы могли и не задаваться вопросом, просто необычная ситуация активировала инвариантное представление в вашем мозгу. Вы увидели аналогию между двумя обычно несвязными вещами. Я мог бы, например, вдруг увидеть, что возникновение научной идеи аналогично продаже бизнес-идеи, или что проведение политических реформ подобно воспитанию детей. Если я поэт – вуаля! – у меня есть новая метафора. Если я ученый или инженер – у меня есть новое решение для давнишней проблемы. Творчество – это смешивание и сопоставление паттернов для любых вещей, которые вы видите или должны увидеть в вашей жизни. Говорят «this is kinda like that». Для выполнения этой задачи есть нейронные механизмы во всех областях кортекса.

## 7.4. Являются ли одни люди более творческими, чем другие?

Попутный вопрос, который мне часто задают: «Если у всех людей мозг с рождения обладает творческими механизмами, почему у нас различная способность к творчеству?». Теория «память-предсказание» указывает на два возможных ответа. Один связан с природой, другой – с воспитанием.

С точки зрения воспитания у всех различный жизненный опыт. Следовательно, каждый строит отличные от других модели мира и знания о нем в своем кортексе, и будет находить другие аналогии и делать другие предсказания. Если я слушал много музыки, я смогу спеть песню в другой тональности и сыграть простую мелодию на другом инструменте. Если я почти не слушал музыку, я не смогу осуществить этого. Если я изучал физику, я смогу объяснить поведение повседневных объектов на основе законов физики. Если я вырос рядом с собаками, у меня будут способности видеть аналогии относительно собак и я смогу лучше предсказывать их поведение. То, что одни люди более творческие в социальных ситуациях, другие в языках, или в математике, или в дипломатии – все это определяется окружением, в котором они выросли. Наши предсказания, и, следовательно, наши таланты построены на опыте.

В главе 6 я описал, как знания постепенно опускаются вниз по кортикальной иерархии. Чем больше вы подвергаетесь воздействию определенных паттернов, тем больше знаний об этих паттернах реформируются на более низких уровнях. Это позволяет вам изучать отношения на более высоком уровне абстракции в верхней части иерархии. В этом суть того, что называют «быть экспертом». Эксперт – это некто, кто путем практики и постоянного обучения научился узнавать паттерны более тонкие, чем те, что может распознать неспециалист, такие как форма крыла на машинах конца 50-х или размер пятна на клюве чайки. Эксперты могут распознавать паттерны поверх других паттернов. Разумеется, есть физическое ограничение на то, что мы можем запомнить, обусловленное размером кортекса. Но у людей кортекс гораздо больше по сравнению с другими видами и у нас потрясающая гибкость в том, что мы можем запомнить. Все это зависит от того, что нас окружало в нашей жизни.

С природной точки зрения мозг демонстрирует физическую изменчивость. Определенно некоторые из различий предопределены генетически, такие как размер областей (у отдельных людей наблюдается различие размеров области V1 в три раза) и полушарные различия (у женщин имеется тенденция к более толстым соединениям между левой и правой частями мозга, чем у мужчин). У некоторых людей мозг содержит больше нейронов или другие типы соединений. Маловероятно, что творческий гений Альберта Эйнштейна был чисто следствием стимуляции окружения патентного бюро, в котором он работал в юности. Недавние исследования его мозга – который считался потерянным, но был найден в сохранности несколько лет назад – обнаружили, что его мозг количественно был необычным. В нем было больше вспомогательных клеток, называемых глиальными, на один нейрон, чем обычно. Был обнаружен необычный паттерн из борозд в теменных долях – в области, которая считается ответственной за важные математические способности и пространственное мышление. Она была также на 15 процентов шире, чем обычно у других. Возможно мы никогда не узнаем, почему Эйнштейн таким умным и творческим, но можно биться об заклад, что часть его таланта была унаследована через генетические факторы.

Каким бы не было различие между блестящими людьми и обычными – все мы творческие. И путем практики и обучения мы можем улучшить наши знания и таланты.

## 7.5. Можем ли мы натренировать себя так, чтоб стать более творческими?

Да, вполне определенно. Я обнаружил, что есть путь стимулировать обнаружение необычных аналогий, когда работаешь над проблемой. Во-первых, вы заранее должны предполагать, что есть ответ на проблему, которую вы пытаетесь решить. Люди слишком легко сдаются. Вам необходимо быть убежденным, что решение ждет, чтобы его открыли, и вы должны продолжать думать о проблеме продолжительный период времени.

Во-вторых, вы должны позволить своему разуму блуждать свободно. Вы должны дать своему мозгу время и пространство для нахождения решения. Нахождение решения проблемы – это буквально нахождение паттерна в мире, или паттерна, хранящегося в вашем кортексе, который аналогичен проблеме, над которой вы уже работали. Если вы застряли над проблемой, модель «память-предсказание» советует найти другой способ взглянуть нее, чтобы увеличить вероятность увидеть аналогии в прошлом опыте. Если вы будете просто сидеть и пристально искать снова и снова – вы далеко не продвинетесь. Попробуйте взять части вашей проблемы и упорядочить их по-другому – буквально и фигурально. Когда я играю в Scrabble, я постоянно переставляю фишки. Я не надеюсь, что буквы случайно составят новое слово, но я надеюсь, что различные комбинации букв напомнят мне слова или части слов, которые могли бы быть частью решения. Если вы видите, что ваш рисунок не имеет смысла, попробуйте нарисовать его сверху вниз, изменить цвета или изменить перспективу. Например, когда я думал о том, как различные паттерны в V1 могут привести к инвариантному представлению в IT, я застрял. Тогда я перевернул проблему и спросил, как постоянные паттерны в IT могли бы привести к различным предсказаниям в V1. Инвертирование проблемы немедленно помогло, в конце концов приведя меня к моей уверенности, что V1 не должна рассматриваться как единая кортикальная область.

Если вы застряли над решением проблемы – отойдите ненадолго. Займитесь чем-нибудь другим. Затем начните снова, перефразируя проблему заново. Если вы сделаете это достаточное количество раз – рано или поздно что-то изменится. Это может занять дни или недели, но все равно это произойдет. Цель – найти аналогичную ситуацию где-то в прошлом или в настоящем опыте. Для успешного решения вы должны часто думать над задачей, но также заниматься другими делами, так, чтобы кортекс имел возможность найти аналогичные знания.

Вот другой пример того, как переупорядочивание проблемы привело к новому решению. В 1994 мои коллеги и я пытались решить, как ввести текст в наладонный компьютер. Все сосредоточились на программном распознавании рукописного текста. Говорили, «смотрите, вы пишете на клочке бумаги, вы должны суметь сделать это же на экране компьютера». К несчастью, это оказалось очень трудным. Это одна из тех вещей, которые компьютеры пока не могут, хотя для мозга это очень просто. Причина в том, что мозг использует память и текущий контекст для предсказания написанного. Слова и буквы, которые не получается распознать самих по себе, легко распознаются в контексте. Сопоставления паттернов в компьютере недостаточно для решения этой задачи. Я разработал несколько компьютеров, которые использовали распознавание традиционных рукописных символов, но они это делали недостаточно хорошо.

Я несколько лет боролся над тем, как заставить распознающую программу работать лучше, и не двигался с места. Однажды я отошел и решил взглянуть на проблему с другой перспективы. Я поискал аналогичную проблему. Я сказал себе: «Как мы вводим текст в настольный компьютер? Мы печатаем его на клавиатуре. Как мы учимся печатать на клавиатуре? Да, действительно, это не легко. Это было придумано недавно и необходимо долгое время, чтоб научиться. Печатание пальцами на клавиатуре в стиле печатной машинки сложно и не интуитивно, это совершенно не похоже на писание от руки – но миллионы людей научились этому. Почему? Потому что это работает». Мои размышления продолжались по аналогии, «Возможно я могу предложить систему текстового ввода, которая не обязательно будет интуитивной, но люди будут использовать ее, потому что это работает».

Буквально, именно такой процесс я прошел. Я использовал действие печатания на клавиатуре как аналогию к решению того, как ввести текст стилусом на дисплее. Я понял, что люди готовы научиться сложной задаче (печатанию), потому что это надежный и быстрый способ ввести текст в машину. Следовательно, если б мы смогли создать новый метод ввода текста с помощью стилуса, который будет быстрым и надежным, люди использовали бы его, даже если это потребует обучения. Таким образом, я разработал алфавит, который бы надежно транслировал написанное в компьютерный текст, мы назвали его Граффити. В системах распознавания традиционного рукописного текста, когда компьютер неправильно интерпретирует написанное – вы не знаете почему. Но система Граффити всегда производит корректные буквы, несмотря на то, что вы делаете ошибки в написании. Наш мозг не любит непредсказуемость, вот почему люди не любят системы распознавания традиционного рукописного текста.

Множество людей считает Граффити совершенно дурацкой идеей. Я прошел через все то, как по их мнению должен работать компьютер. Заклинанием тех дней было то, что компьютер должен адаптироваться к пользователю, но не наоборот. Но я был уверен, что люди приняли бы новый способ ввода текста по аналогии с клавиатурой. Граффити обернулось великолепным решением и было широко принято. До сих пор я слышу, как некоторые люди заявляют, что компьютеры должны адаптироваться к пользователям. Это не всегда верно. Наш мозг предпочитает системы, которые непротиворечивы и предсказуемы, и мы можем обучиться новым умениям.

## 7.6. Может ли творчество завести меня в тупик? Могу ли я обмануть сам себя?

Неправильные аналогии всегда опасны. История науки кишит примерами красивых аналогий, которые оказывались неверными. Например, известный астроном Кеплер убедил себя в том, что орбиты шести известных тогда планет определяются платоновскими телами. Платоновское тело – это всего лишь трехмерная форма, которая может быть сконструирована полностью из правильных многоугольников. Их существует ровно пять: тетраэдр (четыре равносторонних треугольника), секстаэдр (шесть квадратов, ака куб), октаэдр (восемь равносторонних треугольников), додекаэдр (двенадцать правильных пятиугольников) и икосаэдр (двадцать равносторонних треугольников). Они были открыты древними греками, которые были охвачены идеей отношений математики и космоса.

Как и все ученики школы Ренессанса, Кеплер был глубоко подвержен влиянию греческих идей. Ему казалось, что не может быть простым совпадением то, что есть пять платонических тел и шесть планет. Как он написал в своей книге *The Cosmic Mystery* (1596 год): «Динамический мир представлен телами с плоской поверхностью. Таковых пять: однако, если их рассматривать как границы, они разделяют 6 вещей: отсюда шесть планет, вращающихся вокруг солнца. Это также причина, почему есть только шесть планет». Он видел красивую, но совершенно ложную аналогию.

Кеплер пришел к вычислению орбит планет в терминах вложенных платонических тел, в центре всех которых находится солнце. Он взял сферу, определенную орбитой Меркурия, в качестве основания и описал вокруг нее октаэдр. Вершины октаэдра задали сферу большего диаметра, которая давала орбиту Венеры. Вокруг орбиты Венеры он описал икосаэдр, чьи вершины дали орбиту Земли. Последовательность продолжилась: додекаэдр, описанный вокруг орбиты Земли дал орбиту Марса, тетраэдр вокруг орбиты Марса дал орбиту Юпитера, и куб вокруг орбиты юпитера дал орбиту Сатурна. Это было элегантно и красиво. Имея астрономические данные ограниченной точности тех дней, он смог убедить себя, что эта схема работает. (Годами позже Кеплер понял, что он ошибался, после того, как он использовал высокоточные астрономические данные своего покойного коллеги Тихо Браге, который доказал, что орбиты планет – эллипсы, а не окружности).

Ошибка Кеплера служит предупреждением для ученых и, несомненно, для всех мыслителей. Мозг это орган, который строит модели и делает творческие предсказания, но эти модели и предсказания могут оказаться как обманчивыми, так и верными. Наш мозг всегда смотрит на паттерны и находит аналогии. Если правильные корреляции не могут быть найдены, мозг с легкостью может приять ложную. Псевдонаука, предрассудки, вера и нетерпимость часто коренятся в ложных аналогиях.

## 7.7. Что такое сознание?

Это один из тех вопросов, которые опасаются нейроученые, причем излишне по моему мнению. Некоторые ученые, такие как Кристоф Кох, готовы биться над сутью сознания, но большинство рассматривают этот вопрос на границе философии с псевдонаукой. Я думаю он заслуживает рассмотрения, хотя бы по той причине, что людям это любопытно. Я не могу дать полностью удовлетворяющий ответ, но я думаю, что память и предсказание могут быть обращены к части его. Во-первых, есть привкус головоломки в том, как оно рассматривается в разговорах.

Не так давно я был на научной конференции в прекрасном местечке ЛонгАйлендСаунд. Был ранний вечер, когда большинство из нас приняли по стаканчику вина сидя на пирсе перед водой и болтая перед ужином и вечерней сессией. Через некоторое время разговор коснулся темы сознания. Как я уже сказал, нейроученые обычно не говорят об этом, но мы были в прекрасной обстановке, было выпито некоторое количество вина, и тема разгорелась.

Женщина-ученый из Британии разглагольствовала о ее идеях о сознании, и сказала: «Конечно, мы никогда не поймем сознание». Я не согласился: «Сознание – это не такая большая проблема. Я думаю, сознание это похоже просто наличие кортекса». Наступила тишина, затем быстро последовал аргумент, когда несколько ученых попытались убедить меня в моей очевидной ошибке. «Вы должны признать, что мир кажется таким живым и красивым. Как вы можете отрицать сознание, которое воспринимает мир? Вы должны признать, что вы хотите чего-то особенного». Чтобы доказать свою точку зрения, я сказал: «Я не знаю, о чем вы говорите. Из того факта, как вы говорите о сознании, я делаю вывод, что я отличаюсь от вас. Я не чувствую того, что чувствуете вы, значит я бессознателен. Похоже я зомби». Зомби часто упоминаются, когда философы говоря о сознании. Зомби такие же, как люди, но бессознательные. Они ходячие живые мясные машины, но без царя в голове.

Ученый из Британии посмотрела на меня. «Конечно же у вас есть сознание!».

«Нет, я так не думаю. Вам так может казаться, но я не являюсь сознательным человеком. Не волнуйтесь об этом, мне и так хорошо».

Она сказала: «Значит вы не испытываете каких либо чувств?» и махнула рукой в сторону искрящейся воды, где опускалось солнце и небеса переливались оранжево-розовым оттенком.

«Да, я все это вижу. И что?»

«Тогда как вы объясните свои субъективные ощущения?»

Я ответил, «Да, я знаю, что я здесь. У меня есть воспоминания о вещах, похожих на этот вечер. Но я не чувствую, что должно быть что-то особенное, так что, если вы испытываете что-то особенное, возможно я бессознателен». Я попытался, чтоб она точнее определила свои сверхъестественные и необъяснимые мысли о сознании. Я попытался, чтоб она точно определила сознание.

Мы продолжали этот поток аргументов до тех пор, пока не настало время ужина. Я не думаю, что я изменил чье либо мнение о существовании и значении сознания. Но я попытался, чтоб они поняли, что большинство людей думают о сознании, как о некотором виде волшебного соуса, который добавляется к физическому мозгу. У вас есть мозг, состоящий из нейронов, и они выделяют сознание – в этом суть человека. С этой точки зрения сознания полностью отделено от мозга. Вот почему у зомби есть мозг, но нет сознания. У них есть все механические штуки – нейроны и синапсы, но у них нет волшебного соуса. Они могут все, что может человек. Со стороны вы не можете отличить зомби от человека.

Мысль о том, что сознание это что-то дополнительное к материи произрастает из ранних верований в жизненный поток – некую силу, дающую жизнь всему живому. Люди верят, что им необходима эта жизненная сила, чтобы объяснить разницу между камнями и растениями или металлом и живым человеком. Некоторые люди в это все еще верят. В наше время нам достаточно известно о разнице между неживой и живой материей, чтобы понять, что нет никакого особого соуса. Теперь мы знаем достаточно много о ДНК, о синтезе белков, транскрипции генов и метаболизме. Хотя мы знаем еще не все механизмы живых систем, нам достаточно известно о биологии, чтобы отказаться от волшебства. Аналогично, люди больше не думают, что необходима магия или какие-то духи, чтобы шевелить мускулами. У нас есть длинные белки, которые притягивают молекулы одну к другой. Вы можете обо всем об этом прочитать.

Тем не менее большинство людей продолжают верить, что сознания чем-то отличается и не может быть объяснено сведением к биологическим терминам. С другой стороны я не изучал сознание. Я ее прочел все философские мнения. Но у меня есть некоторые идеи о том, что смущает людей в их рассуждениях. Я верю, что сознание - это похоже просто наличие неокортекса. Но мы можем кое что еще. Мы можем разбить сознание на две категории. Одна подобна самоанализу – обыденное понятие о наличии сознания. Это относительно легкий путь для понимания. Другое это «*qualia»* - идея, что чувства, ассоциированные с ощущениями, не зависят от сенсорной информации. Qualia – это сложнее.

Когда большинство людей произносят слово *сознание*, они обычно ссылаются на первую категорию. «Были ли вы в сознании, когда прошли мимо меня не поздоровавшись?». «Были ли вы в сознании, когда свалились в кровать прошлой ночью?». «Вы без сознания, когда спите». Некоторые люди говорят, что такая форма сознания это то же самое, что и осознание. Второе ближе к истине, но я думаю, что осознание охватывает его не совсем точно. Я полагаю, это значение сознания является синонимом формированию декларативных знаний. Декларативная память – это знания, которые вы можете вспоминать и о которых можете рассказать кому-то другому. Вы можете выразить их вербально. Если вы спрашиваете меня, где я был в прошлые выходные, я могу рассказать вам. Это декларативные знания. Если вы спрашиваете меня, как удержать в равновесии велосипед, я могу сказать вам, что нужно держать руль и жать на педали, но я не могу объяснить точно, как сделать это. Удержание велосипеда в равновесии осуществляется в основном нервной активностью в старом мозге, так что это не декларативные знания.

Проведем небольшой мысленный эксперимент, чтоб показать, каким образом наши повседневные замечания о сознании являются тем же самым, что и формирование декларативных знаний. Вспомните, что считается, что вся память заключается в физических изменениях в синапсах и нейронах, которые они соединяют. Следовательно, если у меня был бы способ обратить эти физические изменения, ваша память могла бы быть стерта. Теперь вообразите, что я могу повернуть переключатель и вернуть ваш мозг точно в такое же состояние, в котором он был в некоторый момент прошлого. Это может быть час назад, двенадцать часов назад, любой промежуток времени. Я просто поворачиваю переключатель в моей обращающей машине, и ваши синапсы и нейроны возвращаются в предыдущее состояние во времени. Делая это, я стираю вашу память о том, что произошло с того момента времени.

Теперь предположим, что вы прожили сегодняшний день и проснулись завтра. Но как только вы проснулись, я повернул переключатель и стер последние двенадцать часов. У вас была бы абсолютно нулевая память о предыдущем дне. С точки зрения вашего мозга, «вчера» никогда не происходило. Я сказал бы вам, что уже среда, а вы протестовали бы, «нет, вторник. Я уверен в этом. Календарь просто поменяли. Но все равно вторник. Зачем вы так шутите надо мной?». Но каждый, кого вы встречали во вторник, сказал бы, что вы были весь день в сознании. Они видели вас, обедали с вами, разговаривали с вами. Разве вы этого не помните? Вы бы сказали нет, этого не было. В конце концов, посмотрев видеозапись вашего обеда, вы постепенно убеждаетесь, что день действительно прошел, даже хотя у вас нет воспоминаний об этом. Как будто вы были целый день зомби, а не в сознании. Однако, вы были в сознании все время. Ваша уверенность в том, что вы были в сознании, исчезла только тогда, когда ваши декларативные знания были стерты.

Этот мысленный эксперимент охватывает эквивалентность между декларативной памятью и нашим повседневным пониманием сознания. Если во время или в конце игры в теннис я спрошу вас, были ли вы в сознании, конечно же, вы скажете да. Если затем я стер бы ваши воспоминания по последних двух часах, вы заявили бы, что были без сознания и не отвечаете за действия, которые происходили в течение игры. В обоих случаях, это была одна и та же игра в теннис. Единственная разница в том, есть ли у вас воспоминания о ней в тот момент, когда я вас спрашиваю. Следовательно, это значение сознания не абсолютно. Оно может изменить после стирания памяти.

Наиболее сложные вопросы о сознании касаются qualia. Qualia часто перефразируется в стиле дзен-буддизма, например «Почему красное – красное, а зеленое – зеленое? Является ли красное для меня таким же, как и для вас? Почему красное эмоционально связывается с определенными чувствами? Это должно быть для меня неотъемлемое качество или чувство. Какое чувство оно вызывает у вас?»

Я обнаружил, что такое описание трудно соотносится с нейробиологией, так что я предпочитаю перефразировать вопрос. Для меня эквивалентным, но все также трудным для объяснения, является вопрос «Почему разные чувства кажутся качественно различными? Почему зрение кажется отличным от слуха, и почему слух кажется отличным от осязания? Если кортекс одинаков везде, если он работает по одним и те м же принципам, если он всего лишь оперирует с паттернами, если нет звука или света в голове, а только паттерны, то почему зрение кажется таким отличающимся от слуха?». Мне трудно описать, как зрение отличается от слуха, но это самоочевидно. Я думаю, для вас также. Аксоны для представления звука и аксоны для представления света во всех практических отношениях идентичны. «Освещенность» и «звучание» не переносятся по аксонам сенсорных нейронов.

У людей с особенностью, называемой синестезией, мозг размывает различия между чувствами – определенные звуки имеют цвет, или определенные текстуры имеют цвет. Это говорит нам, что качественный аспект чувств не является неизменным. Путем некоторых физических модификаций, мозг может придать качественный аспект зрения слуховой информации.

Так каково же объяснение для qualia? Я могу предположить две возможности, ни одну из которых я не нахожу полностью удовлетворительной. Одна заключается в том, что хотя слух, осязание и зрение работают по одним и тем же принципам неокортекса, они по разному обрабатываются под кортексом. Слух полагается на множество специфических для слуха подкорковых структур, которые обрабатывают звуковые паттерны прежде чем они достигнут кортекса. Соматосенсорные паттерны также проходят через множество субкортикальных областей, которые уникальны для чувств осязания. Вероятно qualia, подобно эмоциям, не опосредуется только неокортексом. Если они как-то переплетены с субкортикальными частями мозга, которые имеют уникальную структуру связей, возможно связанную с эмоциональными центрами, это могло бы объяснить, почему мы воспринимаем их по-разному, даже если это не поможет нам объяснить, почему какой-то вид qualia находится на первом месте.

Другая возможность, которую я могу подразумевать, это что структура поступающей информации – отличия в самих паттернах – диктует то, как вы воспринимаете качественные аспекты информации. Природа пространственно-временных паттернов в слуховом нерве отличается от природы пространственно-временных паттернов в оптическом нерве. В оптическом нерве миллионы волокон и он переносит практически пространственную информацию. В слуховом нерве только тридцать тысяч волокон и он переносит больше временнОй информации. Эти отличия могут быть соотнесены с тем, что мы называем qualia.

Мы можем быть уверены, что как бы ни определялось сознание, память и предсказание играют ключевую роль в их создании.

К сознанию относятся и понятия разума и души.

Когда я был ребенком, я не переставал удивляться тому, что было бы, если бы «Я» родился в теле другого ребенка в другой стране, как если бы «Я» каким-то образом не зависело от моего тела. Эти мысли о независимом разуме являются всеобще распространенными и естественным следствием того, как работает неокортекс. Ваш кортекс создает модель мира в своей иерархической памяти. Мысли – это то, что возникает, когда эта модель работает сама по себе; воспоминания ведут к предсказаниям, которая выступает в качестве сенсорной информации, что ведет к новым воспоминаниям и т.д. Наши наиболее абстрактные мысли не приводятся в действие и даже не связаны с реальным миром; они чисто продукт модели. Мы закрываем глаза и тихо размышляем, так что наши мысли не прерываются сенсорной информацией. Конечно, наша модель мира изначально создана изучением реального мира через органы чувств, но когда мы планируем и думаем о мире, мы делаем это в кортикальной модели, а не в реальном мире.

Для кортекса наше тело является всего лишь частью внешнего мира. Вспомните, что мозг находится в тихой и темной коробке. Он знает о мире только через паттерны в сенсорных нервных волокнах. С точки зрения мозга, как устройства для обработки паттернов, он не знает о вашем теле как-нибудь не так, как об остальном мире. Нет какого-либо специального различия между тем, где кончается тело и начинается мир. Но у кортекса нет возможности смоделировать мозг сам по себе, потому что в мозгу нет сенсоров. Таким образом, мы можем видеть, почему наши мысли возникают независимо от нашего тела, почему он склонен к тому, чтоб у нас был независимый разум или душа. Кортекс строит модель вашего тела, но он не может построить модель мозга самого по себе. Ваши мысли, которые располагаются в мозгу, физически отделены от тела и оставшегося мира. Разум независим от тела, но не от мозга.

Мы можем ясно увидеть эти отличия при травме или заболевании. Если кто-то теряет конечность, модель конечности в его мозгу может тем не менее оставаться нетронутой, давая в результате так называемую фантомную конечность, которую он все также может ощущать как часть тела. С другой стороны, если он испытает кортикальную травму, он может потерять модель руки, даже хотя рука сама по себе останется. В этом случае он может испытать то, что известно как синдром чужой конечности, и ощущать дискомфортное, возможно невыносимое чувство, что рука не его собственная и управляется кем-то другим. Некоторые даже настаивают, что конечность должна быть ампутирована! Если наш мозг остается нетронутым, тогда как остальное тело заболевает, у нас ощущение здравого разума заключенного в умирающем теле, хотя то, что мы имеем в действительности – это здоровый мозг, заключенный в умирающем теле. Естественно считать, что наш разум продолжается после смерти тела, но когда мозг умирает, также умирает и разум. Истинность этого становится очевидной, если наш мозг выходит из строя раньше тела. Люди с болезнью Альцгеймера или с серьезным повреждением мозга теряют свой разум даже если из тела остаются здоровыми.

## 7.8. Что такое воображение?

Концептуально, воображение - это очень просто. Паттерны попадают в каждую кортикальную область либо от органов чувств, либо от нижележащих по иерархии областей. Каждая кортикальная область создает предсказания, которые посылаются обратно вниз по иерархии. Чтобы вообразить что-либо, вы только позволяете вашим предсказаниям вернуться назад и стать входной информацией. Ничего не делая физически, вы можете следовать вашим предсказаниям. «Если произойдет это, то должно последовать то, потом третье», и так далее. Мы делаем это, когда готовимся к деловой встрече, играем в шахматы, готовимся к спортивному событию или делаем тысячи других вещей.

В шахматах вы воображаете, как двигаете фигуру на определенную позицию и затем представляете, как будет выглядеть доска после этого хода. С помощью этой мысленной картинки вы предсказываете, что сделает ваш оппонент и как будет выглядеть доска после его хода. Затем вы предсказываете, что сделаете вы, и так далее. Вы проходите по воображаемым ходам и их следствиям. В конце концов на основании воображаемой последовательности событий вы решаете, будет ли начальный ход хорошим или нет. Некоторые спортсмены, такие как горнолыжники, могут улучшить свой результат, если они мысленно повторяют маршрут в своей голове снова и снова. Закрывая глаза и воображая каждый поворот, каждое препятствие и даже нахождение на пьедестале, они увеличивают свои шансы на успех. Воображение – это всего лишь другое название планирования. Именно там наша предсказательная способность окупается сполна. Она позволяет нам узнать, каковы будут следствия наших действий до того, как мы совершим их.

Воображение требует нейронного механизма для превращения предсказания во входную информацию. В главе 6 я предположил, что предсказание возникает в нейронах шестого слоя. Нейроны в этом слое проецируются на нижележащие уровни иерархии, но они также проецируются и вверх, к входным нейронам слоя 4. Таким образом выходная информация области кортекса может стать входной для нее же. Как я уже упоминал ранее, специалист по кортикальному моделированию, Стивен Гроссберг называет эти цепочки для воображения «folded feedback» («закольцованные обратные связи»). Если вы закроете глаза и вообразите гиппопотама, визуальная область кортекса станет активной, точно так же, как будто вы действительно увидели гиппопотама. Вы видите то, что воображаете.

## 7.9. Что есть реальность?

Люди спрашивают с волнением и удивлением, «Вы хотите сказать, что наш мозг создает модель мира? И что эта модель может быть более важной, чем реальность?»

Я говорю «Да, в некоторых случаях это было бы верно».

«Но существует ли вообще мир за пределами моей головы?»

Конечно существует. Люди реальны, деревья реальны, моя кошка реальна, социальные ситуации, в которых вы себя обнаруживаете – реальны. Но ваше понимание мира и ваш отклик на него базируется на предсказаниях, идущих из внутренней модели. В любой момент времени напрямую вы можете ощущать только крошечную часть мира. Эта крошечная часть диктует, какие воспоминания должны быть задействованы, но самой по себе ее не достаточно для построения вашего полного восприятия. Например, прямо сейчас я печатаю в моем кабинете и слышу стук в парадную дверь. Я знаю, что пришла моя мама и я воображаю, как она спускается по лестнице, даже хотя я в действительности не видел и не слышал ее. В моем восприятии не было ничего конкретно связанного с моей мамой. Именно моя модель мира предсказала, что она здесь, по аналогии с предыдущим опытом. Большинство из того, что вы воспринимаете приходит не от органов чувств; оно генерируется вашей внутренней моделью.

Так что вопрос «Что есть реальность?» в основном касается того, как точно наша кортикальная модель отражает истинную природу мира.

Множество аспектов окружающего мира настолько устойчивы, что у рядом живущих людей модели мира похожи. В детстве вы запомнили, что свет, падающий на круглый предмет, производит определенную тень, и что вы можете оценить форму большинства предметов по подсказкам из естественного мира. Вы запомнили, что если вы выроните чашку из рук, гравитация всегда притянет ее на пол. Вы запомнили текстуры, геометрию, цвета и ритмы дня и ночи. Простые физические свойства мира согласованно запоминаются всеми людьми.

Но большая часть модели мира базируется на традициях, культуре и на том, чему научили нас наши родители. Эти части нашей модели менее согласующиеся, и могут радикально отличаться у различных людей. Ребенок, выросший в любящей и заботливой семье, чьи родители отзывались на его или ее эмоциональные нужды, вырастет, вероятно, с ожиданием того, что мир справедливый и любящий. Ребенок, которого обижали один или оба родителя, вероятно будет видеть будущие события опасными и суровыми, и будет считать, что никому нельзя верить – независимо от того, как сложится их дальнейшая жизнь. Большая часть психологии базируется на следствиях раннего жизненного опыта, привязанностях и воспитании, потому что именно тогда в мозгу закладывается фундамент модели мира.

Ваша культура основательно формирует вашу модель мира. Например, изучение показывает, что азиаты и европейцы по разному воспринимают пространство и объекты. Азиаты уделяют больше внимания пространству между объектами, тогда как европейцы в основном уделяют внимание объектам – отличие, которое транслируется в различную эстетику и способы решения проблем. Исследования показали, что некоторые культуры, такие как племена Афганистана и некоторые племена Южной Америки построены на принципах чести, и, как результат, легче воспринимают естественность насилия. Различные религиозные верования, приобретенные в детстве, могут привести к полностью отличающимся моделям морали, отношениям между мужчинами и женщинами, и даже к ценности самой жизни. Несомненно такие отличающиеся модели мира не могут все быть абсолютно корректными, универсальными, даже хотя они могут казаться правильными конкретным людям. Вывод – и хорошее и плохое – это всего лишь то, как вас научили.

Ваша культура (и семейный опыт) обучает вас стереотипам, которые к несчастью, являются неизбежной частью жизни. На протяжении этой книги вы могли бы заменить слово *стереотип* на *инвариантное воспоминание* (или *инвариантное представление*) без существенного изменения смысла. Предсказание по аналогии – это почти тоже самое, что и суждение по стереотипам. Негативные стереотипы имеют ужасные социальные последствия. Если моя теория интеллекта верна, мы не сможем избавить людей от их склонности мыслить стереотипами, потому что стереотипы – это то, как работает кортекс. Использование стереотипов – это неотъемлемое свойство мозга.

Способ избежать вреда, вызванного стереотипами – это учить наших детей узнавать ложные стереотипы, быть сопереживающими и быть скептическими. Нам необходимо продвигать эти критические для мышления вещи в дополнение к внушению известных нам ценностей. Скептицизм, сердце научного метода, единственный известный нам путь отличить факты от фантазии.

\* \* \*

**С**ейчас я надеюсь, что я убедил вас в том, что разум это просто название для того, что делает мозг. Это не отдельная вещь, которая манипулирует или сосуществует с нейронами мозга. Нейроны это просто клетки. Нет мистической силы, которая заставляет конкретные нервные клетки или группы нервных клеток вести себя не так, как они вели бы себя обычно. Поставленные перед фактом, теперь мы можем обратить наше внимание на то, как мы могли бы воплотить способности нервных клеток к запоминанию и предсказанию – наши кортикальные алгоритмы – в кремнии.

# 8. Будущее интеллекта

**Сложно** предсказать, как в отдаленном будущем будут использованы новые технологии. Как мы видели в этой книге, мозг делает предсказания по аналогии с прошлым. Так что у нас природная склонность воображать, что новые технологии будут использованы для того, чтобы делать то же самое, что делали предыдущие технологии. Мы воображаем, используя новый инструмент для того, чтоб делать что-то известное, только быстрее, эффективнее или дешевле.

Примеров предостаточно. Люди называли поезд «железной лошадью», а автомобиль «безлошадной телегой». Десятилетия телефон рассматривался в контексте телеграфа, что-то, что должно быть использовано только для передачи важных новостей или сведений; так продолжалось до 1920 года, пока люди не стали использовать его повсеместно. Фотография изначально использовалась как новая форма портретной живописи. Кино рассматривалось как разновидность сценического искусства, вот почему в кинотеатрах существовал занавес перед экраном большую часть 20-го столетия.

Но окончательное использование новых технологий часто неожиданно и влекут более серьезные последствия, чем сначала может показаться нашему воображению. Телефон был развит в беспроводные сети передачи голоса и данных, позволяя любым двум людям на планете общаться друг с другом голосом, текстом, картинками, не зависимо от того, где они. Транзистор был изобретен в 1947 году в Bell Labs. Людям сразу же стало ясно, что это устройство произведет прорыв, но первым применением было просто улучшение старых приложений: транзисторами заменили вакуумные лампы. Это привело к более надежным и компактным радио и компьютерам, что было важно и замечательно в те дни, но основное отличие был в размерах и надежности машин. Наиболее революционное применение транзистора было открыто позже. Был необходим период постепенных инноваций, прежде чем кто-нибудь смог подумать об интегральных микросхемах, микропроцессоре, процессоре цифрового сигнала или микросхеме памяти. Подобным образом микропроцессор, разработанный в 1970 году, замышлялся для настольных вычислителей. Снова, первым приложением была простая замена существующих технологий. Механические настольные вычислители были заменены на электронные калькуляторы. Микропроцессоры так же рассматривались как замена для устройств на реле, например, как переключатели светофоров. Однако, прошли годы, прежде чем обнаружилась реальная мощь микропроцессоров. Никто в то время не мог предвидеть современный персональный компьютер, сотовый телефон, Интернет, GPS, или любые другие виды современных повседневных информационных технологий.

По тем же приметам, было бы глупо думать, что мы можем предсказать революционные применения мозгоподобной системы памяти. Я ожидаю, что эти интеллектуальные машины полностью улучшат нашу жизнь во всех смыслах. Мы можем быть уверены в этом. Но предсказать будущее технологии более чем на несколько лет невозможно. Чтобы признать это, вам достаточно прочитать некоторые из абсурдных прогнозов, сделанных футуристами. В 1950 году было предсказано, что к 2000 году у каждого из нас в подвале будет атомный реактор и что мы будем проводить отпуск на луне. Но пока мы верим в эти истории, многое может быть выиграно на спекуляциях о том, на что будут похожи интеллектуальные машины. Как минимум, есть некоторые обширные и полезные выводы, которые мы можем сделать относительно будущего.

Вопросы интригующие. Можем ли мы построить интеллектуальные машины и, если можем, на что они будут похожи? Будут ли они похожи на человекоподобных роботов из фантастических произведений, черной или бесцветной коробкой персонального компьютера или чем то еще? Как они будут использоваться? Не будет ли это опасной технологией, которая может нанести нам вред или угрожать нашей личной свободе? Каковы самые очевидные применения интеллектуальных машин, и существует ли способ узнать, каковы будут самые фантастические применения? Каковым будет окончательное воздействие интеллектуальных машин на нашу жизнь?

## 8.1. Сможем ли мы построить интеллектуальные машины?

Да, мы сможем построить интеллектуальные машины, но они могут оказаться не тем, чего мы ожидаем. Хотя это может показаться очевидным, я не верю, что построенные нами интеллектуальные машины будут действовать как люди, или даже взаимодействовать с нами человекоподобным образом.

Одно популярное представление об интеллектуальных машинах пришло к нам из фильмов и книг – это привлекательные, злые или подчас неуклюжие человекоподобные роботы, которые разговаривают с нами о чувствах, мыслях, и играют роль в бесконечных научно-фантастических заговорах. Столетие научной фантастики приучило людей видеть роботов и андроидов как неизбежную и желаемую часть нашего будущего. Поколение, выросшее на образах робота Робби из «*Затерянной планеты»,* R2D2 и C3PO из «*Звездных войн*», и лейтенант-капитан Дэйта из «Стартрека». Даже HAL из фильма «*2001, Космическая Одиссея*», хотя и не обладал телом, был очень человекоподобным, был разработан для людей в их длительных космических буднях как в качестве компаньона, так и в качестве программного автопилота. Роботы для ограниченного применения – наподобие умных автомобилей, автономных минисубмарин для исследования глубин океана, и автоматические пылесосы или газонокосилки – весьма правдоподобны и когда-нибудь появятся. Но андроиды и роботы вроде Капитана Дейта и C3PO скорее всего останутся фантастикой на долгие годы. На это есть несколько причин.

Первая – человеческий разум создается не только неокортексом, но также эмоциональными системами старого мозга и особенностями человеческого тела. Для того, чтоб быть человеком, вам нужны все биологические механизмы, не только кортекс. Чтобы общаться как человек во всех смыслах (чтобы пройти тест Тьюринга), интеллектуальная машина должна была бы иметь жизненный опыт и эмоции реального человека, и жить человекоподобной жизнью. У интеллектуальных машин будет аналог кортекса и набор органов чувств, но остальное – опционально. Было бы увлекательно увидеть интеллектуальную машину, щеголяющую в человекоподобном теле, но оно не обязательно должно было бы иметь разум, отдаленно похожий на человеческий, за исключением того, что мы наделили бы его человекоподобными эмоциональными системами и человекоподобным опытом. Это было бы крайне сложно, и это кажется мне неэффективным.

Вторая – даже при наличии средств и сил, необходимых для постройки и обслуживания гуманоидных роботов, сложно увидеть их практическое применение. Робот-лакей был бы более дорогим и менее полезным, чем человек-ассистент. Хотя робот мог бы быть «интеллектуальным», у него не было бы такого согласия и взаимопонимания, как у человека-ассистента.

И паровые механизмы, и цифровые компьютеры вы качестве зрения роботов никогда не стали бы плодотворными. Аналогично, когда мы размышляем о построении интеллектуальных машин, для большинства людей является естественным снова и снова воображать человекоподобных роботов, но такое вряд ли будет. Роботы – это концепция, порожденная индустриальной революцией и подправленная фантастикой. Мы не должны принимать их в расчет в качестве вдохновения при разработке действительно интеллектуальных машин.

Так на что же будут похожи интеллектуальные машины, если не на ходящих говорящих роботов? Эволюция подсказывает, что если к органам чувств приделать систему памяти, то память стала бы моделью мира и начала бы предсказывать будущее. Заимствуя идею у природы, мы должны строить интеллектуальные машины аналогичным путем. В этом рецепт для построения интеллектуальных машин. Начнем с набора сенсоров для извлечения паттернов из реального мира. Наши интеллектуальные машины могли бы иметь набор органов чувств, отличающийся от человеческого, и возможно даже «существовали» бы в мире, отличном от нашего (подробнее чуть позже). Так что не предполагайте, что у них должна быть пара глазных яблок и пара ушей. Затем, приделаем к этим органам чувств иерархическую системы памяти, которая работает на тех же самых принципах, что и кортекс. Затем мы должны научить систему памяти, так же, как мы обучаем детей. В процессе обучения, интеллектуальные машины построят модель *их* мира, как они видят его через *их* органы чувств. Не будет необходимости или возможности для кого либо запрограммировать правила мира, базы данных, факты или другие высокоуровневые концепции, которые были бы проклятьем искусственного интеллекта. Интеллектуальные машины должны обучаться через наблюдение мира, включая при необходимости информацию от инструктора. Как только интеллектуальная машина создаст модель ее мира, она сможет увидеть аналогии с прошлым опытом, делать предсказания будущих событий, предлагать решения новых проблем и делать эти знания доступными для нас.

Физически, интеллектуальная машина могла бы быть в строенной в самолеты или автомобили, или сидеть неподвижно на полке в компьютерном зале. В отличие от людей, чей мозг должен жить вместе с телом, система памяти интеллектуальной машины могла бы быть расположена в отдалении от ее сенсоров (и от «тела», если б у нее таковое было). Например, у интеллектуальной системы безопасности сенсоры могли бы быть расположены по всему производственному помещению или территории, но иерархическая система памяти, присоединенная к этим сенсорам, могла бы быть заперта в фундаменте здания. Следовательно, физическое воплощение интеллектуальной машины могло бы иметь множество форм.

Нет причин, по которым интеллектуальная машина должна была бы выглядеть, действовать или чувствовать подобно человеку. Что делает ее интеллектуальной – так это то, что она может понимать и взаимодействовать с ее миром посредством иерархической модели в памяти и может думать о мире таким же способом, как думаете вы и я. Как мы увидим, ее мысли и действия могли бы быть полностью отличающимися от человеческих, но тем не менее она оставалась бы интеллектуальной. Интеллектуальность измеряется предсказательной способностью иерархической памяти, а не человекоподобным поведением.

\* \* \*

**Д**авайте обратим наше внимание на наиболее сложную проблему, которая встанет перед нами при построении интеллектуальных машин и создании памяти. Чтобы построить интеллектуальные машины, нам необходимо построить большую систему памяти, которая была бы иерархически организована и работала бы аналогично кортексу. Мы встали бы перед проблемой емкости и большого числа соединений.

Емкость – это первая проблема. Пусть, скажем, в кортексе 32 триллиона синапсов. Если мы представим каждый синапс всего лишь двумя битами (что дает нам 4 возможных значения для синапса) и каждый байт – восемью битами (так что каждый байт представлял бы 4 синапса), то нам было бы необходимо приблизительно 8 триллионов байтов памяти. Жесткий диск современного персонального компьютера содержит примерно 100 гигабайт, так что нам было бы необходимо около 80 современных жестких дисков, чтобы получить то же самое количество памяти, что и в человеческом мозге. (Не беспокойтесь о точных цифрах, потому что это всего лишь предположения). Суть в том, что такое количество памяти определенно можно построить в лаборатории. Нас уже не останавливают тысячные множители, но это все еще не та машина, которую вы могли бы положить в карман или встроить в тостер. Важно то, что количество требуемой памяти не закрывает вопрос, тогда как всего лишь десятилетие назад закрывало бы. Спасает нас тот факт, что нет необходимости воссоздавать человеческий кортекс целиком. Для многих применений достаточно гораздо меньше.

Интеллектуальным машинам будет нужно много памяти. Вероятно, мы начнем их строить, используя жесткие диски или оптические диски, но в итоге мы хотели бы построить их также и в кремнии. Кремниевые чипы маленькие, потребляют мало энергии и прочные. Это всего лишь вопрос времени, когда кремниевые чипы обретут емкость, достаточную для построения интеллектуальных машин. Фактически, у интеллектуальной памяти есть преимущество над обычной компьютерной памятью. Экономика полупроводниковой индустрии базируется на проценте чипов с ошибками. Для большинства чипов даже единичная ошибка делает чип бесполезным. Процент хороших чипов называется выходом. Он определяет, может ли конкретная разработка чипа быть произведена и продана с пользой. Поскольку шансы на ошибку возрастают с увеличением размеров чипа, большинство современных чипов не больше почтовой марки. Индустрия увеличивает количество памяти на единичном чипе не увеличением размеров чипа, а, в основном, делая элементы меньше размером.

Но чипы интеллектуальной памяти будут изначально толерантны к дефектам. Вспомните, что ни один единичный компонент вашего мозга не содержит каких либо важных элементов данных. Ваш мозг теряет тысячи нейронов каждый день, но ваша ментальная емкость спадает на ничтожную величину в течение вашей жизни. Чипы интеллектуальной памяти будут работать на тех же самых принципах, что и кортекс, так что даже если процент этих чипов будет с дефектами, чипы все равно будут полезны и коммерчески жизнеспособны. Наиболее вероятно, изначальная толерантность мозгоподобной памяти к ошибкам позволит дизайнером разрабатывать чипы значительно большие и более плотные, чем современные чипы компьютерной памяти. Результат таков, что мы сможем воплотить мозг в кремнии гораздо раньше, чем показывают текущие тенденции.

Вторая проблема, которую нам необходимо преодолеть – это количество соединений. В реальном мозге очень большое количество субкортикального белого вещества. Как мы упоминали ранее, белое вещество состоит из миллионов аксонов, идущих чуть ниже кортикального слоя, соединяя различные области кортикальной иерархии. Отдельный нейрон в кортексе может соединяться с пятью-десятью тысячами других нейронов. Такой вид сильно параллельного соединения сложно или невозможно реализовать, используя традиционные технологии кремниевого производства. Кремниевые чипы делаются путем наложения нескольких слоев металла, разделенных слоями изоляции. (Этот процесс не имеет ничего общего со слоями кортекса). Слои металла содержат «провода» чипов, и поскольку провода не могут пересекаться внутри одного слоя, общее число проводов ограничено. Это не годится для мозгоподобной системы памяти, где необходимы миллионы соединений. Силиконовые чипы и белое вещество не совсем совместимы.

Много инженерной и экспериментаторской работы необходимо для решения этой проблемы, но нам известны базовые принципы того, как она будет решена. Электрические провода посылают сигналы гораздо быстрее, чем аксоны нейронов. Единичный провод на чипе может быть использован для множества различных соединений, тогда как в мозгу каждый аксон принадлежит только одному нейрону.

Пример из реального мира – система телефонии. Если б мы проложили линии от каждого телефона к каждому другому телефону, поверхность планеты была бы зарыта под джунглями медных проводов. Вместо этого все телефоны используют относительно небольшое количество высокоскоростных каналов. Этот метод работает, пока емкость каждого канала гораздо больше емкости, требуемой для передачи единственного разговора. Телефонная система удовлетворяет этому требованию: единственное оптоволокно может передавать миллионы разговоров одновременно.

В реальном мозге для общения любых двух нейронов существует отдельный аксон, но мы можем построить интеллектуальные машины по принципу телефонной системы, отдавая соединения в совместное использование. Верите или нет, некоторые ученые многие годы думают над тем, как решить проблему мозгового чипа. Даже хотя операции в кортексе оставались загадкой, исследователи знали, что когда-нибудь мы решим эту головоломку, и затем мы должны предстать перед вопросом количества соединений. Нет необходимости рассматривать здесь различные подходы. Достаточно сказать, что количество соединений может стать огромнейшим техническим препятствием, с которым мы столкнемся при построении интеллектуальных машин, но мы должны суметь преодолеть его.

Как только технологические вопросы будут преодолены, не останется фундаментальных проблем, препятствующих построению действительно интеллектуальных систем. Да, есть множество вопросов, которые должны быть решены, чтоб сделать эти системы маленькими, недорогими и с низким энергопотреблением, но больше ничего не будет стоять на нашем пути. Ушло пятьдесят лет на то, чтоб от компьютеров размером с комнату прийти к карманным компьютерам. Но поскольку мы начинаем с продвинутой технологической позиции, тот же самый путь для интеллектуальных машин должен быть пройден гораздо быстрее.

## 8.2. Должны ли мы строить интеллектуальные машины?

В течение двадцать первого столетия интеллектуальные машины должны выйти из области научной фантастики и стать фактом. Прежде чем мы приступим к этому, мы должны подумать об этических вопросах, не превзойдет ли опасность интеллектуальных машин их возможные преимущества?

Перспектива машин, которые могут думать и действовать сами по себе, волновала людей многие годы. Это понятно. Новые области знания и новые технологии всегда пугали людей, когда только-только появлялись. Человеческое творчество позволяет нам вообразить ужасные пути, на которых новая технология может унаследовать наши тела, переоценить наше предназначение, или перечеркнуть саму ценность человеческой жизни. Но история показывает, что такие мрачные картины почти никогда не идут по тому пути, как мы ожидаем. Когда начиналась индустриальная революция, мы боялись электричества (помните Франкенштейна?) и паровых машин. Машины, которые обладают своей собственной энергией, которые могут вести себя сложным образом, кажутся сверхъестественными и в то же время потенциально зловещими. Но электричество и энергия внутреннего сжигания больше не кажутся странными и зловещими. Они такая же часть нашего окружения, как воздух и вода.

Когда началась информационная революция, вскоре возник страх перед компьютерами. Было бесчисленное множество научно-фантастических произведений о мощных компьютерах или компьютерных сетях, которые внезапно сознавали себя и затем оборачивались против своих органических хозяев. Но сейчас, когда компьютеры стали интегрированными в повседневную жизнь, этот страх кажется абсурдным. У Компьютер у вас дома, или Интернет имеют такие же шансы спонтанно обернуться сознательными, как и кассовый аппарат.

Конечно, любая технология может быть применена с добрым или злым умыслом, но некоторые более предрасположены к злоупотреблению или трагическому исходу, чем другие. Атомная энергия опасна, воплощена ли она в ядерной боеголовке или на электростанции, потому что единственная катастрофа или единственное злоупотребление может искалечить или убить миллионы людей. И хотя ядерная энергия очень полезна, ей имеются альтернативы. Транспортные технологии могут принять форму танков и истребителей, или может принять форму автомобилей и пассажирских самолетов, и несчастный случай или злоупотребление могут навредить множеству людей. Но транспортные средства возможно и более присущи современной жизни и менее опасны, чем ядерная энергия. Разрушения, вызванные одиночным злоупотреблением самолета гораздо меньше, чем ядерной бомбы. Есть множество технологий, которые в основном полностью благотворны. Например, телефон. Его тенденция соединять людей превышает любые негативные эффекты. То же самое для электричества и здравоохранения. По моему мнению, интеллектуальные машины будут одной из наименее опасных из технологий, наиболее благотворной технологией, которую мы когда либо разрабатывали.

Тем не менее, некоторые мыслители, наподобие Билла Джоя, одного из соучредителей Sun Microsystems, опасаются, что мы можем разработать интеллектуальных роботов, которые могли бы избежать нашего контроля, кишели бы по Земле, и переделали бы ее в соответствии с их собственными намерениями. У меня в воображении возникает картинка тех живых метелок из *Sorcerer's Apprentice*, регенерирующих себя из своих щепок и неустанно делающих все, чтоб привести к катастрофе. Аналогичным образом, некоторые оптимисты ИИ предлагают тревожащие пророчества. К примеру, Рей Курцвейл говорит о дне, когда нанороботы проберутся в ваш мозг, запишут каждый синапс и каждое соединение, и затем передадут всю информацию суперкомпьютеру, который реконфигурирует себя в вас! Вы станете «программной» версией самого себя, которая станет практически бессмертна. Эти два предсказания о машинном интеллекте, сценарий обезумевающих интеллектуальных машин и сценарий загрузки мозга в компьютер, кажутся на первый взгляд очень жуткими.

Построение интеллектуальных машин – это не то же самое, что построение само копирующихся машин. Между ними нет какой бы то ни было логической связи. Ни мозг ни компьютер не могут напрямую само копироваться, и мозгоподобная система памяти не отличается от них. Пока одним из достоинств интеллектуальных машин будет наша способность их массового производства, мир избежит из самокопирования на манер бактерий или вирусов. Самокопирование не ведет к интеллекту, и интеллект не ведет к самокопированию.

Более того, я серьезно сомневаюсь в том, что мы когда либо сможем скопировать наш разум в машину. Сейчас, насколько мне известно, нет действительных или воображаемых методов, способных записать триллионы деталей, составляющих ваше «Я». Нам потребуется записать и воссоздать все части нервной системы и тела, не только неокортекс. И нам необходимо будет понять, как все это работает. Однажды, мы определенно сможем сделать это, но проблемы уходят далеко за понимание того, как работает кортекс. Понимание неокортикального алгоритма и встраивание его в машины с нуля - это одно, но сканирование в несметном количестве деталей живого мозга и воспроизведение этого в машине – это совсем другое.

\* \* \*

**П**осле упоминания о саморепликации и копировании разума, у людей возникает другое отношение к интеллектуальным машинам. Могут ли интеллектуальные машины как то угрожать большой популяции людей, как ядерная бомба? Может ли их присутствие привести к захвату полномочий небольшой группой злых людей? Или могут ли машины стать злыми и обернуться против нас, подобно безжалостным негодяям из фильмов *Терминатор* или *Матрица*?

Ответ на эти вопросы отрицательный. Как информационные устройства, мозгоподобные системы памяти должны стать в число наиболее полезных технологий, которые мы когда либо разрабатывали. Но подобно автомобилям и компьютерам, они всего лишь инструмент. Всего лишь обретение интеллектуальности совершенно не обозначает обретение способностей к разрушению имущества или манипулированию людьми. И точно также, как мы не доверили бы контроль над ядерным арсеналом одному человеку или одному компьютеру, мы должны будем проявить осторожность и не полагаться слишком на интеллектуальные машины.

Это приводит нас к вопросу недружелюбия. Некоторые люди полагают, что быть интеллектуальным – это в основном то же самое, что иметь человеческий разум. Они боятся, что интеллектуальные машины обидятся на «порабощение», поскольку люди ненавидят быть рабами. Они боятся, что интеллектуальные машины попытаются захватить мир, потому что интеллектуальные люди в течение всей истории пытались захватить мир. Но эти опасения основаны на ложной аналогии. Они основаны на объединении интеллекта – неокортикального алгоритма – с эмоциональными мотивациями старого мозга – страх, паранойя и страсть. Но у интеллектуальных машин не будет таких способностей. У них не будет личных амбиций. Они не будут желать богатства, общественного признания, или чувственного удовлетворения. У них не будет потребностей, склонностей или плохого настроения. У интеллектуальных машин не будет ничего похожего на человеческие эмоции, если только мы специально не заложим этого в них. Наилучшее применение интеллектуальных машин будет там, где человеческий интеллект испытывает трудности, в тех областях, в которых наши чувства неадекватны, или в деятельности, которая скучна для нас. У этих видов деятельности минимальное эмоциональное содержание.

Интеллектуальные машины будут распространены от простых систем узкого применения до очень мощных сверхчеловеческих интеллектуальных систем, но пока мы не захотим сделать их человекоподобными – они не станут таковыми. Возможно когда-нибудь нам придется установить ограничения на то, что люди могут делать с интеллектуальными машинами, но этот день очень далек, и когда он настанет, этические вопросы скорее всего будут относительно легче по сравнению с моральными вопросами настоящего времени относительно генетики и ядерной технологии.

## 8.3. Зачем строить интеллектуальные машины?

Теперь к вопросу Что будут делать интеллектуальные машины?

Меня часто просят рассказать о будущем мобильных компьютеров. Организаторы конференций нередко просят меня описать, на что будут похожи наладонники или сотовые телефоны через пять или двадцать лет. Они хотят услышать мое видение будущего. Я не могу этого сделать. Я пытаюсь избежать того, что все вместе я называю одним словом «V». Чтобы убедить в этом, я однажды вышел на сцену в шляпе волшебника и с хрустальным шаром. Я объяснил, что никто не может видеть будущее детально. Хрустальный шар – это фикция, и любой, кто претендует на точное знание того, что произойдет в грядущие годы, обречен на провал. Самое лучшее, что мы можем сделать вместо этого – это понять общую тенденцию. Если вы понимаете общую идею, вы вполне сможете отследить ее, как бы она ни развивалась по мере выявления деталей.

Наиболее известный пример технологической тенденции это закон Мура. Гордон Мур верно предсказал, что количество элементов, которые могут быть размещены на кремниевом чипе, должно удваиваться каждые полтора года. Мур не сказал, относится ли это к чипам памяти, процессорам или чему-то еще. Он не сказал, в каком виде продукции эти чипы будут использованы. Он не предсказывал, будут ли чипы в пластиковых или керамических корпусах или наклеенные прямо на плату. Он не сказал ничего о различных процессах, используемых при производстве чипов. Он привязался к наиболее общей тенденции, какой мог, и оказался прав.

В настоящее время мы не можем предсказать окончательное использование интеллектуальных машин. Просто нет способа узнать конкретные детали. Если я, или кто-то еще, предсказывает в деталях, что будут делать эти машины, мы неизбежно будем обмануты. Однако, мы можем сделать нечто большее, чем просто пожимать плечами. Есть две линии идей, которые могут быть полезны. Одна – вообразить наиболее краткосрочное использование мозгоподобной системы памяти – очевидное, но менее интересная вещь, которую можно попробовать для начала. Второй подход – подумать о долгосрочных тенденциях, подобных закону Мура, которые могут помочь нам вообразить применения таких машин, которые, возможно, могли бы стать частью нашего будущего.

Давайте начнем с некоторых краткосрочных применений. Это вещи, кажущиеся очевидными, подобно замене радиоламп на транзисторы или построению калькуляторов на микропроцессорах. И мы сможем начать с обозрения некоторых областей, в которых пытались применить ИИ, но не смогли – распознавание речи, техническое зрение и умные автомобили.

\* \* \*

**Е**сли вы когда-нибудь пытались использовать программы для распознавания речи для ввода текста в персональный компьютер, вы знаете, насколько они могут быть неуклюжими. Подобно Китайской Комнате Серла, у компьютера нет понимания того, что ему сказали. За время использования таких продуктов я был разочарован. Если в комнате есть хоть какой-нибудь шум, от упавшего карандаша до чьего-нибудь разговора, на экране появляются лишние слова. Процент ошибок распознавания очень высок. Часто слова, которые послышались программе, не имели смысла. «Remember to fell Mary that the bog is ready to be piqued up». Ребенок знает, что это неправильно, но не компьютер. Аналогично, так называемые естественно-языковые интерфейсы были целью для компьютерщиков многие годы. Идея в том, чтоб вы могли сказать компьютеру или другим устройствам сказать, чего вы хотите – и пусть машина делает свою работу. Персональному цифровому помощнику, или PDA, вы могли бы сказать «Перенести занятия баскетболом для моей дочери на десять часов утра в воскресенье». Такое невозможно сделать на приемлемом уровне с помощью традиционного ИИ. Даже если компьютер смог бы распознать каждое слово, для того, чтоб выполнить задачу ему потребовалось бы знать, в какую школу ходит ваша дочь, что, возможно, вы имели в виду ближайшее воскресенье, и, может быть, что такое «занятия баскетболом» потому что назначение могло бы только говорить «Menlo vs. St. Joe». Или, вероятно, вы хотели бы, чтоб компьютер слушал радиопередачу на предмет упоминания определенного продукта, даже если диктор описывает этот продукт не используя его название. Вы и я знали бы, о чем он говорит, но компьютер – нет.

Эти и многие другие приложения требуют, чтоб машина могла слышать и понимать устный язык. Но компьютеры не могут выполнять такую задачу, потому что они не понимают, что было сказано. Они сопоставляют звуковые паттерны со словарными шаблонами механически, не зная, что обозначают слова. Представьте себе, если бы вы научились звучание отдельных слов иностранного языка, но не значение слов, и я попросил бы вас перевести разговор с другого языка. В процессе перевода вы бы не понимали, о чем речь, но пытались бы ухватить отдельные слова. Однако, слова перекрываются и взаимодействуют, и частично звуки подавляются шумами. Вам бы показалось черезвычайно сложно отделить слова и узнать их. Именно с этим препятствием борются современные распознающие программы. Инженеры обнаружили, что используя вероятности перехода слов они могут улучшить качество программ в чем-то. Например, они используют правила грамматики для того, чтоб выбрать один из омонимов. Это очень простая форма предсказания, но системы все еще остаются слабо интеллектуальными. Современные программы распознавания речи успешно работают только в очень ограниченных ситуациях, в которых количество слов, которые вы можете сказать в любой момент, ограничено. Но люди очень легко выполняют множество задач, связанных с языком, потому что наш кортекс понимает не только слова, но предложения и контекста, в котором они произнесены. Мы предчувствуем мысли, фразы и отдельные слова. Наша кортикальная модель мира делает это автоматически.

Так что мы можем ожидать, что системы памяти, подобные кортексу, трансформируют ошибочное распознавание речи в надежное понимание речи. Вместо программирования вероятностей переходов отдельных слов, иерархическая память будет отслеживать ударения, слова, фразы, смысл и использовать их для интерпретации того, что было сказано. Подобно человеку, такая интеллектуальная машина могла бы отличать различные случаи применения речи – например, разговор между вами и вашим другом в комнате, телефонный разговор, команды редактирования для текстового редактора. Такие машина построить нелегко. Для полного понимания человеческого языка, машина должна иметь опыт и изучать, что делает человек. Так что даже если пройдет много лет, прежде чем мы сможем построить интеллектуальные машины, мы будем способны в ближайшем будущем улучшить производительность существующих систем распознавания речи, используя память, подобную кортексу.

Зрение предлагает другой набор приложений, которые ИИ не может решить, но которые интеллектуальные системы должны уметь решать. Сегодня нет машин, которые могут смотреть на естественное изображение – мир перед вашими глазами, или картинка с камеры – и описывать, что она видит. Есть несколько успешных применений машинного зрения, которое работает в очень ограниченных областях, такие как визуальное выравнивание чипов на печатной плате или сопоставление лиц по базе данных, но в настоящее время для компьютера невозможно идентифицировать различные объекты или анализировать изображение более общо. Для вас не составляет труда осмотреть комнату и найти место, где можно присесть, но не просите компьютер сделать это. Представьте, что вы смотрите на экран с изображением с камеры безопасности. Можете вы отличить человека, стучащего в дверь с подарком в руке от человека с монтировкой? Конечно можете, но такое различение за пределами способностей современных программ. Следовательно мы нанимаем человека, который следит за экраном камеры безопасности все время в поисках подозрительного. Для такого человека сложно все время оставаться начеку, тогда как интеллектуальная машина могла бы выполнять такую задачу неустанно.

В конце концов, давайте взглянем на транспортировку. Машины становятся все более сложными. У них есть GPS для отображения маршрута от A к B, сенсоры для включения фар, когда становится темно, акселерометры для срабатывания подушки безопасности, сенсоры предупреждающие о том, что вы можете наехать задним ходом на какой то предмет. Есть даже автомобили, которые ездят автономно по специальным путям или в идеальных условиях, хотя они коммерчески недоступны. Но чтобы эффективно и безопасно вести машину по любым типам дорог или в любых дорожных условиях, требуется большее, чем набор сенсоров и обратные связи контуров управления. Чтоб быть хорошим водителем, вы должны понимать дорожную обстановку, других водителей, принцип работы автомобиля, сигнальные огни и массу прочих вещей. Вы должны быть способны понимать предупреждающие знаки или отмечать, когда другая машина движется слишком опасно. Вы должны видеть поворотники другой машины и предвидеть, что она вероятно сменит полосу, или, если сигнал включен уже несколько минут, предположить, что водитель вероятно забыл, что сигнал включен и не будет менять полосу. Вы должны понять, что клубы дыма далеко впереди могут обозначать, что произошла авария и следовательно вы должны замедлить движение. Водитель видит мячик, катящийся по дороге и автоматически понимает, что за ним может выбежать ребенок, чтоб догнать его, и интуитивно жмет на тормоз.

Скажем, мы хотим построить действительно умную машину. Первое, что мы должны сделать – это выбрать набор сенсоров, которые позволили бы машине ощущать мир. Мы могли бы начать с камеры для зрения, возможно несколько камер, спереди и сзади, и микрофоны для слуха, но мы могли бы также обеспечить ее радаром или ультразвуковыми сенсорами, которые могли бы точно определять расстояние до других объектов и скорость и в светлых и в темных условиях. Суть в том, что мы не должны полагаться или ограничивать себя только теми чувствами, которые использует человек. Кортикальный алгоритм гибок, и, пока мы корректно применяем систему иерархической памяти, для нее не должно иметь значения, какие типы сенсоров мы инсталлируем. Наша машина могла бы, теоретически, быть лучше в ощущении мира дорожного движения, чем мы, потому что ее набор сенсоров может быть выбран в соответствии с задачей. Затем сенсоры были бы приделаны к достаточно большой системе иерархической памяти. Разработчики машины натренировали бы память умной машины путем демонстрации ей условий из реального мира, так чтоб она обучалась строить модель ее мира точно также, как делают люди – только более ограниченной области. (Например, машине нужно знать о дорогах, но не о лифтах или самолетах). Память машины изучила бы иерархическую структуру дорожного движения и дорог, так что она смогла бы понимать и предвидеть, что скорее всего должно произойти в ее мире движущихся автомобилей, дорожных знаков, препятствий или пересечений. Разработчики машины могли бы разработать систему памяти такую, чтоб она действительно вела машину или просто наблюдала за тем, что происходит, когда вы ведете. Он мог бы давать советы или наоборот принимать советы в сложных ситуациях, подобно советчику, который бы вас не раздражал. Как только память один раз полностью натренирована и машина может понимать и разбираться во всем, что может произойти, инженеры имели бы возможность постоянной настройки памяти, так что все машины, сходящие со сборочного конвейера вели бы себя одинаково, или они могли бы разработать память для продолжения обучения после того, как машина продана. И с помощью компьютера, но не человеком, память могла бы быть перепрограммирована обновленной версией, которая работала бы и в каких то новых ситуациях.

Я не говорю о том, что мы определенно построим умные автомобили или машины, которые будут понимать язык и обладать зрением. Но это хорошие примеры устройств, которые мы могли бы исследовать и разрабатывать, и кажется возможным построить их.

\* \* \*

**Л**ично меня меньше всего интересуют очевидные применения интеллектуальных машин. Для меня истинное преимущество и восхищение новыми технологиями – это найти для них применения, который были ранее невообразимыми. В каких случаях интеллектуальные машины могли бы удивить нас, и какие фантастические возможности возникли бы через некоторое время? Я уверен, что иерархическая память, подобно транзистору и микропроцессору, преобразует нашу жизнь к лучшему невероятным способом, но каким? Один способ, как мы могли бы мельком увидеть будущее интеллектуальных машин, заключается в том, чтоб подумать над аспектами хорошо масштабируемых технологий. ТО есть, какие атрибуты интеллектуальных машин будут становиться дешевле и дешевле, быстрее и быстрее, или меньше и меньше. Вещи, которые растут экспоненциально, моментально уходят за пределы нашего воображения и наиболее вероятно играют ключевую роль в наиболее радикальной эволюции технологии будущего.

Примеры технологий, которые росли экспоненциально многие годы, включают кремниевые чипы памяти, жесткие диски, технологии расшифровки ДНК, и оптоволоконные технологии. Эти моментально масштабирующиеся технологии были основой для многих новых продуктов и видов деятельности. В другой форме также хорошо масштабировалось программное обеспечение. Нужные программы, однажды написанные, могли быть скопированы множество раз и виртуально не имели стоимости.

В отличие от этого, некоторые технологии, такие как батарейки, моторы, и традиционная робототехника, плохо масштабируются. Несмотря на многочисленные усилия и неизменные улучшения, манипуляторы, построенные сегодня, не намного лучше чем построенные много лет назад. Развитие робототехники идет постепенно и скромно, ничего подобного экспоненциальным кривым роста разработки чипов или распространения программного обеспечения. Манипулятор, построенный в 1985 году за миллион долларов, не может быть построен сегодня за десять долларов. Подобным образом, сегодня батарейки не намного лучше чем они были десять лет назад. Вы могли бы сказать, что они в два или три раза лучше, но не в тысячи раз и прогресс идет очень медленно. Если бы емкость батареек возрастала с такой же скоростью, как емкость жестких дисков, то сотовые телефоны и другая электроника вообще не требовали бы зарядки, а легковесные электромобили, пробегающие тысячи миль за одну зарядку, были бы обыденностью.

Так что нам следует думать о том, какие аспекты мозгоподобной системы памяти будут масштабироваться значительно за пределы наших биологических мозгов. Эти атрибуты могут подсказать, куда в конечном счете придут технологии. Я вижу четыре атрибута, которые должны превысить наши собственные: скорость, емкость, возможность копирования и сенсорные системы.

***Скорость***

Тогда как нейроны работают с временами порядка миллисекунд, кремниевые чипы оперируют с временами порядка наносекунд (и также продолжают становиться быстрее). Эта разница в миллионы раз, на шесть порядков. В следствие этого между органическим и кремниевым мозгом должно быть отличие в скорости. Интеллектуальные машины будут способны думать в миллионы раз быстрее, чем человеческий мозг. Такой разум мог бы читать библиотеки целиком или изучать огромные, сложные массивы данных – задачи, которые могут занять у меня или у вас целые годы – за несколько минут, получая тот же самый результат. В этом нет ничего магического. Биологический мозг эволюционировал с двумя ограничениями по времени. Одно – это скорость, с которой нейроны могу работать, а другое – это скорость, с которой изменяется мир. Для биологического мозга было бы бесполезно думать в миллион раз быстрее, если окружающий мир изменяется медленно. Но в кортикальном алгоритме нет ничего, что заставляло бы его работать медленно. Если бы интеллектуальная машина общалась и взаимодействовала с человеком, она должна была бы замедлить свою работу до человеческой скорости. Если бы она читала книгу, переворачивая страницы, было бы ограничение на то, как быстро она могла бы читать. Но когда она взаимодействовала бы с электронным миром, она могла бы функционировать гораздо быстрее. Две интеллектуальные машины могли бы поддерживать общение в миллион раз быстрее, чем два человека. Вообразите прогресс интеллектуальных машин, которые решают математические или научные проблемы в миллион раз быстрее, чем человек. За десять секунд она могла бы выдать столько же идей по проблеме, сколько вы могли бы за месяц. Никогда не устающая, не поддающийся скуке разум на такой скорости однозначно был бы полезен в таких применениях, которые мы еще не можем вообразить.

***Емкость***

Несмотря на впечатляющую емкость памяти человеческого кортекса, интеллектуальные машины могу быть построены так, что превзойдут ее. Размер нашего мозга ограничен несколькими биологическими факторами, включая отношение черепа ребенка к диаметру таза матери, высокая метаболическая стоимость работающего мозга (ваш мозг всего 2 процента от вашего тела по весу, но использует около 20 процентов вдыхаемого кислорода), и низкая скорость нейронов. Но мы можем построить интеллектуальные системы памяти любого размера, и, в отличие от слепого, извилистого процесса эволюции, мы можем проявить предусмотрительность и специфические намерения к деталям дизайна. Может оказаться, что емкость человеческого неокортекса станет относительно скромной за десятилетие с текущего момента.

При построении интеллектуальных машин мы могли бы увеличить емкость их памяти несколькими способами. Увеличение глубины иерархии приведет к более глубокому пониманию – способности видеть высокоуровневые паттерны. Увеличение емкости областей позволит машине помнить больше деталей, или воспринимать с более высокой точностью, подобно тому, как у слепого человека гораздо лучше чувство осязания и слуха. Добавление новых чувств к сенсорной иерархии позволит устройству создавать лучшую модель мира, как я укажу далее.

Было бы интересно посмотреть, есть ли верхний предел того, насколько большой может стать интеллектуальная система памяти и в каких измерениях. Возможно, устройство может стать слишком хаотичным, чтоб его использовать, или оно могло бы перестать работать при достижении некоторого теоретического предела. Возможно человеческий мозг уже близок к максимальному теоретическому размеру, но мне это кажется слишком неправдоподобным. Человеческий мозг стал большим сравнительно недавно на по эволюционной временной шкале, и нет оснований полагать, что мы на некотором стабильном максимуме. Какой бы ни оказалась пиковая емкость интеллектуальной системы памяти, человеческий мозг определенно не достиг ее. Она, вероятно, не близко.

Один способ посмотреть, что такие системы могли бы сделать – это посмотреть на известные ограничения человеческой производительности. Эйнштейн несомненно был черезвычайно умным, но его мозг был все таки мозгом. Мы можем предположить, что его экстраординарная интеллектуальность была в основном следствием физических отличий его мозга и обычного человеческого мозга. Что сделало Эйнштейна таким необыкновенным, так это то, что человеческий геном не часто производит мозг, подобный его мозгу. Однако, когда мы будем разрабатывать кремниевый мозг, мы сможем его построить как захотим. Он мог бы быть уровня Эйнштейна, или даже умнее. Другой экстремальный пример – savants – могут показать нам другое возможное измерение интеллекта. Savants – это умственно отсталые люди, проявляющие замечательные способности, такие как фотографическая памяти или способность выполнять сложные математические вычисления с молниеносной скоростью. Их мозг, хотя не типичный, все еще мозг, работающие по тому же самому кортикальному алгоритму. Если нетипичный мозг может иметь ошеломляющие способности к запоминанию, то, теоретически, мы могли бы добавить эти способности нашему искусственному мозгу. Эти экстремальные ментальные человеческие способности не только показывают, что возможно воссоздать; они предоставляют направление, в котором мы могли бы превзойти наилучшие человеческие способности.

***Возможность копирования***

Каждый новый органический мозг должен быть выращен и натренирован с нуля, этот процесс занимает десятилетия человеческой жизни. Каждый человек должен открыть для самого себя основы координации конечностей тела и групп мышц, удержания равновесия и движения, запомнить общие свойства многочисленных объектов, животных, других людей; названия вещей и структуру языка; правила семьи и общества. Когда эти основы усвоены, наступают годы и годы формального обучения. Каждый человек должен пройти в жизни один и тот же набор кривых обучения, необходимых для построения модели мира в кортексе, даже хотя они были пройдены бесчисленное количество раз другими людьми.

Интеллектуальным машинам нет необходимости проходить этот длинный путь обучения, поскольку чипы и другие устройства хранения могут быть скопированы многократно и содержимое может быть легко перенесено. В этом смысле интеллектуальные машины могли бы копироваться как программное обеспечение. Как только единичный прототип системы был удовлетворительно настроен и натренирован, он мог бы быть скопирован многократно по нашему желанию. Могут уйти годы на разработку чипа, конфигурирование аппаратуры, тренировки, пробы и ошибки при совершенствовании системы памяти для умного автомобиля, но как только получен конечный продукт, можно начать его массовое производство. Как я уже упоминал ранее, мы могли бы выбрать, позволить копиям продолжать обучение или нет. Для некоторых приложений мы хотели бы ограничить интеллектуальные машины, чтоб они работали только протестированным и известным образом. Как только умная машина узнает все, что необходимо, мы не хотели бы, чтоб у нее развились вредные привычки или чтоб она поверила в ложные аналогии, которые ей показались бы. Мы ожидали бы, что все одинаковые машины вели бы себя одинаково. Но для других приложений, мы хотели бы, чтоб мозгоподобная система памяти сохраняла способность продолжать обучение. Например, интеллектуальная машина, разработанная для поиска математических доказательств, нуждалась бы в способности обучаться на опыте, применять старые догадки к новым проблемам, и была бы обобщенно гибкой и открытой.

Была бы необходима возможность совместного использования компонентов обучения, так же как это делается для программного обеспечения. Интеллектуальная машина конкретной разработки могла бы быть перепрограммирована с новым множеством соединений, ведущих к отличному поведению, как если бы я мог загрузить новый набор соединений в ваш мозг и изменить вас с англо-говорящего на франко-говорящего, или с профессора политических наук на музыковеда. Люди могли бы обменивать и строить работу других. Скажем, я разработал и натренировал машину с превосходной визуальной системой, а другой человек разработал и натренировал машину с превосходным слухом. При правильной разработке мы могли бы скомбинировать преимущества обеих систем без тренировки снизу вверх. Совместное использование жизненного опыта таким способом просто невозможно для людей. Направление разработки интеллектуальных машин могло бы эволюционировать тем же путем, что и компьютерная индустрия, с сообществами людей, тренирующих интеллектуальные машины на специализированные знания и способности, и людей, продающих и обменивающих результаты конфигурации памяти. Перепрограммирование интеллектуальной машины не отличалось бы от запуска новой видеоигры или инсталлирования нового программного обеспечения.

***Сенсорные системы***

У людей есть набор чувств. Эти чувства замурованы глубоко в наших генах, в наших телах и в субкортикальных соединениях нашего мозга. Мы не можем изменить их. Иногда мы используем технологию, чтобы усилить наши чувства, такие как очки ночного видения, радары или космический телескоп Хаббл. Эти высокотехнологические инструменты являются хитрыми трюками преобразования данных, но не новой моделью восприятия. Они преобразуют информацию, которую мы не можем ощущать, в визуальную или слуховую, которую уже мы можем интерпретировать. Тем не менее, необходимо отдать должное изумительной гибкости нашего мозга, которая проявляется в том, что мы можем смотреть на экран радара и понимать то, что он представляет. Большинство видов животных демонстрируют различные чувства, такие как эхолокация у летучих мышей и дельфинов, способность пчел видеть поляризованный и ультрафиолетовый свет, способность чувствовать электрическое поле у некоторых рыб.

Интеллектуальные машины могли бы воспринимать мир через любые органы чувств, как существующие в природе, так и чисто разработанные человеком. Сонар, радар и инфракрасное зрение очевидные примеры нечеловеческих видов органов чувств, которые мы могли бы захотеть использовать в интеллектуальных машинах. Но это только начало.

Гораздо более интересным был бы способ, которым интеллектуальные машины воспринимали бы мир экзотических, чуждых чувств. Как мы видели, неокортикальный алгоритм фундаментально занимается поиском паттернов в мире. Для него нет предпочтения к физическим источникам этих паттернов. Пока информация, поступающая в кортекс, является неслучайной и имеет определенную статистическую структуру, интеллектуальная система будет формировать инвариантные воспоминания и предсказания на их основе. Нет причины, по которой эти паттерны обязаны быть аналогичными чувствам животных, или даже вообще приходить из реального мира. Я ожидаю, что именно в области экзотических органов чувств лежит революционное использование интеллектуальных машин.

Например, мы могли бы разработать сенсорную систему, которая охватывает весь земной шар. Вообразите метеорологические датчики, расположенные примерно через каждые пятьдесят миль по континенту. Эти сенсоры были бы аналогичны сенсорам сетчатки. В любой момент времени, два смежных сенсора погоды имели бы высокую корреляцию их активности, подобно двум смежным нейронам сетчатки. Есть большие метеорологические объекты, такие как шторма и фронты, которые движутся и изменяются во времени, точно также как движутся и изменяются во времени визуальные объекты. Приделывая такой сенсорный массив к большой кортикальноподобной памяти, мы могли бы позволить системе научиться предсказывать погоду, точно так же, как мы учимся узнавать визуальные объекты и предсказывать, как они движутся во времени. Система видела бы локальные погодные паттерны, существующие часы, дни, годы. Размещая сенсоры поближе в некоторых регионах, мы могли бы создать некоторое подобие фовеа, позволяя интеллектуальному погодному мозгу понимать и предсказывать микроклимат. Наш погодный мозг понимал бы и думал бы о глобальной погодной системе так же, как я и вы понимаем и думаем об объектах и людях. Метеорологи сегодня занимаются чем то похожим. Они собирают записи с различных распределенных датчиков и используют суперкомпьютеры для моделирования климата и прогноза погоды. Но этот подход, который фундаментально отличается от того, как работали бы интеллектуальные машины, сродни тому как компьютер играет в шахматы – тупо и без понимания – тогда как интеллектуальная погодная машина сродни человеку, играющему в шахматы – вдумчиво и с пониманием. Интеллектуальная погодная машина открыла бы паттерны, которые недоступны человеку. Только в 1960 году был открыт погодный феномен, известный как El Niño. Погодный мозг мог бы найти гораздо больше паттернов, подобных El Niño, или обучиться тому, как предсказывать торнадо или муссоны гораздо лучше, чем человек. Располагая огромное количество погодных данных в форме, которая недоступна для понимания человеком, наш погодный мозг ощущал бы и думал о погоде непосредственно.

Другая сильно распределенная система сенсоров могла бы позволить нам построить интеллектуальную машину, понимающую и предсказывающую миграцию животных, демографические изменения, распространение болезней. Вообразите, что у нас есть сенсоры, распределенные по сетям линий электропередачи. Интеллектуальная машина, присоединенная к этим сенсорам наблюдала бы спады и потоки электропотребления таким же образом, как мы видим потоки дорожного движения, или движение людей в аэропорту. Через постоянное наблюдение этого человек обучается предсказывать эти паттерны – просто спросите об этом служащих, ездящих на транспорте или службу безопасности аэропорта. Аналогично интеллектуальный монитор сети электропередачи был бы способен лучше чем человек предсказывать потребности в электроэнергии, предсказывать ситуации, ведущие к перегрузке. Мы могли бы комбинировать сенсоры для погоды и человеческой демографии для того, чтобы предвидеть политические волнения, стихийные бедствия или вспышки болезней. Подобно очень умному дипломату, интеллектуальные машины могли бы играть роль в сокращении конфликтов и человеческих страданий. Вы могли бы подумать, что интеллектуальной машине нужны были бы эмоции, чтобы предсказывать паттерны, связанные с человеческим поведением, но я так не думаю. Мы не родились с набором культурных, религиозных ценностей; мы обучились им. И точно также, как я могу научиться понимать мотивацию людей с ценностями, отличными от моих, интеллектуальная машина могла бы постичь человеческие мотивации и эмоции, даже если сама по себе машина не имеет эмоций.

Мы могли бы изучать структуру бытия. Теоретически возможно сделать сенсоры, которые могли бы представлять паттерны в клетках или больших молекулах. Например, важная проблема сегодня – это понять, как форма молекулы белка может быть предсказана из последовательности аминокислот, составляющих этот белок. Умея предсказывать процесс свертывания и взаимодействия белков ускорило бы развитие медицины и вылечило бы многие болезни. Инженеры и ученые создают трехмерные визуальные модели белков, пытаясь предсказать, как поведут себя эти сложные молекулы. Но это слишком сложно. С другой стороны, супер интеллектуальная машина с набором сенсоров, специально подобранных для этой задачи, могла бы ответить на такие вопросы. Если это кажется слишком неестественным, вспомните, что мы не удивились бы, если человек смог бы решить эту проблему. Наша неспособность ухватить всю суть может быть базируется, в основном, на несоответствии между человеческими органами чувств и физическим феноменом, который мы хотим понять. Интеллектуальные машины могли бы иметь различные органы чувств и память, больше чем у человека, позволяющие им решать те проблемы, которые мы не можем решить.

С правильным набором чувств и слегка реструктурированной кортикальной памятью интеллектуальные машины могли бы жить и размышлять в виртуальных мирах, используемых в математике и физике. Например, большинство попыток в математике и науке требуют понимания того, как ведут себя объекты в мире, в котором больше чем три измерения. Теория струн, которая изучает структуру самого пространства, представляет Вселенную, как имеющую десять или более измерений. Для людей представляет большую сложность думать о математических проблемах в четырех или более измерений. Возможно интеллектуальная машина соответствующего дизайна могла бы понять многомерные пространства таким же образом, как мы понимаем трехмерные пространства, и, следовательно, была бы экспертом в предсказании того, как они себя поведут.

В конце концов мы могли бы использовать связку интеллектуальных машин в большой иерархии, точно так же как наш кортекс объединяет слух, осязание и зрение в высокоуровневую иерархию. Такая система автоматически училась бы моделировать и предсказывать мыслительные паттерны в популяции интеллектуальных машин. С распределенными системами передачи данных, такие как Интернет, отдельные интеллектуальные машины могли бы быть распределены по земному шару. Большая иерархия изучала бы более глубокие паттерны и видеть более сложные аналогии.

Цель этих размышлений показать, что есть множество способов, которыми мозгоподобные машины могли бы опередить наши способности. Они могли бы думать и учиться в миллион раз быстрее, чем можем мы, помнить огромное количество детальной информации или видеть невероятно абстрактные паттерны. У них могли бы быть более чувствительные сенсоры, чем у нас, или более распределенные сенсоры, или сенсоры для очень микроскопических феноменов. Они могли бы думать в трех, четырех или большем количестве измерений. Ни одна из этих интересных возможностей не зависит от того, что интеллектуальные машины подражают или действуют подобно людям, и они не требуют сложной робототехники.

Теперь мы можем полностью увидеть, как Тест Тьюринга сравнивая интеллект с человеческим поведением, ограничил наше видение возможного. Поняв в первую очередь, что такое интеллект, мы можем построить интеллектуальные машины, которые будут гораздо более ценны, чем простое копирование человеческого поведения. Интеллектуальные машины будут поразительным инструментом и сильно расширят наши знания о Вселенной.

\* \* \*

**К**огда сбудется что-либо из этого? Построим мы интеллектуальные машины через пятьдесят лет, через двадцать или через пять? В мире высоких технологий есть высказывание, что изменения идут дольше, чем вы ожидаете в краткосрочной перспективе, но возникают быстрее, чем вы ожидаете в долгосрочной перспективе. Я видел это много раз. Кто-то выскакивает на конференции, объявляет новую технологию и заявляет, что она будет в каждом доме через четыре года. Оказывается, что он ошибался. Четыре года превращаются в восемь, и люди начинают думать, что это никогда не произойдет. Спустя некоторое время, когда всем кажется, что идея совсем умерла, она начинает возрождаться и становится большой сенсацией. Что-то подобное должно произойти с интеллектуальными машинами. Поначалу прогресс кажется медленным, но потом начинает быстро набирать обороты.

На конференциях нейроученых мне нравится обходить зал и просить каждого высказать свое мнение о том, когда у нас будет работающая теория кортекса. Некоторые люди – меньше 5 процентов – говорят «никогда» или «у нас она уже есть» (неожиданный ответ). Другие 5 процентов говорят «через 10 лет». Половина оставшихся говорят от 10 до 50 лет, или «в течение моей жизни». Оставшиеся говорят от 50 до 200 лет, или «уже после моей жизни». Я на стороне оптимистов. Мы в течение десятилетий жили в «медленном» периоде, так что многим людям кажется, что прогресс в теоретической нейронауке и интеллектуальных машинах окончательно застрял. Опираясь на прогресс последних 50 лет естественно предположить, что мы никогда не приблизимся к ответу. Но я верю, что мы на поворотной точке и прогресс тронется с места.

Возможно ускорить будущее, чтобы приблизить поворотную точку. Одна из целей этой книги убедить вас, что при наличии корректных теоретических основ мы можем добиться ускоренного прогресса в понимании кортекса – что с моделью «память-предсказание» в качестве руководства мы можем дешифровать детали того, как работает мозг и наше мышление. Это знание, необходимое для построения интеллектуальных машин. Если это верная модель, прогресс может вскоре продолжиться.

Так что хотя я не могу точно предсказать, когда эра интеллектуальных машин станет реальность, я думаю, что если достаточное количество людей возьмутся за решение проблемы сегодня, мы сможем создать полезный прототип и эмулятор кортекса всего за несколько лет. В течение десяти лет, я надеюсь, интеллектуальные машины станут одной из самых горячих областей технологии и науки. Я не хочу уточнять, потому что я знаю, как легко недооценить время, требуемое для того, чтоб произошло что-то важное. Так почему же я столь оптимистичен в оценке скорости прогресса в понимании мозга и построении интеллектуальных машин? Моя вера коренится в основном на том, что я потратил уже довольно много времени на работу по проблеме интеллекта. Когда я впервые увлекся изучением мозга, я почувствовал, что решение этой головоломки может наступить при моей жизни. В течение многих лет я тщательно наблюдал спад ИИ, восхождение и падение нейронных сетей, и Декаду Мозга в 90-х годах. Я видел, как эволюционировали отношение к теоретической биологии и в особенности теоретической нейронауке. Я видел, как идеи предсказания, иерархического представления и время вползали в лексикон нейронауки. Я видел прогресс в моем собственном понимании и понимании у моих коллег. Я загорелся ролью предсказания 18 лет назад и с тех пор несколькими способами проверял ее. Поскольку я был погружен в нейронауку и компьютерную область свыше двух десятилетий, возможно мой мозг построил высокоуровневую модель того, как возникают технологические и научные изменения, и что модель предсказывает быстрый прогресс. Сейчас поворотный момент.

# Эпилог

**Астроном** Карл Саган любил говорить, что понимание чего-либо не уменьшает его интересность и загадочность. Множество людей боятся, что научное понимание повлечет за собой компромисс с удивительностью, как если бы знание высасывало бы вкус и цвет жизни. Но Саган был прав. Истина в том, что с пониманием мы обретаем больше комфорта в нашей роли во вселенной и одновременно вселенная становится более полноцветной и загадочной. Быть крошечным пятном в бесконечном космосе, живым, сознательным, интеллектуальным и творческим – это более интересно, чем жить на плоской ограниченной Земле в центре маленькой вселенной. Понимание того, как работает наш мозг не уменьшает интересности и загадочности вселенной, нашей жизни, нашего будущего. Наше изумление станет только глубже по мере применения этих знаний к пониманию самих себя, построению интеллектуальных машин и затем овладевания новыми знаниями.

Следовательно поиск понимания мозга и построение интеллектуальных машин это достойная попытка и логически следующий шаг для человечества.

Этой книгой я надеюсь соблазнить молодых инженеров и ученых к изучению кортекса, к принятию модели «память-предсказание» и построению интеллектуальных машин. С ее высоты искусственный интеллект был большим продвижением. У него были журналы, образовательные программы, книги, бизнес-планы и предприниматели. Нейронные сети аналогично создали огромное возбуждение, как область, возникшая в 80-х. Но научные основы, лежащие в ИИ и нейронных сетях не годились для построения интеллектуальных машин.

Я убежден, что сейчас у нас есть более многообещающий путь. Если вы учитесь в колледже или высшей школе, и эта книга сподвигла работать над этой технологией – построить первые действительно интеллектуальные машины, помочь этой индустрии стартовать – я поддерживаю вас в этом устремлении. Сделайте это реальностью. Один из ключевых моментов предпринимательского успеха в том, что вы должны безрассудно уйти с головой в новую область прежде чем станет ясно, что успех вам обеспечен на 100 процентов. Очень важно правильно выбрать время. Если вы начнете слишком рано, вам придется бороться. Если вы дождетесь, пока нерешительность уйдет – будет слишком поздно. Я верю, что сейчас самое время начать разрабатывать и строить кортикоподобные системы памяти. Эта область будет черезвычайно важна и для науки и для коммерции. Intel и Microsoft новой индустрии, построенные на иерархической памяти, начнут свою деятельность где-то в течение ближайших десяти лет. Предпринять попытку на этом этапе может быть финансово рискованно или интеллектуально требовательно, но цена попытки всегда такова. Я надеюсь, что вы присоединитесь ко мне вместе с остальными, кто принял вызов чтобы создать одну из величайших технологий, которые когда-либо видел мир.

# Приложение: проверяемые предсказания

**Каждая** теория должна вести к проверяемым предсказаниям, поскольку экспериментальная проверка единственно верный путь к определению правильности новой идеи. К счастью, модель «память-предсказание» основана на биологии и ведет к нескольким специфическим и новым предсказаниям, которые могут быть проверены. В этом приложении я перечислил предсказания, которые могут опровергнуть и/или подтвердить предположения, сделанные в этой книге. Эти материалы более сложные, чем в главе 6, и определенно не требуются для понимания остальной части книги. Некоторые предсказания могут быть сделаны только на бодрствующих животных или людях, потому что тесты включают ожидание и предсказание появления стимулов. Предсказания не упорядочены по важности.

**Предсказание 1**

Мы должны обнаружить нейроны во всех областях кортекса, включая первичный соматосенсорный кортекс, которые демонстрируют особую активность в предчувствии сенсорных событий, в противоположность реакции на сенсорное событие.

Например, лаборатория Тони Зейдора из Cold Spring Harbor Laboratory обнаружили нейроны в первичной слуховой коре крыс, которые возбуждаются, когда крыса ожидает услышать звук даже если звука нет (персональные соответствия). Это должно быть общим свойством кортекса. Мы должны найти аналогичную предсказывающую активность в визуальном кортексе и соматосенсорном. Нейроны, которые возбуждаются в предчувствии сенсорных событий являются определением предсказания, основной предпосылкой теории «память-предсказание».

**Предсказание 2**

Чем более пространственно-специфическим является предсказание, тем ближе к первичной сенсорной коре должны находиться нейроны, активизирующиеся в предчувствии события.

Если обезьяна обучена последовательностям визуальных паттернов, таких что она может предсказать определенный визуальный паттерн в конкретный момент, мы должны найти нейроны, демонстрирующие усиленную активность, когда ожидается предсказываемый паттерн (подтверждение предсказания 1). Если обезьяна обучена отличать лица, но не знает точно, какое именно лицо и как должно появиться, то мы должны ожидать обнаружения предчувствующих нейронов в области распознавания лиц, но не в нижестоящих визуальных областях. Однако, если обезьяна сосредотачивается на цели и обучена ожидать определенные паттерны в точном месте ее визуального поля, то мы должны обнаружить предчувствующие нейроны в V1 или поблизости от V1. Активность, представляющая предсказание, распространяется вниз по кортикальной иерархии так далеко, насколько может, в зависимости от специфики предсказания. Иногда она может пройти весь путь до первичных сенсорных областей, в других случаях она останавливается в высших областях. Аналогичные результаты должны наблюдаться и в других сенсорных модальностях.

**Предсказание 3**

Нейроны, демонстрирующие усиленную активность в предчувствии сенсорной информации, должны располагаться преимущественно в кортикальных слоях 2, 3 и 6, и предсказание должно останавливать свое движение вниз по иерархии в слоях 2 и 3.

Предсказание, которое идет вниз по кортикальной иерархии, делает это на нейронах слоев 2 и 3, которые затем проецируются на слой 6. Эти нейроны в слое 6 проецируются широко по слою 1 в областях ниже по иерархии, активируя другие множества нейронов в слоях 2 и 3, и так далее. Следовательно, нейроны в этих эти слоях (2, 3 и 6) располагаются именно там, где мы должны обнаружить предчувствующую активность. Вспомните, что нейроны в слоях 2 и 3 представляют множество возможно активных колонок; это возможные предсказания. Активные нейроны в слое 6 представляют небольшое количество колонок; это конкретные предсказания по данной области кортекса. По мере продвижения предсказания вниз по иерархии, активность в конечном итоге будет останавливаться в слоях 2 и 3. Например, скажем, крыса научилась предчувствовать один из двух звуковых тонов. Основываясь на внешнем стимуле, крыса знает, когда она должна услышать один из этих двух тонов, но она не может предсказать, какой именно из них. В этом сценарии мы должны ожидать увидеть предчувствующую активность в слоях 2 и 3, в колонках, которые представляют оба тона. В слое 6 той же области не должно быть активности, потому что крыса не может предсказать, какой конкретно тон она услышит. Если в другой попытке крыса может предсказать конкретный тон, то мы должны увидеть активность в слое 6, в колонке, которая отвечает за конкретный тон.

Мы не можем полностью исключить возможность обнаружения активности в слоях 4 и 5. Например, очень вероятно, что в этих слоях есть несколько классов нейронов с неизвестными функциями. Следовательно, эти предсказания относительно слабые, но я все же чувствую, что их стоит упомянуть.

**Предсказание 4**

Один класс нейронов в слоях 2 и 3 должен получать информацию преимущественно от нейронов слоя 6 в вышестоящих кортикальных областях.

Часть модели «память-предсказание» заключается в том, что заученные последовательности паттернов, которые возникают вместе, образуют постоянное во времени инвариантное представление, которое я называю «именем». Я предполагаю, что это имя – это набор нейронов в слоях 2 и 3 по всей области кортекса в различных колонках. Набор нейронов остается активным до тех пор, пока следуют члены последовательности (то есть, набор нейронов остается активным остается активным, пока слышатся ноты из мелодии). Этот набор нейронов, представляющий имя последовательности, активизируется через обратную связь от нейронов слоя 6 в вышестоящих областях кортекса. Я убежден, что эти нейроны имени располагаются в слое 2 по причине их близости к слою 1. Но это может быть любой класс нейронов в слоях 2 и 3, которые имеют нейроны в слое 1. Для работы системы именования апикальные дендриты этих нейронов имени должны формировать синапсы преимущественно на аксонах слоя 1, которые берут начало в слое 6 вышестоящей области. Они должны избегать формирования синапсов на аксонах слоя 1, которые берут начало в таламусе. Таким образом, теория убеждает, что мы должны обнаружить в слоях 2 и 3 класс нейронов с апикальными дендритами в слое 1, имеющие строгое предпочтение к формированию синапсов на аксонах от нейронов слоя 6 вышестоящей области. Другие нейроны с синапсами в слое 1 не должны иметь такого предпочтения. Это сильное, и, насколько мне известно, совершенно новое предсказание.

Дополнительное предсказание – мы должны обнаружить другой класс нейронов в слоях 2 и 3, чьи апикальные дендриты формируют синапсы преимущественно на аксонах, берущих начало в неспецифических областях таламуса. Эти нейроны должны предсказывать следующие члены в последовательности.

**Предсказание 5**

Набор нейронов «имени», описанный в предсказании 4, должен оставаться активным в течение всей запомненной последовательности.

Набор нейронов, остающихся активными в течение заученной последовательности – это определение «имени» предсказываемой последовательности. Следовательно, мы должны обнаружить нейроны, которые остаются активными, даже когда активность остальных нейронов колонки (нейроны в слоях 4, 5 и 6) изменяется. К несчастью, мы не можем сказать, на что должна быть похожа активность нейронов имени. Например, постоянная активность паттерна имени могла бы быть просто единичным спайком, когерентным по всему набору нейронов имени. Следовательно, может оказаться трудным обнаружить эту группу активных нейронов.

**Предсказание 6**

Другой класс нейронов в слоях 2 и 3 (отличающиеся от нейронов имени, упомянутых в предсказаниях 4 и 5) должны быть активными в ответ на неожиданную информацию, но должны быть неактивными в ответ на предсказуемую информацию.

Идея этого предсказания в том, что неожиданные события должны быть переданы вверх по кортикальной иерархии, но когда событие предсказуемо, мы не должны передавать его вверх по иерархии детально, потому что оно предсказывается локально. Следовательно, должен быть класс нейронов в слоях 2 и 3, отличающийся от нейронов имени, описанных в предсказаниях 4 и 5, демонстрирующие активность, когда возникает неожиданное событие, но не активизирующиеся, если событие предчувствовалось. Аксоны этих нейронов должны проецироваться на вышестоящие области кортекса. Я предполагаю один механизм для изменения их активности. Такие нейроны могли бы затормаживаться через интернейроны, активизируемые нейронами имени, но сейчас нет способа сделать точное предсказание механизма. Все, что мы можем сказать, это то, что некоторые нейроны должны демонстрировать такую дифференциальную активность. Это другое сильное, и насколько мне известно, новое предсказание.

**Предсказание 7**

В соответствии с предсказанием 6, непредсказуемые события должны распространяться вверх по иерархии. Чем более новым является событие, тем выше непредсказуемая информация должна распространяться. Совершенно новые события должны достигать гиппокампа.

Строго заученные паттерны предсказываются внизу иерархии, и, наоборот, чем более новой является информация, тем выше она должна распространиться вверх по иерархии. Должно быть возможно разработать эксперимент для детектирования этих различий. Например, человек мог бы слушать незнакомую, но простую мелодию. Если он слышит ноту, которая хотя и неожиданна, но удовлетворяет стилю музыки, неожиданная нота должна вызвать изменения в активности слухового кортекса, вверх до некоторого уровня по кортикальной иерархии. Однако, если вместо того, чтоб услышать ноту, удовлетворяющую стилю музыки, он слышит совершенно абсурдный звук, например, треск, мы должны ожидать, что изменения активности от этого звука пройдут до самого верха иерархии. Результаты должны переключиться, если он ожидает услышать треск, но вместо этого слышит ноту. Должно быть возможным протестировать предсказание с помощью фМРТ на человеке.

**Предсказание 8**

Неожиданное понимание должно привести к целому каскаду предсказательной активности, которая распространяется вниз по кортикальной иерархии.

Момент «ага», когда головоломный сенсорный паттерн наконец то узнается – такое как узнавание далматинца на рисунке 12 – начинается, когда область кортекса пробует сопоставить новую информацию. Если соответствие идет в локальной области – предсказание передается вниз в быстрой последовательности по всем нижестоящим областям. Если это корректная интерпретация стимула, то каждая нижестоящая область иерархии уляжется в корректное предсказание в быстрой последовательности. Тот же самый эффект должен возникать при рассмотрении картинки с двумя интерпретациями, такие как силуэт вазы, похожий на два лица, или куб Неккера (изображение куба, которое можно интерпретировать в двух ориентациях). Каждый раз, когда восприятие такого изображения изменяется, мы должны видеть распространение нового предсказующего потока вниз по иерархии. На самом нижнем уровне, скажем, в области V1, колонка, представляющая линейный сегмент изображения, должна оставаться активной при любом восприятии картинки (предполагая, что глаза не двигаются). Однако, мы могли бы увидеть, что некоторые нейроны в этих колонках переключают свое состояние. То есть, один и тот же низкоуровневый элемент существует в каждой картинке, но внутри колонки могут быть активными разные нейроны при различных интерпретациях. Основной момент в том, что мы должны увидеть распространение предсказующего потока вниз по иерархии, когда изменяется высокоуровневое восприятие.

Аналогичное распространение предсказания должно возникать при каждой саккаде по известному визуальному объекту.

**Предсказание 9**

Модель «память-предсказание» требует, чтобы пирамидальные нейроны могли детектировать точное совпадение синаптических импульсов на тонких дендритах.

Многие годы считалось, что нейроны могли бы быть просто интеграторами, суммирующими информацию со всех синапсов для определения того, должен ли нейрон возбудиться. Сегодня в нейрофизиологии много неопределенности относительно того, как ведет себя нейрон. Некоторые люди до сих пор придерживаются идеи, что нейрон просто интегратор, и большинство моделей нейронных сетей построены на нейронах, которые работают именно таким образом. Есть также множество моделей нейронов, предполагающих, что нейрон ведет себя так, как если бы каждая дендритная секция оперировала бы независимо. Модель «память-предсказание» требует, чтобы нейроны были способны обнаруживать совпадения только на нескольких активных синапсах в узком временном промежутке. Модель могла бы работать даже с единственным синапсом, потенциированным достаточно, чтоб вызвать возбуждение нейрона, но более вероятно, что должно быть два или больше активных синапсов, расположенных рядом на тонком дендрите. Таким образом нейрон с тысячами синапсов может научиться возбуждаться на множество различных более точных паттернов. Это не новая идея, и есть основания поддержать ее. Это, однако, радикальное отклонение от стандартной модели, используемой многие годы. Если будет показано, что нейрон не возбуждается на точные паттерны, будет сложно удержать модель «память-предсказание» нетронутой. Синапсы на толстых дендритах или вблизи тела нейрона не обязаны работать таким образом, только множественные синапсы на тонких дендритах.

**Предсказание 10**

Представление продвигается вниз по иерархии по мере обучения.

Я утверждаю, что через постоянное обучение кортекс запомнил бы последовательность в иерархически более низких областях кортекса. Это естественным образом следует из того, как память о последовательности изменяет входной паттерн, передаваемый в вышестоящую область. Из этого процесса есть несколько следствий. Одно в том, что мы должны обнаружить нейроны, отвечающие на сложные стимулы, ниже по кортикальной иерархии после усиленного обучения, и выше по иерархии после минимального обучения. Например, у человека я ожидал бы обнаружить нейроны, отвечающие на печатные буквы, в такой области, как IT, после обучения распознаванию отдельных букв. Но после обучения чтению целых слов, я ожидал бы обнаружить нейроны, отвечающие буквам, в различных частях V4 в дополнение к IT. Аналогичные результат должен быть достигаем и у других видов, в других областях и на другие стимулы. Другое следствие такого процесса обучения в том, что места, где возникают воспоминания и где детектируются ошибки, должны также перемещаться. То есть, ощущения сильно заученных паттернов должно распространяться на меньшее расстояние вверх по иерархии. Это должно быть обнаруживаемо с помощью аппаратуры отображения. Мы должны суметь детектировать изменения во времени реакции на определенные стимулы, потому что информация не обязательно идет до самой верхушки кортекса, чтоб быть распознанной и вспомненной.

**Предсказание 11**

Инвариантное представление должно быть найдено во всех кортикальных областях.

Широко известно, что существуют нейроны, высокоселективно отвечающие на информацию инвариантно ко множеству деталей. Наблюдают нейроны, отвечающие на лица, руки, Билла Клинтона и т.п. Модель «память-предсказание» предсказывает, что все области кортекса должны формировать инвариантные представления. Инвариантные представления должны отражать все сенсорные модальности нижестоящих областей кортекса. Например, если б у меня в визуальном кортексе был нейрон Билла Клинтона, он отвечал бы на любое изображение Билла Клинтона. Если б нейрон Билла Клинтона был у меня в слуховом кортексе, он отвечал в любом случае, когда я слышал бы имя «Билл Клинтон». Поэтому я ожидал бы обнаружить в ассоциативных областях нейроны, получающие и визуальную и слуховую информацию и отвечающие либо на изображение, либо на произнесение имени Билла Клинтона. Мы должны найти инвариантное представление во всех сенсорных модальностях, и даже в моторном кортексе. В моторном кортексе нейроны отображали бы сложные моторные последовательности. Чем выше по моторной иерархии, тем более сложные и более инвариантные представления должны быть. (Недавние исследования похоже обнаружили нейроны, активизирующие движение руки ко рту у обезьян). Это не новое предсказание. Большинство исследователей верят в общую идею, что инвариантные представления формируются во многих местах кортекса. Однако, даже хотя я обсуждал это как факт, это пока не было продемонстрировано по всем областям. Модель «память-предсказание» предсказывает, что мы увидим такие нейроны во всех областях кортекса.

\* \* \*

Упреждающие предсказания – это один из способов, которым может быть проверена модель из данной книги. Я уверен, что есть и другие. Однако, невозможно доказать корректность теории. Можно доказать только ее некорректность. Так что даже если все перечисленные выше предсказания окажутся верными, это не будет доказательством корректности гипотезы «память-предсказание», но это будет сильным подтверждением теории. Обратное также верно. Если некоторые из вышеперечисленных предсказаний окажутся ложными, это не обязательно опровергнет всю теорию. Для некоторых предсказаний есть альтернативные способы, которыми может достигаться необходимое поведение. Например, есть другие способы, которыми могли бы создаваться имена последовательностей. Это приложение намерено только показать, что модель ведет к нескольким предсказаниям, и, следовательно, может быть проверена. Разработка экспериментов – это трудная работа и она потребовала бы гораздо больше обсуждения, чем может быть приведено в данной книге. Также было бы неплохо, если б мы могли найти способ протестировать эту теорию с помощью техник отображения, таких как фМРТ. Есть множество лабораторий, занимающихся отображением мозга, и эти эксперименты могут быть выполнены относительно быстрее по сравнению с прямой записью активности нейронов.

# Библиография

**Большинство** научных книг и журнальных статей содержат длинную библиографию, которая столько же служит перечислению вклада других авторов, сколько помогает читателю. Поскольку данная книга предназначена для широкого круга читателей, включая тех, у кого нет предварительных знаний нейрофизиологии, я избегал писать книгу в академическом стиле. Аналогично, эта библиография разработана в основном, чтобы помочь неспециалистам, которые хотят знать больше. Я не перечислил все существенные опубликованные исследования, и не пытался доверять отдельным людям, сделавшим фундаментальные открытия в этой области. Вместо этого я выборочно перечислил те вещи, которые, по моему мнению, стали бы хорошим материалом для читателей, заинтересованных узнать больше о мозге. Я также включил несколько вещей, которые мне кажутся полезными в основном для специалистов. Вы можете найти всесторонние обсуждения большинства из этих статей в WWW. Дополнительные библиографические материалы могут быть найдены на веб-сайте этой книги, http://www.onintelligence.org.

К несчастью, вы найдете только некоторые ссылки на общие теории мозга, потому что, как я уже написал в прологе, по этой теме мало написано, и даже с меньшим количеством конкретных предположений, чем изложено в этой книге.

**История ИИ и нейронных сетей**

Baumgartner,

Peter, and Sabine Payr, eds. *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists* (Princeton, N. J.: Princeton University Press, 1995).

В этой книге содержатся интересные интервью с множеством ведущих мыслителей в области ИИ, нейронных сетей и когнитивной науки. Это легкий и интересный обзор недавней истории и образа мышления об интеллекте.

Dreyfus,

Hubert L. *What Computers Still Can't Do: A Critique of Artificial Reason* (Cambridge, Mass.: MIT Press, 1992).

Жесткая критика ИИ, изначально опубликованная под названием *What Computers Can't Do* и переизданная годы спустя с другим названием. Это всесторонний обзор истории ИИ, написанный одним из его сильных критиков.

Anderson,

James A., and Edward Rosenfeld, eds. *Neurocomputing, Foundations of Research* (Cambridge, Mass.: MIT Press, 1988).

Эта большая книга является аннотированной коллекцией важных статей по нейронным сетям и теории мозга, охватывающая годы с 1980 по 1987, представленных в хронологическом порядке. Она содержит статьи МакКаллока и Питтса, Дональда Хебба, Стива Гроссберга и многих других с примечанием редактора к каждой статье. Это легкий способ прочитать множество важных исторических статей в этой области.

Searle,

J. R. "Minds, Brains, and Programs," *The Behavioral and Brain Sciences*, vol. 3 (1980): pp. 417–24.

Представляет известный аргумент «Китайской Комнаты» в противовес вычислительной модели разума. В WWW вы можете найти множество описаний и дискуссий мысленного эксперимента Серла.

Turing,

A. M. "Computing Machinery and Intelligence," *Mind*, vol. 59 (1950): pp. 433–60.

Представляет известный «Тест Тьюринга» для детектирования наличия интеллекта. Опять же, многочисленные ссылки и обсуждения Теста Тьюринга могут быть найдены в WWW.

Palm,

Günther. *Neural Assemblies: An Alternative Approach to Artificial Intelligence* (New York: Springer Verlag, 1982).

Чтобы понять, как работает кортекс и как он хранит последовательности паттернов, полезно знать об автоассоциативной памяти. И хотя про автоассоциативную память было много написано, я не нашел ни одного печатного источника, который бы представлял простое суммарное изложение того, что я считаю важным. Палм является одним из пионеров в этой области. Эту его книгу очень трудно достать и не легко читать, но она покрывает основы автоассоциативной памяти, включая память для последовательностей.

**Неокортекс и Общая нейрофизиология**

Следующие книги рекомендуются для тех, кто хочет узнать больше о нейробиологии и неокортексе.

Crick,

Francis H. C. "Thinking about the Brain," *Scientific American*, vol. 241 (September 1979): pp. 181–88. Also available in *The Brain: A* Scientific American *Book* (San Francisco: W. H. Freeman, 1979).

Эта статья, которая заставила меня заинтересоваться мозгом. Хотя она 25-летней давности, я все еще нахожу эту статью Френсиса Крика вдохновляющей.

Koch,

Christof. *Quest for Consciousness: A Neurobiological Approach* (Denver, Colo.: Roberts and Co., 2004).

Каждый год публикуются несколько посвященных мозгу книг. Эта книга Кристофа Коха о сознании, но она покрывает многие сопутствующие статьи о мозге, нейроанатомии, нейрофизиологии и сознании. Если вы ищете базовое введение в нейробиологию и науку о мозге в одной книге, эта книга должна стать хорошим началом.

Mountcastle,

Vernon B. *Perceptual Neuroscience: The Cerebral Cortex* (Cambridge, Mass.: Harvard University Press, 1998).

Великолепная книга, посвященная все и вся о неокортексе. Она хорошо написана, ясно изложена и, хотя техническая, я нахожу ее приятной для чтения. Это одно из лучших введений в неокортекс.

Kandel,

Eric R., James H. Schwartz, Thomas M. Jessell, eds. *Principles of Neural Science*, 4th ed. (New York: McGraw-Hill, 2000).

Это однотомная энциклопедия по всем нейронным вещам. Эта книга не для чтения на ночь, а хороший источник ссылок. Она обеспечивает детальное введение во все области нервных систем, включая нейроны, органы чувств и нейромедиаторы.

Shepherd,

Gordon M., ed. *The Synaptic Organization of the Brain*, 5th ed. (New York: Oxford University Press, 2004).

Эта книга была мне полезна, хотя я предпочел бы более ранние издания с одним автором. Это технический ресурс по всем частям мозга, особенно по синапсам. Я использую ее как источник ссылок.

Koch,

Christof, and Joel L. Davis, eds. *Large-scale Neuronal Theories of the Brain* (Cambridge, Mass.: MIT Press, 1994).

Очень мало написано об общих теориях мозга. Эта книга – компиляция статей по указанной теме, хотя большинство статей в этом томе не соответствуют цели, предполагаемой по названию. Эта книга дает обзор различных подходов, предпринятых людьми для понимания того, как в целом работает мозг. В этой книге вы можете найти кусочки модели «память-предсказание».

Braitenberg,

Valentino, and Almut Schüz. *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2nd ed. (New York: Springer Verlag, 1998).

Эта книга описывает статистические свойства мозга мыши. Я знаю, что это звучит неинтересно, но это освежающая и полезная книга. Она дает численные описания мозга.

**Специфические статьи по нейронаукам**

Следующие статьи являются оригинальными источниками для некоторых важных концепций, описанных в этой книге. Большинство из них могут быть найдены только в университетских библиотеках или в интернете.

Mountcastle,

Vernon B. "An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System," in Gerald M. Edelman and Vernon B. Mountcastle, eds., *The Mindful Brain* (Cambridge, Mass.: MIT Press, 1978).

Именно в этой статье я впервые прочел предположение Монткастла о том, как на общих принципах целиком работает неокортекс. Монткастл также предполагает, что кортикальные колонки – это базовые модули вычислений. Эти идеи являются и исходными положениями и вдохновением для теории, предложенной в этой книге.

Creutzfeldt,

Otto D. "Generality of the Functional Structure of the Neocortex," *Naturwissenschaften,* vol. 64 (1977): pp. 507–17.

После того, как я закончил написание *On Intelligence*, я обратил внимание на эту статью, в которой, подобно Монткастлу, давался общий кортикальный алгоритм. Она была опубликована чуть раньше Монткастла и элегантно дополняет его.

Felleman,

D. J., and D. C. Van Essen. "Distributed Hierarchical Processing in the Primate Cerebral Cortex," *Cerebral Cortex*, vol. 1 (January/February 1991): pp. 1–47.

Это неоклассическая статья, описывающая иерархическую организацию визуального кортекса. Модель «память-предсказание» построена на предположениях, которые относятся не только к визуальной системе, но и весь неокортекс рассматривает как иерархическую структуру.

Sherman,

S.M., and R.W. Guillery. "The Role of the Thalamus in the Flow of Information to the Cortex," *Philosophical Transactions of the Royal Society of London*, vol. 357, no. 1428 (2002): pp. 1695–708.

Дает обзор таламической организации и излагает гипотезу Шермана-Гиллери, в которой таламус служит информационным шлюзом между кортикальными областями. Я обдумывал эту идею в главе 6 в секции, озаглавленной «Альтернативный путь по иерархии».

Rao,

R. P., and D. H. Ballard. "Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-field Effects," *Nature Neuroscience*, vol. 2, no. 1 (1999): pp. 79–87.

Я включил эту статью как пример недавнего исследования, которое касается предсказания и иерархии. Статья Рао и Балларда представляет модель обратных связей в кортикальной иерархии, в которой нейроны в вышестоящих областях пытаются предсказать паттерны активности в нижестоящих областях.

Guillery,

R. W. "Branching Thalamic Afferents Link Action and Perception," *Journal of Neurophysiology*, vol. 90 (2003): pp. 539–48.

Young,

M. P. "The Organization of Neural Systems in the Primate Cerebral Cortex," *Proceedings of the Royal Society: Biological Sciences*, vol. 252 (1993): pp. 13–18.

Эти две великолепно написанные книги дают подтверждение, что моторное поведение и сенсорное восприятие тесно связаны и являются частью одного и того же процесса. Гиллери утверждает, что все сенсорные кортикальные области играют роль в моторном поведении, а Юнг показывает, что моторный кортекс и соматосенсорный кортекс так плотно связаны, что должны рассматриваться как одна система. Я кратко обсуждал эту идею в главе 6.

# Благодарности

**Когда** кто-либо меня спрашивает, «зачем ты живешь?», я никогда не знаю, что сказать. По правде сказать, я сделал очень немного. Но меня окружают люди, которые, кажется, сделали очень многое. Мой вклад заключается в том, чтоб время от времени побуждать их, и при необходимости пытаться перенаправить усилия команды по новому пути. Успехом, которого я добился в своей карьере, я в основном обязан тяжелому труду и интеллекту моих коллег.

У меня был шанс встретить множество ученых и почти все из них научили меня чему-то, и, следовательно, они все внесли свой вклад в идеи этой книги. Я благодарен всем, но здесь могу упомянуть только некоторых. Бруно Ольшаузен, который сочетает работу в Редвудском Институте Нейронаук (RNI) и в Калифорнийском Университете в Дэвисе – это ходячая энциклопедия нейронаук. Он постоянно подсказывал мне, чего я не знаю и советовал мне пути для исправления моего невежества, что является одной из ценных вещей, которые может сделать человек. Билл Софтки, также из RNI, был первым человеком, подсказавшим мне идеи о сокращении времени в кортикальной иерархии и о свойствах тонких дендритов. Рик Грейнджер из Калифорнийского Университета в Ирвине подкинул мне идею о памяти на последовательности и о том, какую роль может играть таламус. Боб Найт из Калифорнийского Университета в Беркли и Кристоф Кох из Калифорнийского Технологического Института оказали пользу при формировании Редвудского Института Нейронаук во множестве других научных вещей. Все служащие RNI поставили мне задачу и вынудили меня усовершенствовать мои идеи; большинство предположений этой книги были прямым результатом встреч и дискуссий в RNI. Всем им большое спасибо.

Донна Дубински и Эд Коллига были моими деловыми партнерами много лет. С помощью их тяжелого труда и поддержки я смог стать предпринимателем, одновременно работая над теорией мозга. Донна обычно говорит, что одна из ее целей – сделать наш бизнес успешным, так чтоб я смог потратить больше времени на теорию мозга. Этой книги не было бы, если б не Донна и Эд.

Я не смог бы написать *On Intelligence* Без посторонней помощи. Джим Ливайн, мой агент, верил в эту книгу еще до того, как она была написана. Без такого агента, как Джим, не написать книгу. Он познакомил меня с Сандрой Блэйксли, моим соавтором. Я хотел, чтоб эта книга была доступна для широкой аудитории и Сэнди была необходима для достижения этого. Я принимаю любые упреки за сложные разделы. Мэтью Блэйксли, сын Сэнди и также автор научных книг, привел несколько примеров, использованных в этой книге и предложил термин *memory-prediction framework* (теория «память-предсказание»). Было приятно работать со всем семейством Генри Хольта. Я хотел бы особенно поблагодарить Джона Стерлинга, президента и издателя Генри Хольта. Я встречался с Джоном лицом к лицу только один раз, и несколько раз мы говорили по телефону. Это все, что ему потребовалось, чтобы оказать сильное воздействие на структуру этой книги. Он мгновенно понял суть того, с чем я должен был столкнуться предлагая теорию интеллекта и затем он предложил, как должна быть написана книга.

Я хотел бы поблагодарить моих дочерей, Энн и Кейт, за то что они не хныкали, пока их папа потратил множество выходных за клавиатурой компьютера. И, наконец, я хотел бы поблагодарить мою жену, Дженет. Быть замужем за мной нелегко. Я люблю ее больше, чем нейронауку.

# Об авторах

ДЖЕФ ХОКИНС

Один из наиболее успешных и высоко уважаемых разработчиков компьютеров и предпринимателей Силиконовой Долины. В настоящее время технический директор palmOne, он основал Palm Computing и Handspring, и создал Институт Нейронаук в Редвуде для содействия исследованиям памяти и познавательных способностей. Он является членом Национальной Академии Инженеров и входит в состав научной комиссии лаборатории Cold Spring Harbor. Он живет в северной Калифорнии.

САНДРА БЛЭЙКСЛИ

писала о науке и медицине для *New York Times* свыше тридцати лет и является соавтором бестселлера по психологии и браку *Phantoms in the Brain* В.С.Рамачандрана и Джуди Уоллерштейн. Она живет в Санта Фе, Нью Мексико.