

**РЭЙ КУРЦВЕЙЛ**

выдающийся изобретатель и футуролог,  
технический директор GOOGLE

# ЭВОЛЮЦИЯ РАЗУМА

ИЛИ

БЕСКОНЕЧНЫЕ ВОЗМОЖНОСТИ ЧЕЛОВЕЧЕСКОГО МОЗГА,  
ОСНОВАННЫЕ НА РАСПОЗНАВАНИИ ОБРАЗОВ



«Среди знакомых мне людей Рэй Курцвейл лучше всех умеет предсказывать будущее искусственного интеллекта».

**БИЛЛ ГЕЙТС**



Вы смогли скачать эту книгу бесплатно на законных основаниях благодаря проекту «**Дигитека**». [Дигитека](#) — это цифровая коллекция лучших научно-популярных книг по самым важным темам — о том, как устроены мы сами и окружающий нас мир. Дигитека создается командой научно-просветительской программы «[Всенаука](#)». Чтобы сделать умные книги доступными для всех и при этом достойно вознаградить авторов и издателей, «Всенаука» организовала всенародный сбор средств.

Мы от всего сердца благодарим всех, кто помог освободить лучшие научно-популярные книги из оков рынка! Наша особая благодарность — тем, кто сделал самые значительные пожертвования (имена указаны в порядке поступления вкладов):

Дмитрий Зимин  
Алексей Сейкин  
Николай Кочкин  
Роман Гольд  
Максим Кузьмич  
Арсений Лозбень  
Михаил Бурцев  
Ислам Курсаев  
Артем Шевченко  
Евгений Шевелев  
Александр Анисимов  
Михаил Калябин  
Роман Мойсеев  
Никита Скабцов  
Святослав Сюрин  
Евдоким Шевелев

Мы также от имени всех читателей благодарим за финансовую и организационную помощь:

Российскую государственную библиотеку  
Компанию «Яндекс»  
Фонд поддержки культурных и образовательных проектов «Русский глобус».

*Этот экземпляр книги предназначен только для вашего личного использования. Его распространение, в том числе для извлечения коммерческой выгоды, не допускается.*



**RAY KURZWEIL**

**HOW  
TO CREATE  
A MIND:**

**THE SECRET OF HUMAN  
THOUGHT REVEALED**

**РЭЙ КУРЦВЕЙЛ**

**ЭВОЛЮЦИЯ  
РАЗУМА**

**ИЛИ**

**БЕСКОНЕЧНЫЕ ВОЗМОЖНОСТИ ЧЕЛОВЕЧЕСКОГО МОЗГА,  
ОСНОВАННЫЕ НА РАСПОЗНАВАНИИ ОБРАЗОВ**



**БОМБОРА™**

Москва 2018

УДК 612+004.8  
ББК 28.707.3+32.813  
К93

Ray Kurzweil

HOW TO CREATE A MIND: THE SECRET OF HUMAN THOUGHT REVEALED

© 2012 by Ray Kurzweil. All rights reserved including the right of reproduction in whole or in part in any form. This edition published by arrangement with Viking, a member of Penguin Group (USA) Inc.

**Курцвейл, Рэй.**

**К93** Эволюция разума, Или бесконечные возможности человеческого мозга, основанные на распознавании образов / Рэй Курцвейл ; [пер. с англ. Т. П. Мосоловой]. — Москва : Эксмо, 2018. — 352 с. — (Большая наука).

В своей книге «Эволюция разума» известный американский футуролог, технический директор Google раскрывает бесконечный потенциал возможностей в сфере обратного проектирования человеческого мозга. Самый известный прогноз автора: в 2045 году должна произойти технологическая сингулярность, момент, после которого спрогнозировать что-то будет почти невозможно, т.к. к тому времени искусственный интеллект станет настолько мощным, что мозг не сможет его понять. Технический прогресс человечества выйдет на новый уровень. «Эволюцию разума» невозможно пропустить всем тем, кто интересуется развитием искусственного интеллекта или просто будущим цивилизации.

УДК 612+004.8  
ББК 28.707.3+32.813

Все права защищены. Книга или любая ее часть не может быть скопирована, воспроизведена в электронной или механической форме, в виде фотокопии, записи в память ЭВМ, репродукции или каким-либо иным способом, а также использована в любой информационной системе без получения разрешения от издателя. Копирование, воспроизведение и иное использование книги или ее части без согласия издателя является незаконным и влечет уголовную, административную и гражданскую ответственность.

Научно-популярное издание

БОЛЬШАЯ НАУКА

**Рэй Курцвейл**

**ЭВОЛЮЦИЯ РАЗУМА**

Главный редактор *Р. Фасхутдинов*

Ответственный редактор *Ю. Лаврова*. Художественный редактор *П. Петров*

В коллаже на обложке использованы фотографии: Sarah Holmlund, Veles Studio / Shutterstock.com  
Используется по лицензии от Shutterstock.com.

В оформлении обложки использована иллюстрация:

Aleksei Derin / Shutterstock.com

Используется по лицензии от Shutterstock.com

**ООО «Издательство «Эксмо»**

123308, Москва, ул. Зорге, д. 1. Тел.: 8 (495) 411-68-86. Home page: [www.eksmo.ru](http://www.eksmo.ru) E-mail: [info@eksmo.ru](mailto:info@eksmo.ru)

Өндіруші: «ЭКСМО» АҚБ Баспасы, 123308, Мәскеу, Зорге көшесі, 1 үйі. Тел.: 8 (495) 411-68-86.

Home page: [www.eksmo.ru](http://www.eksmo.ru) E-mail: [info@eksmo.ru](mailto:info@eksmo.ru). Тауар белгісі: «Эксмо»

**Интернет-магазин** : [www.book24.ru](http://www.book24.ru)

**Интернет-дуken** : [www.book24.kz](http://www.book24.kz)

Импортер в Республику Казахстан ТОО «РДЦ-Алматы».

Қазақстан Республикасындағы импорттаушы «РДЦ-Алматы» ЖШС.

Дистрибутор и представитель по приему претензий на продукцию, в Республике Казахстан: ТОО «РДЦ-Алматы»

Қазақстан Республикасында дистрибутор және өнім бойынша арыз-талаптарды қабылдаушының өкілі «РДЦ-Алматы» ЖШС,  
Алматы қ., Домбровский көш., 3-а, литер Б, офис 1.

Тел.: 8 (727) 251-59-90/91/92. E-mail: [RDC-Almaty@eksmo.kz](mailto:RDC-Almaty@eksmo.kz)

Өнімнің жарамдылық мерзімі шектелмеген.

Сертификация туралы ақпарат сайты: [www.eksmo.ru/certification](http://www.eksmo.ru/certification)

Сведения о подтверждении соответствия издания согласно законодательству РФ

о техническом регулировании можно получить на сайте Издательства «Эксмо»: [www.eksmo.ru/certification](http://www.eksmo.ru/certification)

Өндірген мемлекет: Ресей. Сертификация қарастырылмаған

Подписано в печать 31.05.2018. Формат 60x90<sup>1/16</sup>.

Печать офсетная. Усл. печ. л. 22,0.

Тираж экз. Заказ



ISBN 978-5-04-089904-3



9 785040 899043 >

ISBN 978-5-04-089904-3



В электронном виде книги издательства вы можете  
купить на [www.litres.ru](http://www.litres.ru)

**ЛитРес:**  
Электронная библиотека



© Мосолова Т.П., перевод на русский язык, 2015

© Оформление. ООО «Издательство «Эксмо», 2018

# СОДЕРЖАНИЕ

Введение .....	6
Глава первая. Мысленные эксперименты над миром .....	19
Глава вторая. Мысленные эксперименты над мышлением .....	33
Глава третья. Модель новой коры; теория мысленного распознавания образов .....	43
Глава четвертая. Биологическая новая кора .....	89
Глава пятая. Старый мозг .....	109
Глава шестая. Трансцендентные способности .....	127
Глава седьмая. Цифровая кора — модель биологической коры ....	141
Глава восьмая. Мозг как компьютер .....	205
Глава девятая. Мысленные эксперименты по анализу разума .....	229
Глава десятая. Закон ускорения отдачи в применении к мозгу .....	285
Глава одиннадцатая. Критика .....	305
Эпилог .....	319
Комментарии .....	326

## ВВЕДЕНИЕ

Просторней голубых Небес  
Мой мозг во много раз —  
В себя он с легкостью вместит  
И небосвод, и вас.  
Намного глубже моря он,  
Хоть море глубоко —  
Как губка, целый океан  
Впитает он легко.  
Он нужен Богу — чтобы Бог  
Творенье взвесить смог —  
И если с гирей он не схож,  
То лишь как звук и слог.

*Эмили Дикинсон\**

**Р**азум — самое важное явление во Вселенной; он способен выходить за границы физических законов и трансформировать мир. Человеческий разум позволил нам преодолеть ограничения нашей биологической природы и изменить самих себя. Человек — единственное животное, которое на это способно.

История человеческого разума началась с момента возникновения Вселенной, могущей создавать и кодировать информацию. Развитие Вселенной — тоже удивительная история. Современные физические модели зарождения Вселенной содержат десятки констант, которые должны иметь строго определенные значения. Иначе не было бы атомов, не было бы звезд и планет, не было бы мозга и книг о мозге. С трудом верится, что законы

---

\* Эмили Дикинсон (1830–1886) — американская поэтесса. Стихотворение в пер. А. Гаврилова. *Здесь и далее прим. пер.*

физики так точны, что могут предсказать эволюцию информации. Однако в противном случае мы с вами не могли бы говорить на эту тему. Там, где одни находят Божий промысел, другие видят мультивселенную, в которой происходит эволюция одних вселенных и гибель других, не несущих никакой информации. Но вне зависимости от того, каким образом образовалась наша с вами Вселенная, мы начинаем рассказ с того, что наш мир основан на информации.

Чтобы изложить историю эволюции, придется прибегать к все более и более абстрактным понятиям. Атомы (особенно атом углерода, образующий информационно богатые структуры за счет четырех связей с другими атомами) складывались в сложные молекулы. Постепенно физика породила химию.

Через миллиард лет появилась сложная молекула, называемая ДНК, а в ней записаны «программы», в соответствии с которыми образуются живые организмы. Так химия породила биологию.

Со все возрастающей скоростью в организмах эволюционировали коммуникационные сети, которые мы называем нервной системой и которые способны координировать все более и более сложные функции тел, а также их поведение, позволяющее организмам выжить. Нейроны нервной системы образовали головной мозг, способный осуществлять сложную мыслительную деятельность. Так возникла нейробиология, а мозг стал центром хранения и обработки информации. От атомов и молекул мы добрались до ДНК и мозга. Следующий шаг сделал только человек.

Мозг млекопитающих обладает одной способностью, которой нет ни у каких других животных: мы можем мыслить иерархически, понимать строение форм и рисунков, состоящих из различных элементов, представлять эти структуры в виде символов и использовать символы в еще более сложных структурах. Данный процесс реализуется частью головного мозга, называемой новой корой (неокортексом). У человека эта способность

развита настолько сильно, что можно говорить не о рисунках и формах, а об *идеях*. Путем бесконечного рекурсивного процесса мы способны создавать еще более сложные идеи. Этот широкий спектр рекурсивно связанных идей мы называем *знанием*. Только *Homo sapiens* обладает знаниями, которые эволюционируют, растут по экспоненциальному закону и передаются от одного поколения другому.

Наш мозг создал еще один уровень абстракции, на котором используется наш разум и отставленный большой палец руки; с помощью таких приобретений мы изменяем мир и создаем орудия труда. Эти орудия — новая форма эволюции, и на ее уровне нейробиология породила технологию. И благодаря нашим орудиям возможности расширения наших знаний беспредельны.

Первым изобретением человечества стал рассказ — мы создали разговорный язык, который позволил нам передавать мысли словами. Позднее, научившись выражать мысли с помощью специфических символов, мы создали письменность. Письменный язык значительно расширил способность нашего мозга накапливать и расширять знания, составленные из рекурсивно структурированных идей.

Существуют разные мнения относительно того, могут ли другие животные, например шимпанзе, выражать иерархические идеи с помощью речи. Шимпанзе способны выучить ограниченный набор языковых символов и использовать их для общения с дрессировщиками. Однако очевидно, что существуют пределы сложности структурных знаний, которыми могут оперировать обезьяны. Произносимые ими предложения представляют собой определенные простые последовательности существительных и глаголов, и им не дана присущая человеку способность беспредельного усложнения языковых конструкций. В качестве занимательного примера сложности человеческой речи советуем прочесть хотя бы одно из удивительных многостраничных предложений Габриэля Гарсиа Маркеса: его новелла «*Последнее путешествие корабля-призрака*» состоит из одного единственного

предложения, которое прекрасно воспринимается как на испанском, так и на английском языке<sup>1\*</sup>.

Главная мысль трех моих предыдущих книг («Эпоха мыслящих машин» (The Age of Intelligent Machines), написана в 1980-х и опубликована в 1989 г.; «Эпоха духовных машин» (The Age of Spiritual Machines), написана в конце 1990-х и опубликована в 1999 г., и «Сингулярность уже близка» (The Singularity Is Near), написана в начале 2000-х и опубликована в 2005 г.) заключается в том, что эволюционный процесс имеет тенденцию ускоряться (за счет усложнения абстракций), и его результаты по сложности и возможностям изменяются экспоненциально. Я называю этот феномен «законом ускорения отдачи» (ЗУО), и он имеет отношение как к биологической, так и к технологической эволюции. Самый яркий пример действия закона заключается в предсказуемом экспоненциальном росте объема памяти и производительности информационных технологий. Эволюция технологий привела к созданию компьютера, который, в свою очередь, позволил в значительной степени расширить наши знания и связать между собой информацию из разных областей знания. Интернет также является подходящим и выразительным примером способности иерархической системы охватывать большой объем информации, сохраняя при этом свою структуру. Сам мир устроен по принципу иерархии: деревья имеют ветви, на ветвях растут листья, на листьях есть жилки. Дома имеют этажи, на этажах расположены комнаты, в комнатах есть двери, окна, стены и полы.

Мы создали такие инструменты, которые позволили нам понять наше собственное биологическое строение. Мы быстро учимся обратному проектированию (воспроизведению) информационных процессов, лежащих в основе биологических

---

\* Рассказ многократно переводился и на русский язык; см. например, перевод Р. Рыбкина в сборнике «Последнее путешествие корабля-призрака» (М.: Пресс Лтд, 1994).

законов, включая законы функционирования нашего собственного мозга. Теперь у нас в руках есть «объектный код» жизни в виде человеческого генома, что само по себе является удивительным примером экспоненциального роста знаний: за последние двадцать лет объем секвенированных генетических последовательностей в мире ежегодно увеличивался почти вдвое. С помощью компьютеров мы можем моделировать процессы трансляции последовательностей ДНК в последовательности аминокислот, которые складываются в трехмерные белковые структуры, являющиеся основой всей биологии. По мере экспоненциального роста компьютерных возможностей увеличивается сложность белковых структур, упаковку которых мы можем моделировать. Мы также способны моделировать взаимодействия между белками, происходящие за счет трехмерных межатомных сил. Углубление наших биологических знаний является важнейшим условием открытия секретов разума, которым наградила нас эволюция, и создания с помощью этих биологических концепций еще более разумных технологий.

В настоящее время тысячи ученых и инженеров чрезвычайно активно работают над проектом, посвященным изучению механизма самого совершенного интеллектуального процесса — функционирования человеческого мозга. Возможно, это самый важный проект в истории машинной цивилизации. В книге «Сингулярность уже близка» я писал о том, что одним из следствий закона ускорения отдачи является отсутствие других разумных существ. Если бы они были, мы бы их заметили, учитывая сравнительно быстрый переход цивилизации от слабо развитой технологии (только представьте себе, что в 1850-х гг. в Америке самым быстрым способом доставки корреспонденции был «Пони-экспресс»<sup>\*</sup>) к технологии, которая вышла бы за пределы их планеты<sup>2</sup>. С этой точки зрения, обратное проектирование

---

<sup>\*</sup> Почтовая служба фирмы «Рассел, Мэйджорс энд Уодделл», которая использовала перекладных лошадей и индейских пони.

человеческого мозга можно рассматривать в качестве важнейшего проекта во всей Вселенной.

Цель проекта заключается в изучении принципов работы человеческого мозга и использовании этой информации для того, чтобы лучше понимать самих себя, ремонтировать мозг, если это необходимо, и — что непосредственно относится к теме данной книги — создавать еще более разумные машины. Вспомним, что инженерный подход заключается в многократном усилении природных эффектов. Рассмотрим, к примеру, достаточно слабое явление, описываемое законом Бернулли. Этот закон утверждает, что воздух оказывает чуть меньшее давление на движущуюся искривленную поверхность, чем на движущуюся плоскую поверхность. Ученые еще не до конца поняли, как математика закона Бернулли описывает закономерности подъема крыла, а инженеры ухватились за это тонкое различие, сконцентрировали его и создали авиацию.

Данную книгу я посвятил изложению принципа, который называю «теорией мысленного распознавания образов» (ТМРО; The Pattern Recognition Theory of Mind, PRTM) и который, я готов поспорить, описывает основной алгоритм функционирования новой коры мозга (ответственной за понимание, память и критическое мышление). Я рассказываю о том, как последние исследования в области нейробиологии, а также наши собственные мысленные эксперименты привели нас к неизбежному выводу, что этот принцип действует во всех отделах новой коры. Инженерное использование принципов ТМРО и ЗУО позволит значительно расширить мощь нашего собственного разума.

На самом деле этот процесс уже активно претворяется в жизнь. Можно назвать сотни задач и функций, которые считались исключительной прерогативой человека и которые сейчас могут выполняться компьютерами, причем обычно гораздо точнее и с гораздо большей производительностью. Каждый раз, когда вы отправляете электронную почту или звоните по

мобильному телефону, разумные алгоритмы передают информацию оптимальным образом. Когда вам делают электрокардиограмму, вы получаете поставленный компьютером диагноз, конкурирующий с диагнозом врача. То же самое справедливо для изображений клеток крови. Разумные алгоритмы обнаруживают жульничество с кредитными картами, ведут и сажают самолеты, управляют разумными системами вооружения, помогают создавать продукты и отслеживать их запасы, собирают продукцию на автоматизированных производствах и мастерски играют в такие игры, как шахматы.

Миллионы людей следили за тем, как компьютер компании IBM по имени Ватсон играл в «Джеопарди!»\* на человеческом языке и набрал больше очков, чем два лучших в мире игрока вместе. Нужно отметить, что Ватсон не только читал и «понимал» тонкости вопросов (например, каламбуры и метафоры), но находил необходимую для ответа на вопрос информацию среди миллионов страниц документов на человеческом языке, таких как «Википедия» и другие энциклопедии. Ему пришлось разбираться практически во всех областях человеческой интеллектуальной деятельности, таких как история, наука, литература, искусство и т. д. Сейчас IBM работает с компанией Nuance (прежнее название Kurzweil Computer Products — моя первая компания) над созданием новой версии Ватсона. Эта машина будет читать медицинскую литературу (включая все медицинские журналы и важнейшие медицинские блоги) и станет специалистом по диагностике и консультантом по медицинским вопросам. Для этого используется технология компании Nuance, позволяющая воспринимать медицинскую терминологию. Некоторые наблюдатели утверждают, что Ватсон на самом деле не «понимает» вопросов викторины или текстов в энциклопедиях, а только производит «статистический анализ». В книге я расскажу, что математические технологии, используемые для

---

\* Американская телевизионная викторина.

создания искусственного интеллекта (как у Ватсона или Сири\*), очень похожи на те, что эволюционировали естественным путем и воплотились в новой коре. Если понимание языка и другие процессы, осуществляемые с помощью статистического анализа, не считаются «истинным пониманием», можно сказать, что люди тоже ничего не понимают.

Способность Ватсона разумно воспринимать информацию на человеческом языке вскоре будет использована для создания новых поисковых систем. Люди уже говорят по телефону на своем языке (причем через айфон — именно с помощью Сири; в разработке этой системы также принимала участие компания Nuance). Эти помощники, воспринимающие человеческую речь, станут еще разумнее, когда будут применять такие методы, которыми пользуется Ватсон, да и сам Ватсон продолжает совершенствоваться.

Самодвижущиеся машины Google проехали 200 тыс. миль по многолюдным городам и центрам Калифорнии (и эта цифра, без сомнения, значительно возрастет к тому моменту, когда моя книга окажется на полках реальных и виртуальных магазинов). В современном мире существует множество других примеров использования искусственного интеллекта, а на горизонте возникают все новые и новые примеры.

Еще одно проявление закона ускорения отдачи заключается в том, что возможности пространственного разрешения структур головного мозга и количество информации о мозге ежегодно увеличиваются в два раза. Кроме того, мы теперь можем использовать эти данные для создания рабочих моделей и для моделирования участков мозга. Нам удалось осуществить обратное проектирование ключевых функций слуховой коры, в которой обрабатывается звуковая информация, зрительной коры, где обрабатывается зрительная информация, и мозжечка,

---

\* Speech Interpretation and Recognition Interface (Siri) — персональный помощник и вопросно-ответная система, разработанная компанией Apple.

отвечающего за определенные навыки (например, за поимку летящего мяча).

Главная задача проекта по изучению и моделированию головного мозга человека заключается в обратном проектировании новой коры мозга, которая отвечает за рекурсивное иерархическое мышление. Кора составляет 80% головного мозга человека и образована из повторяющихся структур, что позволяет нам создавать сколь угодно сложные идеи.

Моя теория мысленного распознавания образов описывает модель, с помощью которой человеческий мозг реализует эту свою способность, используя замечательную структуру, появившуюся в ходе биологической эволюции. Какие-то детали этого механизма мы пока понимаем не полностью, но мы знаем достаточно, чтобы создать алгоритм, приводящий к тем же результатам. Начав с изучения функций новой коры, мы продвинулись настолько, что можем значительно усилить ее возможности — таким же образом, как конструкторы самолетов усилили эффект принципа Бернулли. Принцип функционирования новой коры, безусловно, является важнейшей в мире идеей, поскольку новая кора способна не только аккумулировать знания и навыки, но и создавать новые знания. Ведь именно новая кора в конечном счете отвечает за все рассказы, все песни, все картины, все научные открытия и все другие результаты человеческой деятельности.

Нейробиология остро нуждается в теории, которая смогла бы связать воедино результаты всех обширных и разнообразных наблюдений. Никакая наука не может существовать без единой теории. В следующей главе я расскажу о том, как двое мечтателей соединили физику с биологией — два раздела науки, которые до них считались безнадежно разупорядоченными, и как можно применить эту теорию для изучения головного мозга.

Сегодня часто говорят о сложном строении человеческого мозга. Если осуществлять поиск с помощью Google, вы обнаружите около 30 млн ссылок на эту тему (однако указать точное

число цитирований невозможно, поскольку некоторые сайты цитируются много раз, а другие однократно). Джеймс Уотсон писал в 1992 г., что «головной мозг — это [для нас] последняя и самая серьезная биологическая преграда, самая сложная вещь из всех, что мы до сих пор открыли во Вселенной». И объяснял это тем, что мозг «содержит сотни миллиардов клеток, соединенных между собой триллионами связей. Мозг поражает разум»<sup>3</sup>.

Я согласен с Уотсоном в том, что изучение головного мозга — самая сложная биологическая задача, но тот факт, что в нем миллиарды клеток и триллионы связей, не обязательно указывает на особую сложность его изучения, если только нам удастся идентифицировать понятные (и воспроизводимые) структуры этих клеток и связей, особенно учитывая их избыточность.

Давайте подумаем, что означает быть сложным? Ну, например, сложен ли лес? Ответ зависит от точки зрения. Можно сказать, что лес состоит из тысяч деревьев и каждое дерево отличается от других. Кроме того, на каждом дереве тысячи веток, и все они тоже различны. Затем можно отметить причудливые формы этих ветвей и прийти к выводу, что лес настолько сложен, что и вообразить трудно.

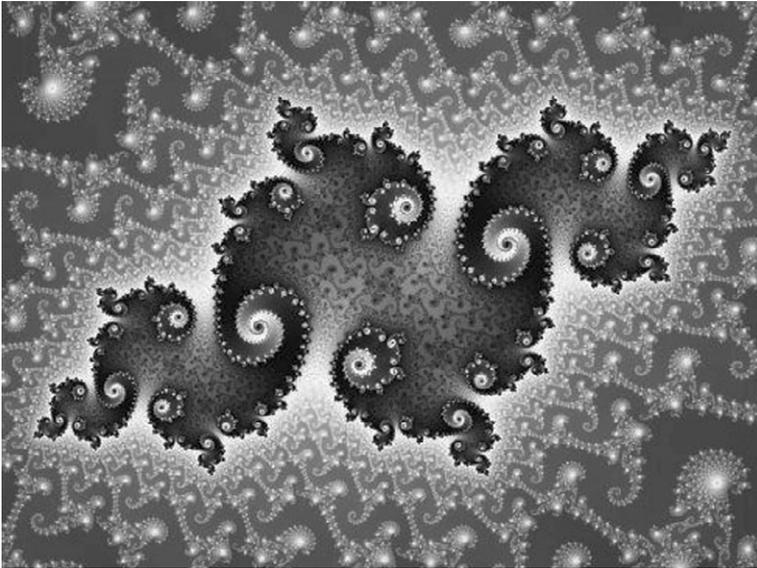
Но такая точка зрения буквально иллюстрирует выражение «не видеть леса за деревьями». Безусловно, все деревья и все ветви различаются, но, чтобы правильно понять принцип устройства леса, следовало бы начать с идентификации повторяющихся рисунков с имеющимися стохастическими (то есть случайными) вариациями. Можно сказать, что принцип устройства леса проще принципа устройства дерева.

То же самое относится и к головному мозгу, для которого характерна такая же невероятная избыточность, особенно в области новой коры. И как я покажу ниже, вполне можно сказать, что один нейрон устроен сложнее, чем вся новая кора.

Моя задача в этой книге заключается не в том, чтобы в миллион первый раз сообщить о сложной структуре мозга, а скорее

в том, чтобы удивить вас мощью его простоты. Я расскажу, как оригинальные базовые механизмы распознавания, запоминания и предсказания, повторяющиеся в новой коре сотни миллионов раз, обеспечивают всю широту наших мыслительных способностей. Как фантастическое разнообразие организмов возникает в результате различных комбинаций букв генетического кода в ядерной и митохондриальной ДНК, так и фантастическое множество идей, мыслей и навыков создается из образов (синаптических связей и потенциалов), распознаваемых специализированными модулями новой коры. Нейробиолог Себастьян Сеунг из Массачусетского технологического института говорит: «Личность определяется не нашими генами, а связями между клетками нашего мозга»<sup>4</sup>.

Нужно понимать разницу между истинной сложностью строения и кажущейся внешней сложностью. Рассмотрим, к примеру, знаменитое множество Мандельброта, которое долгое время считалось символом сложности. Чтобы оценить его кажущуюся сложность, увеличьте его изображение (обратитесь к ссылке, приведенной в комментарии<sup>5</sup>). Вы увидите бесконечное множество рисунков внутри рисунков, и все они различаются между собой. Но строение (формула) множества Мандельброта чрезвычайно просто и описывается последовательностью из знаков:  $Z = Z^2 + C$ . Не нужно полностью понимать смысл функции Мандельброта, чтобы оценить, насколько она проста. Этот закон реализуется многократно и на всех уровнях иерархии. То же самое справедливо для мозга. Его повторяющаяся структура не так проста, как формула множества Мандельброта, но и далеко не так сложна, как можно заключить на основании миллионов ссылок о строении головного мозга. Основная структура новой коры повторяется вновь и вновь на каждой иерархической ступени. Задачу, которую я попытался решить в данной книге, озвучил Эйнштейн, сказавший: «Любой дурак может увеличить и усложнить проблему... но нужно много храбрости, чтобы сделать наоборот».



*Вариант изображения множества Мандельброта — результата многократного повторения простой математической формулы. При приближении к какой-либо точке кажется, что изображение постоянно меняется непредсказуемым образом.*

До сих пор я говорил о мозге. Но что можно сказать о разуме? Например, как связана активность новой коры с сознанием? И сколько осознанных мыслей в нашем мозге, когда мы обдумываем ту или иную проблему? Кажется, иногда бывает больше одной.

Еще один важный вопрос о разуме: что такое свобода воли и есть ли она у нас? Некоторые эксперименты показывают, что мы начинаем реализовывать задуманное даже прежде, чем осознаём, что приняли решение. Означает ли это, что свобода воли — лишь иллюзия?

Наконец, какие атрибуты мозга определяют личность? Тот ли я человек, которым был полгода назад? Ясно, что не совсем тот, но сохранил ли я собственную личность?

Посмотрим, как теория мысленного распознавания образов помогает ответить на эти извечные вопросы.



# Глава первая

## МЫСЛЕННЫЕ ЭКСПЕРИМЕНТЫ НАД МИРОМ

Дарвиновская теория естественного отбора появилась очень поздно в истории развития человеческой мысли. Связано ли ее запоздалое возникновение с тем, что она противоречила очевидным истинам, была абсолютно новым предметом в истории науки, касалась только живых существ и рассматривала лишь причинно-следственные связи, но ничего не говорила о сотворении мира? Я думаю, причина в другом. Дарвин открыл роль отбора — причинной связи, которая очень сильно отличается от пушпульных механизмов, знакомых науке в то время. Происхождение фантастического разнообразия живых существ объяснялось появлением новых признаков, возможно случайным, которые позволяли этим существам выжить. В физике и биологии практически не существовало никаких доказательств причинной функции отбора.

*Б. Ф. Скиннер\**

Ничто так не свято, как чистота вашего разума.

*Р. У. Эмерсон\*\**

---

\* Беррес Фредерик Скиннер (1904–1990) — выдающийся американский психолог и писатель.

\*\* Ральф Уолдо Эмерсон (1803–1882) — американский писатель, поэт, философ и общественный деятель, один из виднейших американских мыслителей.

## **Аналогия из геологии**

В начале XIX в. геологи заинтересовались фундаментальным вопросом. Во всем мире были обнаружены гигантские пущоты и каньоны, такие как Большой каньон в США и каньон Викос в Греции (который считается самым глубоким каньоном в мире). Как возникли такие грандиозные структуры?

Конечно, в каждом случае на дне ущелий протекала вода, которая использовала этот естественный путь, но до середины столетия совершенно абсурдной казалась бы мысль, что эти ручейки и были причиной появления гигантских провалов и расщелин. Однако британскому геологу Чарлзу Лайелю (1797–1875) пришло в голову, что за очень длительный срок именно вода могла вызвать столь значительные изменения ландшафта. Сначала эту идею высмеивали, но за двадцать лет она завоевала всеобщее признание.

Одним из тех, кто внимательно следил за реакцией общественности на гипотезу Лайеля, был английский натуралист Чарлз Дарвин (1809–1882). Что происходило в биологии в 1850-х гг.? Эта область исследований казалась чрезвычайно запутанной, учитывая бесконечное множество видов животных и растений, каждое из которых обладало сложным строением. И в большинстве своем ученые даже не пытались создать общую теорию, объясняющую разнообразие живых организмов. Это разнообразие считалось доказательством божественного происхождения мироздания (и его анализ оказался не под силу ученым, не сумевшим в нем разобраться).

Дарвин подошел к проблеме происхождения видов, используя тезис Лайеля о постепенных изменениях. Эту аналогию в сочетании с собственными экспериментами и наблюдениями Дарвин изложил в своем знаменитом «Путешествии на Бигле»\*.

---

\* «Путешествие натуралиста вокруг света на корабле “Бигль”» (1839); первый русский перевод осуществлен в 1871 г. Е. Бекетовой.

Дарвин утверждал, что в каждом поколении те особи, которые лучше всего приспособлены к условиям своей экологической ниши, дают жизнь следующему поколению.

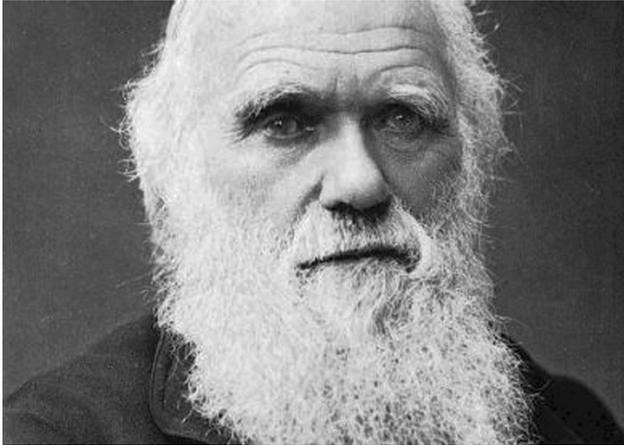
Труд Дарвина «О происхождении видов» появился в продаже 22 ноября 1859 г. В книге Дарвин отдает должное Лайелю: «Я вполне сознаю, что это учение о естественном отборе, поясненное вышеприведенными вымышленными примерами, может встретить те же возражения, которые были впервые выдвинуты против великих идей сэра Чарлза Лайеля о “современных изменениях на земной поверхности, объясняющих нам геологические явления”»; но теперь мы редко слышим, чтобы факторы, которые находятся еще в действии, признавались ничтожными и ничего не значащими, когда идет речь о причинах образования глубочайших речных долин или формирования внутриматериковых длинных скалистых гряд. Естественный отбор действует только путем сохранения и кумулирования малых наследственных модификаций, каждая из которых выгодна для сохраняемого существа; и как современная геология почти отбросила такие воззрения, как, например, прорытие глубокой долины одной делювиальной волной, так и естественный отбор изгонит веру в постоянное творение новых органических существ или в какую-либо большую и внезапную модификацию»<sup>\*1</sup>.

Всегда есть множество доводов, объясняющих неприятие новых и важных идей, и в случае теории Дарвина их несложно идентифицировать. Мысль о том, что мы произошли не от Бога, а от обезьяны, а еще раньше от червя, мало кому понравилась. Многие сочли богохульством считать родственником своего домашнего пса или гусеницу, не говоря уже о растениях,

---

\* Дарвин Ч. Происхождение видов путем естественного отбора, или Сохранение благоприятных рас в борьбе за жизнь (здесь и далее цитаты приводятся в соответствии с русским переводом шестого издания книги; СПб. : Наука, 1991).

по которым она ползает (конечно, очень-очень-очень дальним, но все же родственником).



*Чарльз Дарвин, автор книги «О происхождении видов»,  
обосновавший теорию биологической эволюции.*

Однако идея быстро распространилась, поскольку объединяла то, что долгое время считалось множеством несвязанных наблюдений. В шестом издании книги, которое увидело свет в 1872 г., Дарвин писал следующее: «В качестве напоминания о прежнем положении вещей я сохранил в предшествующих параграфах и в других местах несколько строк, указывающих на то, что натуралисты верят в отдельное сотворение каждого вида, и меня сильно осуждали за то, что я выражался таким образом. Но не подлежит сомнению, что таково было общее убеждение, когда появилось первое издание этой книги. ...Теперь положение совершенно иное, и почти каждый натуралист допускает великий принцип эволюции»<sup>2</sup>.

На протяжении следующего столетия обобщающая идея Дарвина углублялась и развивалась. Всего через десять лет, в 1869 г., швейцарский врач Фридрих Мишер (1844–1895) открыл в клеточном ядре вещество, которое он назвал «нуклеином», а мы теперь называем ДНК<sup>3</sup>. В 1927 г. русский биолог Николай Кольцов

(1872–1940) описал «гигантскую наследственную молекулу», которая состояла из «двух зеркальных нитей, способных реплицироваться по полуконсервативному механизму, используя каждую нить в качестве матрицы». Эти идеи также многими не были приняты. Коммунисты считали их вражеской пропагандой, и неожиданную кончину Кольцова связывают с деятельностью КГБ\*<sup>4</sup>. В 1953 г., почти через сто лет после публикации основополагающего труда Дарвина, американский биолог Джеймс Уотсон (род в. 1928 г.) и английский биолог Френсис Крик (1916–2004) впервые описали структуру ДНК, представляющую собой спираль из двух скрученных между собой длинных молекул<sup>5</sup>.

Следует заметить, что их открытие было основано на так называемой фотографии 51 — рентгенограмме, полученной Розалиндой Франклин и являющейся первым изображением двойной спирали. Учитывая вклад Франклин в открытие двойной спирали, высказывалось мнение о том, что она должна была разделить Нобелевскую премию с Уотсоном и Криком<sup>6</sup>.

Описание молекулы, которая может кодировать биологические программы, подтвердило унифицирующую теорию биологии. Эта молекула просто и элегантно объясняет структуру всего живого. В зависимости от последовательности оснований в нити ДНК в ядре (и, в меньшей степени, в митохондриях) организм станет травинкой или человеческим существом. Данное открытие не умаляет изумительного разнообразия природы, но позволяет нам понять, что все разнообразие определяется вариантами структур, которые записаны в этой уникальной молекуле.

---

\* Николай Константинович Кольцов — выдающийся русский биолог, основатель русской советской школы экспериментальной биологии, автор идеи матричного синтеза хромосом; в 1920 г. был арестован по делу так называемого Тактического центра и приговорен к расстрелу, но освобожден по ходатайству М. Горького; скончался от инфаркта после снятия с должности директора им же созданного Института экспериментальной биологии (теперь Институт биологии развития РАН) и допросов по делу Н. И. Вавилова.

## Прокатимся на луче

В начале XX в. серия другого рода мысленных экспериментов привела к перевороту в физике. В 1879 г. в семье немецкого инженера и домохозяйки родился сын. Говорить он начал только в три года, а в девять лет плохо справлялся со школьной программой. В шестнадцать он мечтал о полете на луче лунного света.

Юноша знал об эксперименте, в 1803 г. проведенном английским математиком Томасом Юнгом (1773–1829) и показавшем, что свет состоит из волн. В те времена считалось, что световые волны распространяются в некоей среде, подобно тому, как морские волны бегут по воде, а звуковые — по воздуху и другим материалам. Ученые называли среду, по которой распространяется свет, «эфиром». Юноша знал и об эксперименте американских ученых Альберта Майкельсона и Эдварда Морли, в 1887 г. пытавшихся подтвердить наличие эфира. Их эксперимент строился на аналогии с плаванием на гребной лодке вверх и вниз по реке. Если вы гребете с постоянной скоростью, скорость движения лодки, измеренная с берега, будет больше, когда вы движетесь по течению, чем когда вы гребете против течения. Майкельсон и Морли считали, что свет движется через эфир с постоянной скоростью (то есть со скоростью света). Они предположили, что скорости света при движении Земли по своей орбите вокруг Солнца в одну и в другую сторону должны различаться (на двукратную величину скорости вращения Земли). Это подтвердило бы существование эфира. Однако никаких различий в скорости солнечного света при любом положении Земли на орбите они не обнаружили. Их эксперимент доказал несостоятельность идеи «эфира», но не объяснил, как же распространяется свет. Ответа на этот вопрос ждали еще около двадцати лет.

Когда немецкий юноша мечтал о путешествиях на световом луче, он рассуждал так, что следует рассмотреть световые волны «остановившимися»: представьте себе, что вы движетесь

параллельно движущемуся поезду с такой же скоростью — тогда вам покажется, что поезд стоит на месте. Поэтому он попытался представить себе, что будет, если двигаться вдоль светового луча, но с несколько меньшей скоростью. Допустим, он движется со скоростью, которая составляет 90% скорости света. Если лучи света можно сравнить с поездами, то он должен был бы видеть свет, движущийся перед ним со скоростью, составляющей 10% скорости света. По крайней мере, это должны были бы видеть наблюдатели на Земле. Но мы знаем, что скорость света постоянна, как показал эксперимент Майкельсона — Морли. Таким образом, наблюдатель должен был видеть свет, движущийся впереди «на полной скорости» света. Тут какое-то противоречие — как такое может быть?

Ответ был получен немецким юношей (как вы могли догадаться, это был Альберт Эйнштейн), когда ему исполнилось 26 лет. Очевидно, для молодого Эйнштейна *замедлилось само время*. Он объяснял свои рассуждения в статье, опубликованной в 1905 г.<sup>7</sup> Если наблюдатели на Земле смогли бы смотреть на часы молодого человека, они бы обнаружили, что они идут в 10 раз медленнее. И когда он вернулся бы на Землю, его часы прошли бы лишь 10% того времени, которое длилось его отсутствия для земных наблюдателей (в данном случае мы не учитываем ускорение и замедление). Однако для самого юноши часы шли нормально, а луч света перед ним двигался со скоростью света. Десятикратное замедление времени (по отношению к земным часам) объясняет кажущееся противоречие. Если рассматривать крайний случай, когда скорость движения равна скорости света, время бы остановилось, хотя двигаться наравне со световым лучом невозможно. Однако теоретически ничто не запрещает двигаться *быстрее* света, и в таком случае время пойдет вспять.

Сначала такие утверждения многим показались абсурдными. Как может время замедляться в зависимости от скорости движения? Действительно, на протяжении 18 лет, которые

прошли со времен эксперимента Майкельсона — Морли, никто не смог прийти к выводу, который казался Эйнштейну таким очевидным. Многие ученые, пытавшиеся разобраться с этой проблемой на протяжении всего XIX в., можно сказать, «падали с лошади», цепляясь за привычные представления о том, как должен быть устроен мир (наверное, лучше сказать «падали со светового луча»).

Второй мысленный эксперимент Эйнштейна заключался в том, что он представил себя и своего брата в космосе. Между ними расстояние 300 тыс. км. Эйнштейн хочет двигаться быстрее, но при этом сохранять между собой и братом одно и то же расстояние. Он подает брату сигнал фонариком каждый раз, когда хочет ускориться. Поскольку он знает, что до брата сигнал дойдет только через секунду, он подает сигнал, ждет секунду и только потом ускоряется. Каждый раз, когда брат получает сигнал, он тут же ускоряется, а так как братья ускоряются точно в одно и то же время, расстояние между ними не меняется.

Но что увидим мы, наблюдая за ними с Земли? Если братья удаляются от нас (Альберт впереди), нам покажется, что до брата свет доходит быстрее, чем за секунду, поскольку он движется навстречу свету. Кроме того, мы увидим, что часы брата замедляются (по мере увеличения его скорости, поскольку он к нам ближе). Ввиду этих двух причин мы будем видеть, что братья постепенно сближаются и наконец сталкиваются. А с точки зрения братьев, они по-прежнему находятся на расстоянии 300 тыс. км друг от друга.

Как это возможно? Ответ — *очевидно* — заключается в том, что расстояния в направлении параллельном (но не перпендикулярном) направлению движения сокращаются. Таким образом, в результате ускорения братья Эйнштейн становятся меньше ростом (при условии, что они передвигаются головой вперед). Это странное заключение, возможно, стоило Эйнштейну большего числа поклонников его теории, чем выводы об изменениях времени.

В том же году Эйнштейн исследовал вопрос о связи материи и энергии. В 1850-х гг. Джеймс Клерк Максвелл показал, что частицы света, называемые фотонами, не имеют массы, но обладают импульсом. В детстве у меня было устройство под названием радиометр (или вертушка) Крукса<sup>8</sup>, которое состояло из четырех лопастей, вращающихся на острие иглы внутри герметичной стеклянной колбы с разреженным воздухом. Лопасты с одной стороны были черными, а с другой — белыми. Белая сторона каждой лопасти отражала свет, а черная поглощала (именно поэтому в жару прохладнее в белой футболке, чем в черной). Когда на вертушку попадал свет, лопасти вращались таким образом, что черные стороны удалялись от источника света. Этот эксперимент показывает, что фотоны обладают достаточным импульсом, чтобы заставить вращаться лопасти прибора<sup>9</sup>:



*Радиометр Крукса — вертушка с четырьмя лопастями, которая вращается под действием света.*

На основании подобных наблюдений Эйнштейн сделал вывод, что импульс зависит от массы: импульс равен массе тела,

помноженной на скорость его движения. Таким образом, паровоз, идущий со скоростью 50 км/ч, обладает гораздо большим импульсом, чем, скажем, насекомое, передвигающееся с той же скоростью. Однако как объяснить наличие положительного импульса у частицы, масса которой равна нулю?

Мысленный эксперимент Эйнштейна состоял в исследовании поведения коробки, находящейся в космосе. Внутри коробки происходит испускание фотона, который движется слева направо. Суммарный импульс системы должен сохраняться, поэтому при испускании фотона коробка смещается влево. Через какое-то время фотон ударяется о правую стенку коробки, возвращая импульс коробке. И вновь общий импульс системы должен сохраняться, так что коробка перестает двигаться.

До сих пор все нормально. Но что видит господин Эйнштейн, находящийся снаружи? Он не видит ничего, что происходит внутри коробки. Никакие частицы (с массой или без массы) в нее не попадают, и никакие частицы из нее не выходят. В соответствии с описанным выше сценарием видно только, что коробка какое-то время движется влево, а затем останавливается. В рассмотренной нами ситуации каждый фотон должен сдвигать коробку влево. Поскольку на коробку не производится никакого внешнего воздействия и ее содержимое не оказывает никакого влияния на внешнюю среду, центр массы коробки не должен смещаться. И движущийся внутри коробки фотон не может сместить центр массы коробки, поскольку он не имеет массы.

Или имеет? Эйнштейн пришел к выводу, что, поскольку фотон определенно обладает энергией и импульсом, он должен иметь массу. Энергия движущегося фотона пропорциональна его массе. Рассчитать коэффициент пропорциональности можно, если учесть, что при перемещении фотона центр массы системы не смещается. Математическим путем Эйнштейн показал, что энергия и масса связаны между собой

простым коэффициентом. Но в том-то и дело, что простым, но огромным: этот коэффициент равен скорости света в квадрате (примерно  $1,7 \times 10^{17} \text{ м}^2/\text{с}^2$ , то есть 17 с шестнадцатью нулями). Так получается<sup>10</sup> знаменитое уравнение Эйнштейна  $E = mc^2$ . Это означает, что одна унция (14 г) массы эквивалентна 6 млн т тротила. Письмо Эйнштейна президенту США Рузвельту, отправленное 2 августа 1939 г., в котором ученый сообщал о возможности создания атомной бомбы, обозначило начало атомного века<sup>11</sup>.

Возможно, вы думаете, что это было очевидно и ранее, учитывая эксперименты с радиоактивными веществами, масса которых уменьшается в процессе радиации. Однако считалось, что радиоактивные элементы содержат некое специфическое высокоэнергетическое «топливо», которое сгорает. Это предположение нельзя назвать в корне неверным. Просто дело в том, что этим «сгорающим топливом» является масса.

Почему я начал книгу с описания мысленных экспериментов Дарвина и Эйнштейна? Во-первых, по той причине, что они раскрывают фантастическую мощь человеческого мозга. Без помощи каких-либо инструментов, за исключением пера и бумаги для изображения схем этих простых мысленных экспериментов и выведения из них простых формул, Эйнштейн перевернул формировавшиеся столетиями представления о физическом мире, оказал глубокое влияние на ход истории (включая Вторую мировую войну) и открыл двери в атомный век.

Безусловно, Эйнштейн опирался на некоторые экспериментальные открытия XIX в., но и эти эксперименты были проведены с использованием достаточно простого оборудования. Также справедливо, что последующие экспериментальные доказательства теорий Эйнштейна потребовали использования более современных технологий, и если бы этого не было сделано, мы не имели бы доказательств справедливости и значимости его

идей. Однако все эти замечания не умаляют того факта, что эти знаменитые мысленные эксперименты раскрывают всю мощь человеческого разума.

Эйнштейна многие считают самым выдающимся ученым XX в. (а Дарвин был бы прекрасным кандидатом на это звание в XIX в.), хотя математический аппарат в основе его теорий достаточно прост. Более важную роль в создании его теорий сыграли сами мысленные эксперименты. Поэтому можно задать вопрос: в чем именно заключалась необычайная способность Эйнштейна? Позже мы обсудим, что происходило в его голове, когда он создавал свои теории, и где именно базируются эти замечательные способности.

Однако эта история также демонстрирует ограниченность человеческого разума. Эйнштейн смог прокатиться и удержаться на солнечном луче (хотя и пришел к выводу, что это невозможно), но ведь тысячи других наблюдателей и мыслителей оказались неспособны выполнить столь простые мыслительные упражнения. Одна из общечеловеческих проблем заключается в том, что большинству людей трудно отбросить идеи современников и переступить через границы этих идей. Но есть и другие недостатки, которые мы обсудим подробнее после того, как рассмотрим механизмы функционирования новой коры головного мозга.

### **Унифицированная модель новой коры**

Основная причина, заставившая меня в качестве вступления рассмотреть наиболее известные мысленные эксперименты в истории человечества, заключается в том, что я хочу использовать такой же подход в изучении головного мозга. Вы увидите, что мы сможем продвинуться очень далеко в анализе функционирования человеческого мозга с помощью собственных несложных мысленных экспериментов. Учитывая тему нашего исследования, мысленные эксперименты мне кажутся очень удачным методом анализа.

## Мысленные эксперименты над миром

Если праздных мыслей молодого человека, ручки и бумаги было достаточно для совершения революционного переворота в физике, мы должны суметь достаточно далеко продвинуться в привычном упражнении — мы ведь упражняем наш мозг постоянно, в каждый момент реальной жизни и во сне.

Затем мы создадим модель мыслительного процесса и посмотрим, до какой степени можем подтвердить ее справедливость, наблюдая за реальным мозгом и самыми совершенными машинами, воспроизводящими мыслительный процесс.



# Глава вторая

## МЫСЛЕННЫЕ ЭКСПЕРИМЕНТЫ НАД МЫШЛЕНИЕМ

Я вообще редко думаю словами. Мысль приходит, и потом я могу попытаться выразить ее в словах.

*Альберт Эйнштейн*

Мозг, который весит три фунта и который вы можете удержать в руке, способен постичь Вселенную размером в сотни миллиардов световых лет.

*Мариан Даймонд\**

Кажется невероятным, что объект весом менее трех фунтов, состоящий из тех же атомов, что и все остальные объекты в этом мире, отвечает практически за все, что делает человек: летает на Луну и отбивает до семидесяти хоум-ранов\*\*, пишет «Гамлета», и строит Тадж-Махал, и даже открывает секреты самого мозга.

*Жоэль Хейвманн\*\*\**

---

\* Мариан Даймонд — профессор анатомии из Калифорнийского университета в Беркли; участвовала в анализе головного мозга А. Эйнштейна.

\*\* Хоум-ран — удар в бейсболе.

\*\*\* Жоэль Хейвманн — журналистка, автор книги о болезни Паркинсона *A Life Shaken: My Encounter with Parkinson's Disease*.

**Я** начал задумываться о мышлении примерно в 1960-х гг., тогда же, когда открыл для себя компьютер. Сегодня трудно найти двенадцатилетнего ребенка, не использующего компьютер, но в то время в моем родном Нью-Йорке таких, как я, было совсем немного. Конечно, те первые экземпляры никак не могли поместиться на ладони, а тот, на котором мне позволили поработать, занимал целую большую комнату. В начале 1960-х гг. на компьютере IBM 1620 я выполнил несколько задач по анализу вариаций данных (статистический тест), собранных при подготовке образовательной программы для младших школьников. Работа над программой складывалась весьма драматично. Алгоритмы и анализируемые данные были достаточно сложны, так что мы не могли предугадать, какие ответы выдаст компьютер. Конечно, ответы определялись нашими данными, но предсказать результат было невозможно. Следует заметить, что *детерминированность* и *предсказуемость* — далеко не одно и то же, и мы еще поговорим об этом позднее.

Я помню тот волнующий момент, когда непосредственно перед окончанием расчетов лампочки на передней панели тускнели, как будто компьютер впадал в глубокую задумчивость. Когда приходили люди, ожидавшие новой серии результатов, я показывал пальцем на слабо мерцающие огоньки и говорил: «Он думает». Это было шуткой и правдой одновременно — *действительно* казалось, что компьютер ищет ответ, и сотрудники лаборатории в какой-то степени относились к машине как к личности. Наверное, это можно назвать антропоморфизмом, но именно тогда я начал серьезно задумываться о связи между мышлением и вычислением.

Чтобы понять, в какой степени мой собственный мозг похож на знакомую мне компьютерную программу, я начал думать о том, что должен делать мой мозг в процессе обработки информации. Я занимался этим вопросом на протяжении пятидесяти лет. Все, что я расскажу дальше о наших современных

представлениях относительно функционирования мозга, будет очень сильно отличаться от стандартной концепции компьютерной обработки информации. Однако на самом деле мозг действительно хранит и обрабатывает информацию, а ввиду универсальности вычислительного процесса между мозгом и компьютером гораздо больше общего, чем может показаться на первый взгляд.

Каждый раз, когда я что-то делаю или о чем-то думаю — чищу зубы, прохожу через кухню, решаю деловые проблемы, играю на музыкальной приставке или обдумываю новую задачу, — я пытаюсь понять, как мне это удастся. И еще больше я думаю о вещах, которые я не умею делать, поскольку определение пределов человеческой мысли обеспечивает не менее важный набор данных. Такое количество размышлений о мышлении вполне могло бы затормозить мою активность, однако, к счастью, эти эксперименты по наблюдению за самим собой позволяют мне усовершенствовать мыслительный процесс.

Чтобы лучше понять механизм работы нашего мозга, давайте сделаем несколько мысленных экспериментов.

*Попробуйте следующее: повторите алфавит.*

Думаю, вы помните его с детства и сделаете это без труда.

*Хорошо, теперь попробуйте повторить алфавит задом наперед.*

Если вы раньше не пытались выучить алфавит таким образом, вполне вероятно, что вы сочтете это задание невыполнимым. Возможно, кто-нибудь, кто провел особенно много времени в начальной школе, где на стенах классной комнаты вывешен алфавит, может напрячь визуальную память и прочесть алфавит задом наперед. Однако даже эта задача окажется непростой, поскольку мы обычно не помним всю картинку целиком. Кажется, прочесть алфавит в одну и в другую сторону одинаково просто, поскольку в нем содержится одинаковая информация, однако мы не можем этого сделать.

Помните номер своего страхового полиса? Если да, попробуйте назвать его задом наперед, не записывая предварительно на бумажке. А как насчет какого-нибудь коротенького детского стишка? Для компьютера это тривиальная задача, а вот нам с ней не справиться, разве что мы начнем заучивать этот обратный порядок букв заново. На этом основании можно сделать кое-какие важные выводы относительно организации человеческой памяти.

Конечно, мы легко справимся с этой задачей, если запишем последовательность, а затем прочтем ее в обратном направлении. В этом процессе мы используем некую технологию — письменность — для преодоления ограничений наших мыслительных способностей. Письменность — очень древнее изобретение человека, второе после разговорного языка. Вот почему мы придумываем орудия — чтобы восполнить собственные недостатки.

Из сказанного выше следует, что **наши воспоминания организованы в определенном порядке и могут воспроизводиться именно в том порядке, в котором запоминались. Мы не можем напрямую воспроизвести сохраненную в памяти информацию в обратном порядке.**

Нам так же трудно воспроизвести последовательность, начиная с середины. Если я заучиваю фортепианную пьесу, обычно я не могу начать с любого места. Есть несколько моментов, с которых я могу продолжать играть дальше, поскольку последовательность моих воспоминаний состоит из фрагментов. Однако если я пытаюсь начать с середины фрагмента, мне придется вернуться к нотам, чтобы включить память.

Далее попробуйте проделать следующий эксперимент: *вспомните свою прогулку, совершенную день или несколько дней назад. Что вы помните?*

Этот мысленный эксперимент лучше всего проводить, если вы ходили гулять совсем недавно — сегодня утром или вчера (это не обязательно должна быть прогулка, это может быть

поездка на машине или любая другая деятельность, в процессе которой вы перемещались в пространстве).

Вполне вероятно, что вы мало что можете вспомнить. Кто был пятым встреченным вами человеком (считая не только знакомых)? Видели ли вы по дороге дубы? Почтовый ящик? Что вы увидели, когда первый раз завернули за угол? Если вы проходили мимо магазинов, что вы видели во второй витрине? Возможно, вам удастся ответить на некоторые из этих вопросов на основании нескольких элементов, которые сохранились в вашей памяти, но скорее всего вы вспомните мало подробностей, хотя все это происходило совсем недавно.

Если вы ходите гулять регулярно, попытайтесь вспомнить первую прогулку (или дорогу на работу) в прошлом месяце. Вероятнее всего, вы не сумеете припомнить никаких особых подробностей, но даже если сможете, то их будет еще меньше, чем подробностей утренней прогулки.

Позднее я обращусь к теме сознания и расскажу о том, что мы склонны приравнивать сознание к событийной памяти. Почему мы считаем, что пребываем в бессознательном состоянии, находясь под наркозом? В первую очередь, по той причине, что мы ничего не помним об этом периоде времени (хотя из данного правила есть удивительные и непонятные исключения). Можно ли сделать вывод, что большую часть времени на прогулке мы находились в бессознательном состоянии? Это вполне законный вопрос, учитывая, что мы не помним практически ничего о том, что видели или о чем думали.

Я, к примеру, могу вспомнить некоторые подробности своей сегодняшней утренней прогулки. Я помню, что думал об этой книге, но не могу сказать, что именно. Помню, что встретил женщину с детской коляской. Женщина мне показалась симпатичной, да и ребенок был очень мил. Я помню две свои мысли по этому поводу: 1) «Малыш очарователен, как мой младший внук» и 2) «Как этот ребенок воспринимает то, что видит вокруг себя?» Я не могу вспомнить то, как эти люди были одеты, или

цвет их волос (моя жена скажет, что она не удивлена). Однако, хотя я не могу описать встреченную женщину, у меня почему-то есть ощущение, что я смог бы найти ее фотографию среди фотографий нескольких других женщин. Таким образом, даже если моя память сохранила что-то, касающееся этой женщины, я не могу визуально представить ни женщину, ни коляску, ни ребенка. В моей голове нет «фотографии» или «видео» с их изображением. Трудно сказать определенно, *что именно* сохранилось в памяти после этой встречи.

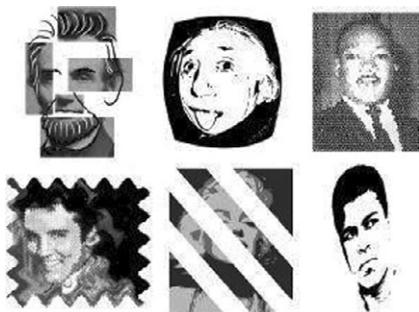
Могу сказать, что некоторое время назад во время прогулки я встречал и других женщин с детьми в колясках. Но в этом случае, мне кажется, я не смог бы даже узнать их на фотографии. Эти воспоминания теперь гораздо менее яркие, чем были непосредственно после встречи.

*Теперь попытайтесь вспомнить людей, с которыми встречались всего один или два раза в жизни. Вы отчетливо их себе представляете?*

Если вы художник или фотограф, возможно, ваша наблюдательность развита достаточно хорошо, но обычно люди не могут визуально представить себе или достаточно подробно описать тех, кого встречали лишь случайно, хотя и способны узнать их по фотографии.

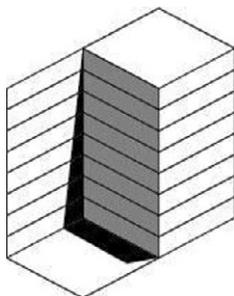
Все это говорит о том, что **в головном мозге не хранятся фото- или видеоизображения и звукозаписи. Наши воспоминания хранятся в виде последовательности образов. Невостребованные воспоминания постепенно стираются.** Когда в полиции составляют портрет подозреваемого в совершении преступления, у свидетелей не спрашивают, какие брови были у злоумышленника. Им показывают серию изображений и просят выбрать наиболее похожий вариант. «Правильный» набор изображений вызывает в памяти свидетеля соответствующую картинку.

Теперь взгляните на лица людей, которых все вы хорошо знаете. *Узнаёте кого-нибудь из них?*



Вы, безусловно, смогли их узнать, хотя фотографии частично скрыты или искажены. Это пример иллюстрирует важнейшую способность человеческого мозга: **мы можем распознавать рисунок, даже если воспринимаем (видим, слышим, чувствуем) лишь его часть и даже если он в некоторой степени изменен.** Наша способность к распознаванию образов, по-видимому, опирается на инвариантные фрагменты рисунка — характеристики, которые не изменяются в условиях реальных вариаций. Видимые искажения, свойственные карикатуре или некоторым видам живописи (например, импрессионизму), подчеркивают характеристики образа (человека или объекта), который мы узнаем даже при изменении других деталей. В настоящее время искусство опережает науку в определении способностей человеческого восприятия. Тот же механизм мы используем, когда узнаём мелодию всего по нескольким нотам.

*Теперь взгляните на это изображение:*



Изображение это воспринимается неоднозначно: черный треугольник может обозначать как углубление, так и выступ. Вначале вы воспринимаете его либо тем, либо иным образом, но при определенном усилии можете переключиться на альтернативный вариант. Однако, если ваш мозг настроился на одно объяснение, может оказаться трудным взглянуть на вещи с другой точки зрения (это относится и к интеллектуальным проблемам). Интерпретация изображения вашим мозгом влияет на ваше восприятие: если вы считаете треугольник выступом, мозг интерпретирует серый участок как тень, и он не кажется таким темным, как когда вы воспринимаете треугольник в качестве углубления.

Таким образом, **наш осознанный опыт восприятия изменяется в зависимости от интерпретации.**

*Подумайте, мы ведь видим то, что ожидаем...*

Я уверен, что вы смогли закончить предложение.

Если бы я дописал последнее слово, вам бы понадобилось лишь скользнуть по нему взглядом, чтобы подтвердить, что вы именно этого и ожидали.

Таким образом, **мы постоянно предсказываем будущее и составляем гипотезы о том, что произойдет. Это ожидание влияет на наше восприятие.** На самом деле главная причина, почему у нас вообще есть мозг, состоит в необходимости предсказывать будущее.

*Рассмотрим каждому из нас хорошо знакомый пример: время от времени по непонятной причине у нас в голове возникают воспоминания многолетней давности.*

Часто это воспоминание о каком-то человеке или событии, о котором вы не думали уже долгое время. Очевидно, что-то стало импульсом, вызвавшим это воспоминание. Иногда ход ваших мыслей можно объяснить. Иногда вы можете восстановить последовательность мыслей, вызвавших это воспоминание, но не способны ее выразить. Часто импульс быстро пропадает, и вам кажется, что мысль пришла «ниоткуда». Ко мне такие случайные

воспоминания часто приходят, когда я занят какими-то рутинными делами, например чищу зубы. Иногда мне удается найти связь: свалившаяся со щетки паста напоминает мне о краске, капавшей с кисточки на уроке рисования в школе. Иногда у меня бывает лишь смутное ощущение, что связь есть, а иногда я не нахожу никакого объяснения. Близкое по сути явление, которое также знакомо каждому, состоит в попытках вспомнить имя или слово. В таких случаях мы сами пытаемся подобрать те импульсы, которые могли бы вызвать в памяти необходимую информацию. Например: *Кто играл королеву Падме в «Звездных войнах»? Дайте подумать, кажется, недавно она же снималась в какой-то драме о балерине... А, это был «Черный лебедь». Ну конечно, Натали Портман!* Иногда мы прибегаем к специфической мнемонике. Например: *Она скорее худая, чем толстая, ну да, Портман, Натали Портман!*\* Некоторые из наших воспоминаний достаточно прочны, так что мы можем непосредственно от вопроса (например, *кто играл королеву Падме?*) перейти к ответу, но иногда приходится перебрать серию импульсов, пока мы не найдем тот, который сработает. Это в значительной степени напоминает поиск нужной ссылки в Интернете. Воспоминания на самом деле могут оказаться утерянными, как странички Интернета, к которым не ведут никакие ссылки (или мы не можем их найти).

*Когда выполняете какое-нибудь привычное действие (натягиваете рубашку) — проследите за собой и подумайте, в какой степени вы каждый раз следуете одной и той же схеме. Мои личные наблюдения (как я уже говорил, я постоянно наблюдаю за самим собой) свидетельствуют, что человек обычно выполняет привычные действия в одном и том же порядке, хотя иногда требуется подключение дополнительных модулей. Например, почти все мои рубашки без запонок, но, когда попадается рубашка с запонками, приходится выполнять дополнительную серию движений.*

---

\* Английское слово *portly* (тучный, толстый) созвучно фамилии Портман.

Стадии всех процессов записаны в моей голове в иерархическом порядке. Прежде чем идти спать, я выполняю стандартную серию процедур. Сначала чищу зубы. Однако этот процесс, в свою очередь, тоже состоит из нескольких этапов, первый из которых заключается в том, чтобы выдавить пасту на щетку. Этот этап тоже подразумевает несколько организованных действий: найти пасту, отвернуть колпачок и т. д. Стадия поиска пасты тоже состоит из более мелких стадий — в первую очередь нужно открыть дверь в ванную комнату. Эта вложенность действий продолжается до мельчайших деталей, так что наш вечерний туалет состоит буквально из сотен маленьких движений. И, хотя мне трудно вспомнить подробности прогулки, состоявшейся всего несколько часов назад, для меня не составляет никакой проблемы назвать все многочисленные этапы подготовки ко сну, которых так много, что, пока я их выполняю, я могу думать совершенно о других вещах. Важно отметить, что в нашей голове этот план не хранится в виде одного длинного списка из сотен движений, но **каждая рутинная процедура записана как сложная иерархическая цепь вложенных действий.**

**Тот же иерархический механизм задействован в нашей способности распознавать объекты и ситуации.** Мы узнаём лица знакомых людей, а также узнаём на этих лицах глаза, нос, рот и т. д. — это иерархия образов, которую мы применяем как для восприятия, так и для действий. Иерархия позволяет нам использовать одни и те же образы. Например, нам не нужно заново изучать, что такое нос или рот, каждый раз, когда мы видим новое лицо.

В следующей главе мы соберем результаты наших мысленных экспериментов в единую теорию, описывающую механизмы работы новой коры. Я покажу, что эти эксперименты отражают важнейшие и общие принципы мышления, одинаковые для поиска зубной пасты и сочинения стихов.

# Глава третья

## МОДЕЛЬ НОВОЙ КОРЫ; ТЕОРИЯ МЫСЛЕННОГО РАСПОЗНАВАНИЯ ОБРАЗОВ

Головной мозг — это ткань. Сложно сплетенная ткань, не похожая ни на что другое во Вселенной, но так же состоящая из клеток, как любая другая ткань. Это, безусловно, высокоспециализированные клетки, но функционируют они по тем же законам, что и любые другие клетки. Можно определить, зарегистрировать и интерпретировать их электрические и химические сигналы, идентифицировать химические молекулы и составить карту нервных сплетений. Короче говоря, мозг можно изучать так же, как изучают почку.

*Д. Х. Хьюбел\**

Предположим, что есть машина, производящая мысли, чувства и ощущения; вообразим, что эта машина увеличилась, но сохранила те же пропорции, так что вы можете войти в нее, как на мельницу. Вы можете осмотреть ее изнутри, но что вы там увидите? Только детали, которые толкают и двигают друг друга, и ничего другого, что могло бы объяснить восприятие.

*Готфрид Вильгельм Лейбниц, 1714*

---

\* Дэвид Хантер Хьюбел (1926–2013) — канадско-американский нейрофизиолог, лауреат Нобелевской премии по физиологии и медицине за 1981 г. «За открытия, касающиеся принципов переработки информации в нейронных структурах».

## Иерархия образов

Я повторял описанные выше простые эксперименты и наблюдения тысячи раз в самых разных ситуациях. Выводы из этих наблюдений неизбежно ограничивают мое представление о том, как должен работать мозг — точно так же, как простые эксперименты XIX в. по анализу времени, пространства и массы сдерживали размышления молодого Эйнштейна об устройстве Вселенной. Далее я остановлюсь на некоторых наблюдениях из нейробиологии, стараясь избегать многих пока еще неизвестных подробностей.

Прежде всего, позвольте мне объяснить, почему в этом разделе я считаю нужным поговорить о неокортексе (что в переводе с латыни буквально означает «новая кора»). Мы знаем, что именно новая кора отвечает за нашу способность обрабатывать информацию и что функционирует она по иерархическому принципу. Животные, не имеющие новой коры (в основном не млекопитающие), в большинстве своем не способны понимать иерархические построения<sup>1</sup>. Понимание и использование иерархической природы реальности являются исключительной способностью млекопитающих и объясняются наличием у них этой новой в эволюционном плане структуры головного мозга. Новая кора отвечает за чувственное восприятие, распознавание видимых объектов и абстрактных понятий, контроль движений, размышления (от пространственной ориентации до рационального мышления) и речь, то есть за все то, что мы называем «разумом».

Новая кора человеческого мозга, его самый внешний слой, представляет собой тонкую, практически двумерную структуру (ее толщина составляет около 2,5 мм). У грызунов она гладкая и имеет толщину почтовой марки. Эволюционной инновацией у приматов стала сложная складчатость этой структуры, образующей на поверхности остального мозга глубокие морщины, складки и бороздки, увеличивающие площадь ее поверхности.

Благодаря этой сложной складчатости новая кора составляет основную часть человеческого мозга — около 80% по массе. Высокий лоб *Homo sapiens* позволяет дополнительно увеличить размер новой коры. В частности, лобная доля мозга отвечает за обработку абстрактной информации и формирование сложных понятий.

Эта тонкая структура мозга состоит из шести основных слоев — от слоя I (внешний) до слоя VI. Аксоны (выходные контакты), отходящие от нервных клеток на уровне слоев II и III, проецируются в другие участки новой коры. Аксоны, отходящие от слоев V и VI, вне новой коры связываются с таламусом, стволом мозга и спинным мозгом. Нейроны слоя IV принимают синаптические (входные) сигналы от нейронов, находящихся вне новой коры, особенно в таламусе. В разных отделах мозга число слоев коры может незначительно изменяться. В моторной коре слой IV очень тонкий, поскольку в этой зоне он практически не принимает сигналов от таламуса, ствола мозга или спинного мозга. Напротив, в затылочной доле (зоне новой коры, ответственной за зрение) слой IV имеет три дополнительных подслоя, поскольку в эту область приходит множество сигналов, в том числе от таламуса.

Важнейшая особенность новой коры — удивительная однородность ее основных структур. Впервые это заметил американский нейробиолог Вернон Маунткасл (1918–2015). В 1957 г. он открыл колончатую структуру новой коры. В 1978 г. он обнаружил закономерность, которая была так же важна для нейробиологии, как эксперименты Майкельсона и Морли в 1831 г. для физики. Он описал удивительно однородную организацию новой коры и высказал гипотезу, в соответствии с которой кора построена по единому принципу, воспроизводимому множеством раз<sup>2</sup>, а в качестве основного модуля этой структуры Маунткасл назвал кортикальную колонку. Изменение высоты некоторых слоев в различных отделах мозга объясняется лишь разным числом связей этих отделов с другими структурами.

Маунткасл предположил, что колонки состоят из мини-колонок, но эта гипотеза не получила поддержки, поскольку не было найдено никаких видимых границ подобных структур. Однако активные экспериментальные исследования показали, что, действительно, в нейронной ткани каждой колонки имеются повторяющиеся единицы. Я считаю, что основной структурной единицей новой коры являются распознающие модули. В отличие от мини-колонок Маунткасла, эти модули не разделены никакими физическими границами, поскольку расположены очень близко друг к другу и связаны между собой, так что кортикальная колонка — это просто агрегат большого числа таких модулей. Со временем распознающие модули могут скручиваться друг с другом, так что сложное соединение модулей, которое мы наблюдаем в новой коре, не определяется генетическим кодом, а формируется постепенно в зависимости от тех образов, которые нам приходится распознавать. Я подробнее остановлюсь на этом позднее; важно, чтобы вы поняли, как организована новая кора.

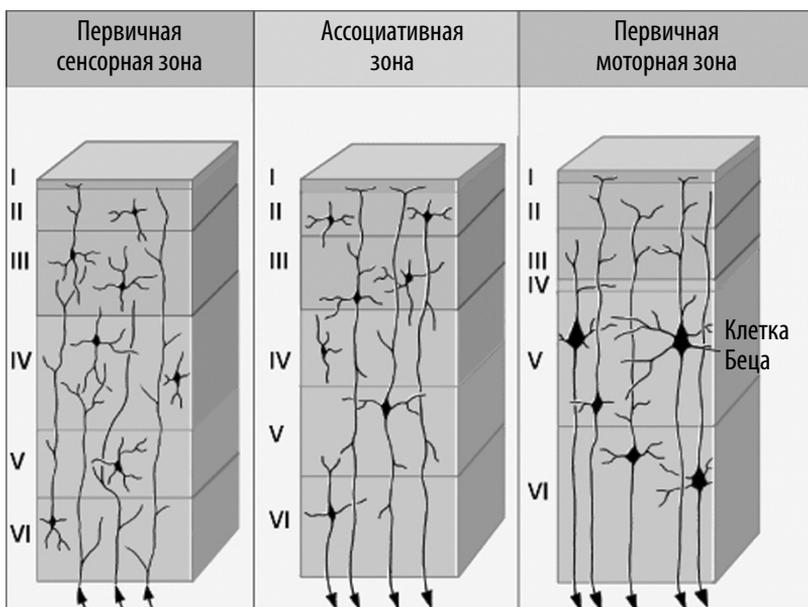
Прежде чем мы продолжим исследовать структуру новой коры, хочу заметить, что чрезвычайно важно осуществлять моделирование поведения сложных систем на правильно выбранном уровне. Хотя химия теоретически основана на физике и может быть выведена из законов физики, на практике сделать это невозможно, поэтому химия создала собственные законы и модели. Аналогичным образом из законов физики можно вывести законы термодинамики, но, если у нас имеется достаточное количество частиц, чтобы мы могли назвать их скопление газом, решение физических уравнений для взаимодействий всех частиц становится безнадежным делом, а вот законы термодинамики отлично работают. Точно так же биология создала свои собственные законы и модели. Единственная клетка поджелудочной железы чрезвычайно сложна, особенно если мы рассматриваем ее поведение на уровне молекул; значительно проще осуществить моделирование функций клеток поджелудочной

железы в регуляции уровня инсулина и пищеварительных ферментов.

Тот же принцип применим к моделированию и изучению функций мозга. Безусловно, важной и необходимой задачей для проектирования мозга является создание моделей взаимодействий на межмолекулярном уровне, однако цель всего процесса главным образом заключается в уточнении модели, которая описывает, каким образом мозг перерабатывает информацию, превращая ее в знания.

Американский ученый Герберт А. Саймон (1916–2001), которого считают одним из отцов теории искусственного интеллекта, очень красочно выразился по поводу того, насколько важно выбрать правильный уровень абстракции для анализа сложных систем. В 1973 г. по поводу созданной им программы искусственного интеллекта ЕРАМ (elementary perceiver and memorizer) Саймон писал следующее: «Представьте себе, что вы хотите понять имеющуюся у меня загадочную программу ЕРАМ. Я могу дать вам две версии этой программы. Первая версия — это та, в которой программа была исходно написана, — со всеми составляющими ее компонентами и подпрограммами... Вторая — полностью трансформированная версия ЕРАМ, переведенная на машинный язык... Я думаю, не стоит долго объяснять, какая версия представляет более краткое, значимое и понятное описание... Третью я вам предлагать не стану, поскольку это окажется даже не программа, а электромагнитные уравнения и граничные условия, которым компьютер как физическая система должен подчиняться. И это будет кульминация сокращения и непостижимости»<sup>3</sup>.

В новой коре человека содержится около полумиллиона кортикальных колонок, каждая из которых составляет около 2 мм в высоту и 0,5 мм в ширину и содержит около 60 тыс. нейронов (таким образом, в сумме в новой коре содержится около 30 млрд нейронов). В общем, каждый распознающий модуль кортикальной колонки содержит около 100 нейронов, а всего в новой коре имеется порядка 300 млн распознающих модулей.



Теперь мы обратимся к рассмотрению механизмов работы распознающих модулей, однако должен заметить, что, честно говоря, совершенно непонятно, с чего же следует начинать. В новой коре все происходит одновременно, так что в этих процессах нет ни начала, ни конца. Мне часто придется упоминать явления, которые я еще не объяснил, а только собираюсь объяснить, так что прошу вас запастись терпением.

Человек лишь в небольшой степени способен к логическим размышлениям, зато прекрасно умеет распознавать образы. Для логического мышления нам приходится использовать новую кору, которая и является большим распознающим модулем. Это далеко не идеальный механизм для осуществления логических преобразований, но другого у нас нет. Сравним, к примеру, как играют в шахматы человек и специальная компьютерная программа. В 1997 г. компьютер DeepBlue, обыгравший чемпиона мира среди людей Гарри Каспарова, был способен за секунду логически анализировать последствия 200 млн комбинаций на

доске (то есть последовательностей ходов и ответных ходов). Сейчас на такое способны некоторые персональные компьютеры. Каспаров в ответ на вопрос, сколько комбинаций он может анализировать за секунду, ответил, что менее одной. Как же в таком случае он вообще мог сражаться против DeepBlue? Ответ заключается в исключительной способности человека распознавать образы. Однако эту способность нужно тренировать — вот почему не каждый из нас мастерски играет в шахматы.

Каспаров знал примерно 100 тыс. шахматных комбинаций. Это реальное число — по нашим расчетам, специалист в какой-либо сфере деятельности должен уметь оперировать примерно 100 тыс. «элементами» знаний. В пьесах Шекспира использовано около 100 тыс. предложений (в которых задействовано около 29 тыс. слов, но большинство из них использовано разными способами). Анализ медицинской экспертной системы, созданной для воспроизведения знаний обычного врача, показал, что врач оперирует примерно 100 тыс. понятий в своей области. Распознавание «элемента» знаний из такого хранилища информации является далеко не простой задачей, поскольку все элементы слегка изменяются при каждом новом обращении к ним.

Вооруженный этими знаниями, Каспаров смотрит на шахматную доску и сравнивает каждую конкретную комбинацию со всеми знакомыми ему 100 тыс. комбинаций, причем все 100 тыс. сравнений он производит одновременно. В этом вопросе наблюдается полный консенсус: все наши нейроны одновременно участвуют в обработке изображения. Это не означает, что все они одновременно *возбуждаются* (в такой ситуации мы, возможно, не могли бы удержаться на ногах), но выполнение их функции подразумевает возможность возбуждения.

Сколько образов может хранить новая кора? Здесь необходимо учесть явление избыточности. Например, лицо любимого человека хранится не в виде единого образа, а записано тысячи раз. Некоторые из этих повторов представляют собой изображения практически одного и того же лица, тогда как другие

показывают его в разных ракурсах, при разном освещении, с разным выражением и т. д. Никакие из этих повторяющихся образов не хранятся в виде истинных рисунков (то есть в виде двумерного набора пикселей). Скорее, они хранятся в виде списков признаков, в которых составные элементы образа сами являются образами. Чуть позже мы подробнее поговорим о том, как устроена эта иерархия признаков.

Если принять, что база знаний специалиста в какой-либо области состоит из 100 тыс. «единиц» знания (то есть образов) при избыточности примерно 100:1, получается, что мозг хранит около 10 млн образов. Эти специфические знания основаны на более общих знаниях и дополняются более глубокими и узкоспециальными знаниями, так что общее число образов повышается до 30 или 50 млн. Однако наши «бытовые» знания, которые мы используем в каждодневной жизни, еще шире; знание «законов улицы» требует от нашей новой коры значительно больше, чем «книжное знание». Если учесть эти бытовые знания и упомянутый выше фактор избыточности, мы получим, что общее число образов, хранящихся в нашей новой коре, превышает 100 млн. Отметим, что фактор избыточности не постоянная величина — часто используемые образы могут повторяться тысячи раз, а новые явления имеют фактор избыточности менее десяти.

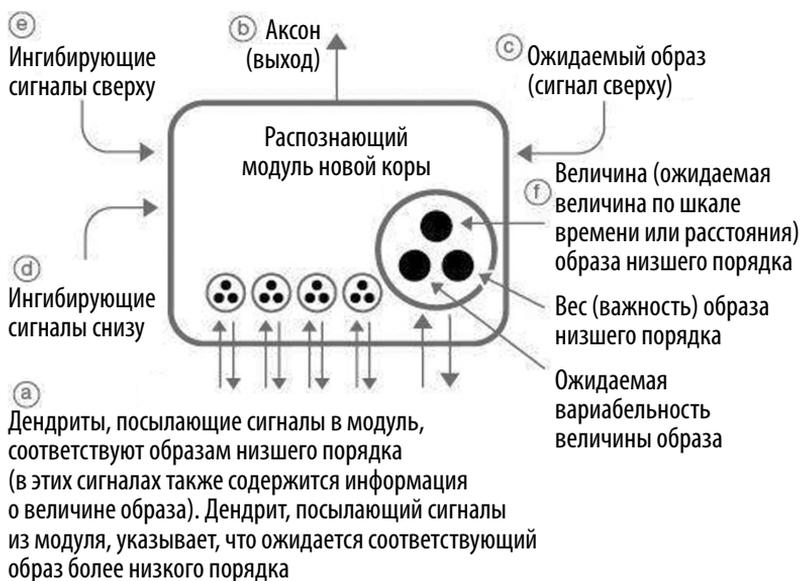
Как я расскажу позднее, наши действия и навыки также составляют образы и тоже хранятся в отделах коры, так что, по моим оценкам, общий объем информации в человеческой новой коре составляет несколько сотен миллионов образов. Эта приблизительная оценка вполне соответствует числу распознающих элементов коры, которое, как мы уже определили, составляет около 300 млн. Поэтому вполне резонно предположить, что функция каждого распознающего элемента новой коры состоит в обработке одной итерации образа (то есть одной копии среди множества повторяющихся копий образов). Наши оценки числа образов, которые может обрабатывать человеческий мозг (с учетом необходимой избыточности), и числа физических

распознающих элементов привели нас к величинам одного порядка. Замечу, что, когда я говорю об «обработке» образа, я имею в виду все процессы, которые мы можем с ним (и его частями) проделать: изучить, предсказать, узнать и использовать (либо путем дальнейшего осмысления, либо путем применения образа физического перемещения).

Процессор, обрабатывающий 300 млн образов, кажется весьма серьезным аппаратом; и действительно, он позволил *Homo sapiens* создать устную и письменную речь, все наши многочисленные инструменты и творения. Одни изобретения вызывали к жизни другие, что привело к экспоненциальному росту информационного содержания технологий, который я описываю с помощью закона ускорения отдачи. Никакие другие существа не способны на такое. Как я уже говорил, ряд животных, включая шимпанзе, по-видимому, обладают рудиментарными способностями понимать и произносить слова, а также использовать примитивные орудия. У них, вообще говоря, тоже есть новая кора, однако их способности ограничиваются небольшим размером коры, особенно лобных долей. Размер нашей новой коры позволил нам создавать еще более мощные инструменты, включая те, которые теперь помогают нам изучать собственный разум. Когда-нибудь наш мозг вместе с созданными нами технологиями позволит сотворить искусственную новую кору, в которой будет не 300 млн процессоров, а значительно больше. Может быть, миллиард? Или триллион?

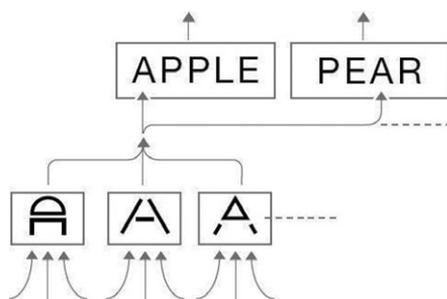
### **Структура образа**

Теория мысленного распознавания образов, которую я хочу вам представить, основана на распознавании образов распознающими модулями новой коры. Эти образы (и модули) организованы иерархическим путем. Ниже я расскажу о том, как возникла эта идея, в том числе поведаю о собственных экспериментах 1980–1990-х гг. и о модели коры Джефа Хокинса и Дайлипа Джорджа, созданной в начале 2000-х гг.



Каждый образ (распознаваемый одним из примерно 300 млн распознающих модулей новой коры) состоит из трех элементов. Первый — это входной сигнал из образов низшего порядка, формирующий основной образ. Нет необходимости в повторении описания каждого из этих образов низшего порядка для каждого образа более высокого порядка, в котором они содержатся. Например, многие образы слов содержат букву «А». Нет нужды повторять описание буквы «А» во всех этих образах слов — они используют одно и то же описание. Это можно сравнить со страницкой в Интернете: в Интернете есть одна страница с описанием буквы «А» (один образ), и все страницы для слов (образы слов), содержащих букву «А», связаны с этой страницей (с образом буквы «А»). Вместо ссылок новая кора использует реальные нервные связи. Аксон от модуля, распознающего букву «А», связан с множеством дендритов — по одному для каждого слова с буквой «А». Но не забывайте и об избыточности коры: для распознавания буквы «А» существует несколько модулей. Любой из них может посылать сигнал модулям, распознающим слова с буквой «А».

Второй элемент каждого образа — его имя. Если речь идет о звуковых образах, этот образ более высокого порядка — просто слово, например «яблоко» (*apple*). Хотя мы напрямую применяем новую кору для понимания и использования речи, большинство образов в коре не являются языковыми образами. «Имя» образа в новой коре — это просто аксон, выходящий из каждого модуля; возбуждение аксона означает распознавание соответствующего образа. Возбуждение аксона означает, что распознающий модуль «называет имя» образа: «Эй, ребята, я увидел слово “яблоко”!»



*Три повторяющихся (но слегка отличных друг от друга) образа «А» включаются в образы более высокого порядка, содержащие букву «А».*

Третий и последний элемент образа — это набор образов более высокого порядка, в состав которых он входит. Для буквы «А» это все слова с буквой «А». Здесь вновь уместно сравнение с веб-страницей. Каждый распознанный образ более низкого порядка запускает распознавание образа более высокого порядка, содержащего первый образ. В новой коре эти связи осуществляются дендритами, соединяющимися с нейронами в каждом распознающем модуле коры. Помним, что каждый нейрон может получать сигналы от множества дендритов, но выдает единственный сигнал на аксон. Этот аксон, однако, может, в свою очередь, передавать сигнал множеству дендритов.

В качестве примера рассмотрим набор знаков, использующихся для изображения печатных букв. Здесь каждый уровень — это образ. Формы — это образы, буквы — образы, слова — тоже образы.

## Глава третья

Каждый образ характеризуется серией входных сигналов, процессом распознавания образа в модуле и выходным сигналом, поступающим к распознающему модулю более высокого порядка.

Из нижней левой точки к центру  
верхней линии:



Из нижней правой точки к центру  
верхней линии:



Горизонтальная перекладина:



Левая вертикаль:



Дуга в нижней области:



Нижняя горизонталь:



Верхняя горизонталь:



Средняя горизонталь:

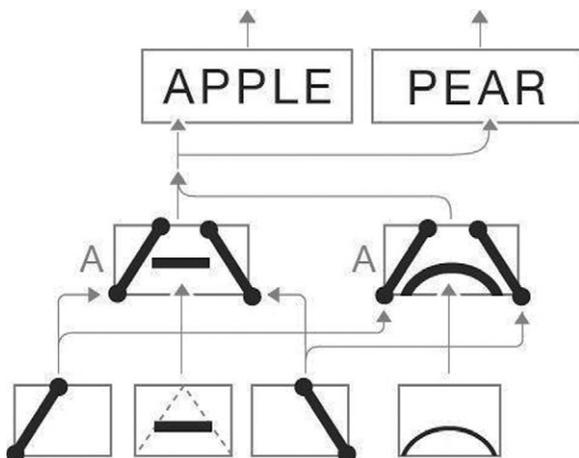


Петля в верхней левой области:



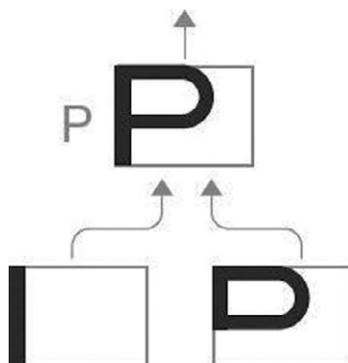
Перечисленные образы составляют образ более высокого порядка, относящийся к категории печатных букв (в новой коре, конечно, таких формальных категорий не существует).

Буква А:



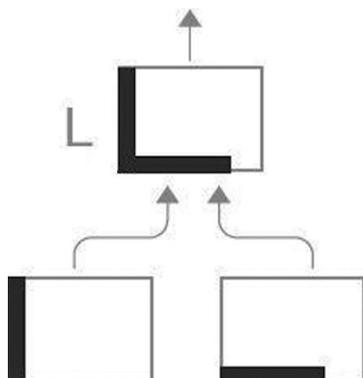
*Два разных образа, составляющих букву «А» и два разных образа более высокого порядка (APPLE и PEAR), в состав которых входит «А».*

Буква Р:



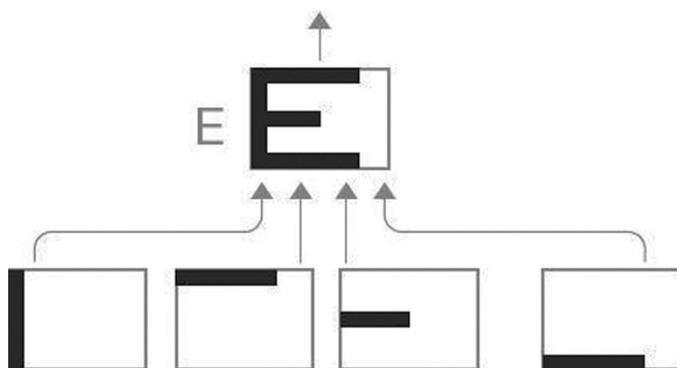
Образы, из которых складывается образ более высокого порядка — буква «Р».

Буква L:



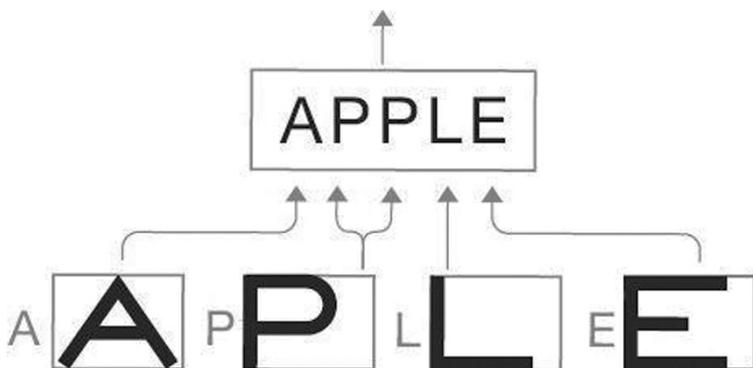
Образы, из которых складывается образ более высокого порядка — буква «L».

Буква E:



Образы, из которых складывается образ более высокого порядка — буква «E».

Эти образы букв соединяются в образ еще более высокого порядка, относящийся к категории слов (в нашем с вами языке — для новой коры это лишь образ определенного порядка): APPLE.



В другой части коры имеет место аналогичная иерархия распознающих модулей, участвующих в обработке образов реальных объектов (а не печатных букв). Если вы смотрите на настоящее яблоко, модули низшего уровня находят округлые формы и образы цвета кожицы, в результате чего происходит возбуждение соответствующего аксона и подается знак: «Эй, ребята, я увидел настоящее яблоко!» Возбуждение аксонов распознающих модулей в слуховой коре, участвующих в определении частоты звуковых сигналов, в ответ на произнесенное кем-то слово «яблоко» даст знак: «Только что прозвучало слово “яблоко”!»

Опять же, помним об избыточности модулей — мы имеем не по одному распознающему модулю для каждого вида яблока (увиденного или услышанного слова «яблоко» и реального яблока). Скорее всего, происходит возбуждение сотен таких модулей, если не больше. Избыточность не только повышает вероятность успешного узнавания всех форм яблока, но и помогает распознавать варианты настоящих яблок. Существуют распознающие модули для узнавания самых разных видов яблок — всех сортов, цветов и форм.

Кроме того, не забываем о том, что описанная выше иерархия является иерархией *понятий*. Распознающие элементы на самом деле не выстраиваются каким-либо иерархическим образом относительно друг друга; новая кора тонкая и по высоте

равна лишь одному распознающему элементу. Концептуальная иерархия создается за счет взаимодействий между отдельными распознающими элементами.

Важным элементом теории мысленного распознавания образов является описание процесса распознавания в каждом распознающем модуле. В модель заложен параметр «веса» входного сигнала каждого дендрита; этот параметр определяет важность данного сигнала для распознавания. Модули характеризуются пороговым значением возбуждения (преодоление этого значения говорит о том, что сигнал успешно узнается соответствующим распознающим модулем). Для возбуждения распознающего модуля не обязательно нужны все входные сигналы. Модуль может возбуждаться, например, при отсутствии входного сигнала с малым весом, но при отсутствии важного сигнала возбуждение вряд ли возможно. Возбуждение распознающего модуля обычно означает следующее: «Образ, за узнавание которого я отвечаю, скорее всего, присутствует».

Однако успешное распознавание модулем соответствующего образа заключается не только в подсчете входных сигналов (и в учете параметра их значимости). Важна также величина сигнала. Каждый входной сигнал, кроме того, описывается параметром, указывающим ожидаемую величину сигнала, и параметром, учитывающим вариабельность этой величины. В качестве примера рассмотрим распознающий модуль, ответственный за узнавание в речи слова *steep* (произносится «стиип» — «крутой», «высокий»). Слово состоит из четырех звуков: [s], [t] [E] и [p]. Звук [t] относится к так называемым зубным согласным; это означает, что звук производится воздухом, нарушающим контакт между языком и верхними зубами. Звук [t] практически невозможно произносить медленно. Глухой звук [p] относится к «взрывным согласным»; он образуется при открытии заблокированного голосового тракта (перекрытого губами в случае [p]). Это тоже быстрый звук. Гласный звук [E] образуется за счет резонанса голосовых связок при открытом рте. Он относится

к долгим гласным звукам, то есть длится гораздо дольше, чем согласные [t] и [p], однако его длительность может изменяться в широких пределах. Звук [s] относится к свистящим согласным; он возникает при прохождении воздуха через щель между сжатыми зубами. Его длительность обычно меньше, чем у долгих гласных звуков, таких как [E], но тоже может изменяться (звук [s] можно произнести быстро, а можно протянуть).

В нашей работе по распознаванию речи мы установили, что для распознавания звуковых образов этот тип информации должен быть закодирован. Например, слова *steep* и *step* (произносится «стэп» — «шаг», «этап») достаточно похожи. Хотя звуки [e] в слове *step* и [E] в слове *steep* звучат немного по-разному (имеют разные резонансные частоты), нельзя достоверно различить эти два слова лишь на этом основании. Гораздо надежнее основываться на длительности звучания гласных в этих двух словах: [e] в слове *step* звучит короче, чем [E] в слове *steep*.

Эту информацию можно закодировать с помощью двух параметров — ожидаемой величины (в данном случае длительности) и степени ее варибельности. В нашем примере звуки [t] и [p] в слове *steep* характеризуются очень малой ожидаемой длительностью и малой варибельностью (это означает, что мы не ожидаем услышать долгих звуков [t] и [p]). Звук [s] характеризуется малой ожидаемой длительностью, но большей варибельностью, поскольку этот звук можно протянуть. Наконец, звук [E] имеет большую ожидаемую длительность и высокую степень варибельности.

В этом примере величина — это длительность, но длительность — лишь одна из нескольких возможных характеристик величины сигнала. В нашей работе по распознаванию знаков мы обнаружили, что для распознавания печатных букв важна соответствующая пространственная информация (например, ожидается, что точка над *i* будет значительно меньше палочки). На более высоком понятийном уровне новая кора имеет дело с самыми разными совокупностями информации, такими как

уровень притягательности, иронии, удовольствия, расстрой-ства, и множеством других. Мы можем найти сходство между еще более различающимися совокупностями, чем Дарвин, кото-рый связал размер геологических разломов с различиями между видами организмов.

Источником данных параметров для головного мозга явля-ется собственный опыт мозга. Когда мы родились, мы ничего не знали о фонемах (звуковых единицах языка), кроме того, раз-ные языки очень сильно различаются по фонетическим харак-теристикам. Это означает, что многочисленные примеры образа записываются в виде параметров для каждого распознающего модуля (поскольку ожидаемое распределение величин входных сигналов определяется в результате множества экспозиций). В некоторых программах искусственного интеллекта такие па-раметры кодируются экспертами (например, лингвисты могут назвать ожидаемую длительность различных фонем). В ходе на-ших исследований мы поняли, что лучше заставить программу самостоятельно определить параметры на основе тренировоч-ных данных (примерно так, как это делает мозг). Иногда мы ис-пользовали смешанный подход, то есть снабжали систему че-ловеческой интуицией (для начальных установок параметров), а затем заставляли ее уточнить эти оценки путем обучения на реальных речевых примерах.

Что же делает распознающий модуль? Он вычисляет вероят-ность (основанную на предыдущем опыте) того, что тот образ, за распознавание которого он отвечает, действительно представ-лен активными входными сигналами. Каждый поступающий на модуль сигнал активен в том случае, если возбужден соответ-ствующий распознающий модуль более низкого порядка (это означает, что произошло распознавание образа более низкого порядка). Каждый входной сигнал также кодирует наблюдаемую величину сигнала (в подходящем измерении — в единицах вре-мени, каких-то физических величин или иных параметров), так что эти величины сравниваются модулем с соответствующими

величинами предыдущих сигналов для расчета вероятности того, что это «правильный» образ.

Как мозг (и система искусственного интеллекта) рассчитывает общую вероятность присутствия образа (за распознавание которого отвечает соответствующий модуль) на основании: 1) входных сигналов (определенной величины), 2) предыдущих параметров величины (ожидаемая величина и ее вариабельность) каждого сигнала и 3) значимости каждого сигнала? Для определения этих параметров и их использования для изучения иерархии образов в 1980-х и 1990-х гг. я и некоторые другие ученые предложили математический метод, называемый методом скрытых моделей Маркова. Мы применили этот подход для распознавания и понимания человеческой речи. Я опишу его в седьмой главе.

В примере с распознаванием слов, идущем от образов низшего порядка к образам высшего порядка, мы видим, как строится иерархия информационного потока от отдельных элементов букв к целым буквам, а затем к словам. Далее процесс распознавания поднимается до уровня фраз и более сложных речевых конструкций. Если мы поднимемся выше еще на несколько десятков уровней, мы дойдем до таких концептуальных образов, как ирония или зависть. Хотя все распознающие модули функционируют одновременно, продвижение по этой иерархической лестнице должно занимать определенное время. Переход с одного уровня на другой занимает от сотых до десятых долей секунды. Эксперименты показывают, что узнавание такого образа средней сложности, каким является человеческое лицо, происходит за десятые доли секунды. Если образ сильно изменен, процесс распознавания может длиться целую секунду. Если бы мозг функционировал последовательно (как обычный компьютер) и распознавал все образы в иерархическом порядке, он должен был бы обрабатывать каждый образ низшего уровня и лишь затем переходить к образам следующих уровней. В таком случае он должен был бы совершать миллионы циклов

распознавания при переходе от одного уровня к другому. Именно это происходит, когда мы моделируем данный процесс на компьютере. Не будем забывать, однако, что компьютеры функционируют в миллионы раз быстрее биологических систем.

Очень важно обратить внимание на то, что поток информации не только поднимается по иерархической лестнице, но и опускается по ней. Этот нисходящий поток информации чрезвычайно важен. Например, если мы читаем слева направо и уже увидели и узнали буквы А, Р, Р и L, модуль, ответственный за распознавание слова apple, предскажет, что, скорее всего, в следующей позиции окажется буква Е. Он отправит сигнал *вниз* модулю, ответственному за распознавание буквы Е, и сообщит: «будь внимателен, вероятно, очень скоро возникнет образ буквы Е». Модуль, распознающий букву Е, скорректирует пороговые параметры в соответствии с высокой вероятностью появления буквы Е. И если на месте следующей буквы возникает образ, напоминающий Е, но искаженный таким образом, что «в нормальных условиях» он не был бы воспринят как Е, теперь распознающий модуль узнает в нем Е, поскольку таковы были его ожидания.

Таким образом, новая кора предсказывает события. Необходимость предсказания будущего — одна из главных причин появления у нас новой коры. На высшем понятийном уровне мы постоянно осуществляем предсказания — кто сейчас войдет в дверь, что именно скажет этот человек в следующий момент, что мы увидим за углом, какими будут результаты наших действий и т. д. Эти предсказания непрерывно осуществляются в новой коре *на всех уровнях* иерархии. Мы часто путаем людей, предметы и слова из-за слишком низкого порогового значения для подтверждения ожидаемого образа.

Кроме положительных сигналов, существуют также отрицательные (ингибирующие) сигналы, указывающие на сниженную вероятность появления того или иного образа. Эти сигналы могут поступать с более низкого уровня иерархии (например, если я вижу у человека в очереди усы, понижается вероятность

того, что это моя жена) или с более высокого уровня (например, я знаю, что моя жена отправилась в путешествие, поэтому человек в очереди — не она). Когда распознающий модуль получает ингибирующий сигнал, он повышает порог распознавания, но возбуждение все еще возможно (так что, если человек в очереди — действительно моя жена, я все-таки смогу ее узнать).

### **Природа данных, поступающих в распознающие модули новой коры**

Давайте более подробно поговорим о том, как кодируются образы. Если образ — это лицо человека, данные о нем представляются как минимум в двух измерениях. Но мы не можем сказать, что первой подается информация, скажем, о глазах, потом о носе и т. д. То же самое относится и к восприятию звука. Музыкальная информация также имеет как минимум два измерения. Пьесу может исполнять не один, а несколько инструментов или голосов. Кроме того, одна нота, исполненная на сложном инструменте, таком как фортепьяно, содержит несколько частот. Голос человека одновременно раскладывается на множество составляющих с разной энергией и частотой. Так что звуковой образ в каждый конкретный момент времени может быть сложным, плюс растягиваться во времени. Тактильные сигналы тоже двумерные, поскольку кожа — двумерный чувствительный орган, и, опять-таки, возможно изменение этих сигналов в третьем, временном, измерении.

Таким образом, кажется очевидным, что образы передаются на распознающие модули новой коры в виде двумерных или трехмерных сигналов. Однако структура новой коры такова, что входные сигналы не могут иметь больше одного измерения. Наши исследования в области создания искусственных систем распознавания образов (речевых и зрительных) показывают, что мы можем представлять (и действительно представляем) двумерные и трехмерные явления в виде одномерных списков. О том, как это происходит, я расскажу в седьмой главе, а теперь

мы просто должны принять, что входной сигнал на каждый распознающий модуль представляет собой одномерный список, даже если сам образ не является одномерным.

Здесь я должен обратить ваше внимание на тот факт, что образы, которые мы научились распознавать (например, образ конкретной собаки или «собаки» вообще, музыкальную ноту или музыкальную пьесу), организованы в соответствии с тем же самым механизмом, что и наши воспоминания. Наши воспоминания на самом деле представляют собой организованные в виде списков образы (и каждый пункт в каждом списке соответствует отдельному образу в иерархии новой коры), которые мы выучили и распознали под действием соответствующего стимула. Воспоминания хранятся в новой коре, чтобы быть узнаваемыми.

Единственное исключение из этого правила относится к низшему понятийному уровню, на котором входные сигналы образа несут специфическую сенсорную информацию (например, изображение, поступающее от зрительного нерва). Однако даже этот образ низшего уровня в значительной мере трансформируется в более простые образы к моменту его поступления в кору головного мозга. Списки образов, составляющих воспоминание, организованы в хронологическом порядке, и мы можем «вспомнить наши воспоминания» только в этом порядке, вот почему нам бывает трудно обратить нашу память вспять.

Воспоминание возникает в ответ на какую-либо мысль или другое воспоминание (что одно и то же). Действие этого механизма запуска воспоминания можно пронаблюдать на примере восприятия образа. Если мы различили буквы А, Р, Р и L, образ слова APPLE предсказывает, что мы сейчас увидим букву Е, и запускает образ Е, который теперь является ожидаемым. Таким образом, наша новая кора «думает», что видит букву Е, еще до того, как мы ее увидели на самом деле. Если такое конкретное взаимодействие в коре привлекло наше внимание, мы будем думать о букве Е до того, как увидим ее, и даже если ее не увидим. Аналогичный механизм запускает воспоминания.

Обычно существует целая цепочка подобных связей. Даже если нам кажется, что мы понимаем, какие воспоминания (то есть образы) вызвали в памяти старые воспоминания, следует понять, что воспоминания (образы) не имеют языковых или визуальных ярлыков. Вот почему иногда кажется, что старые воспоминания возникают внезапно. Возможно, они хранились, не активируясь, годами и были запущены по тому же механизму, как страничка в Интернете активируется по ссылке. И, подобно страничке в Интернете, которая может остаться «сиротой», если больше нет других страниц, дающих на нее ссылку, то же может происходить и с нашими воспоминаниями.

Наши мысли активируются направленным или ненаправленным образом в соответствии с описанными выше кортикальными связями. При ненаправленной активации эти связи действуют самостоятельно, то есть мы не пытаемся направить их в какую-то определенную сторону. Например, по такому принципу осуществляются некоторые виды медитации (в частности, практикуемая мной трансцендентальная медитация). Так же устроены сны.

При направленном способе мышления мы пытаемся пойти более упорядоченным путем, вызывая воспоминание (например, какую-то историю) или решая проблему. Однако и в одном, и в другом случае нам приходится пробираться сквозь списки образов, хранящихся в нашей новой коре. Таким образом, общий ход нашей мысли — весьма разупорядоченный процесс, ярко отраженный Джеймсом Джойсом в его манере «потока сознания».

Когда мы мыслим, используя наши воспоминания (истории или образы) — касаются ли они встреченной на прогулке молодой мамы с ребенком в коляске или первого знакомства с собственной женой, — эти воспоминания состоят из последовательностей образов. Поскольку образы не помечены словами, звуками или картинками, при попытке вспомнить какое-то событие нам приходится восстанавливать изображения в голове, поскольку никаких реальных изображений не существует.

Если бы мы захотели «прочсть» чьи-то мысли и точно описать, что же происходит в новой коре, было бы чрезвычайно сложно интерпретировать воспоминания этого человека — вне зависимости от того, были бы это хранящиеся в коре образы, ожидающие того момента, когда их вызовут в качестве воспоминаний, или те, которые в данный момент активно обрабатываются в виде текущих мыслей. Мы бы «увидели» одновременную активацию миллионов распознающих модулей. Через сотую долю секунды мы бы обнаружили, что активирован другой набор примерно такого же числа модулей. Каждый такой образ был бы списком других образов, а каждый из тех — списком третьих и так далее до самых примитивных образов на самом нижнем уровне. Было бы чрезвычайно сложно определить, что означают образы высоких порядков, если не скопировать *всю* информацию на каждом уровне новой коры. Таким образом, каждый образ в коре имеет какое-то значение исключительно в свете всей информации, находящейся ниже него на иерархической лестнице. Более того, другие образы того же уровня и более высоких уровней тоже важны для интерпретации конкретного образа, поскольку обеспечивают контекст для его интерпретации. Поэтому истинное чтение мыслей требовало бы не только поиска активированных аксонов в голове человека, но и анализа всей новой коры со всеми воспоминаниями, что помогло бы интерпретировать активацию аксонов.

Когда речь идет о нас самих, мы «знаем», что означают наши мысли или воспоминания, но они не существуют в виде легко объяснимых мыслей или наборов изображений. Если мы хотим поделиться ими с другими людьми, нам нужно облечь их в речевую форму. Эту задачу тоже выполняет новая кора с помощью распознающих модулей, натренированных на образах, которыми мы пользуемся для речевого общения. Язык сам по себе обладает строгой иерархией и эволюционировал таким образом, чтобы использовать иерархическую природу новой коры, которая, в свою очередь, отражает иерархическую природу

реальности. Врожденная способность человека познавать иерархическую структуру языка, о которой писал Ноам Хомский\*, отражает структуру новой коры. В статье 2002 г. Хомский называет рекурсию уникальной особенностью человеческой речи<sup>4</sup>. Согласно Хомскому, рекурсия — это способность соединить мелкие элементы в более крупный элемент, а затем использовать этот более крупный элемент в качестве составной части еще одной структуры, и так до бесконечности. Действуя по такому принципу, мы можем выстраивать сложные структуры предложений, фраз и абзацев из ограниченного набора слов. Хотя Хомский в этой статье не говорил о структуре мозга, описанная им функция — именно то, что делает новая кора.

Низшие виды млекопитающих активно используют новую кору в своей каждодневной жизни. Люди приобрели дополнительные способности благодаря значительному увеличению размера коры в связи с формированием устного и письменного языка. Одни люди владеют этими способностями лучше, другие хуже. Если мы вынуждены пересказать какую-то историю много раз подряд, мы начнем заучивать фразы в виде набора отдельных предложений. Но даже в этом случае наши воспоминания являются не прямой последовательностью слов, а скорее языковыми структурами, которые мы должны переводить в специфические фразы при каждом пересказе. Вот почему всякий раз мы пересказываем историю чуточку иначе (если только не заучиваем точную последовательность слов как единый образ).

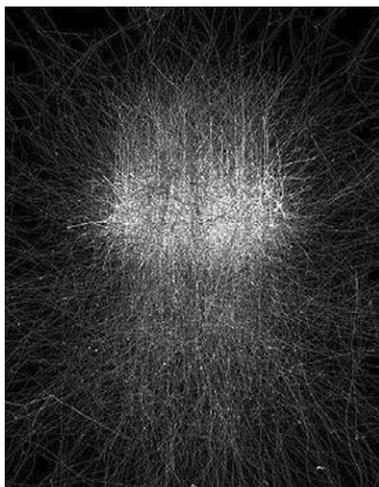
При описании специфических мыслительных процессов нельзя забывать об избыточности коры. Как я уже говорил, важные элементы нашей жизни, вне зависимости от того, относятся ли они к категории ощущений, речи или воспоминаний, записаны в нашей голове не по одному разу. Каждый важный образ на

---

\* Ноам Хомский (Наум Чомски; род. в 1928) — американский лингвист, публицист, политолог и философ; автор классификации языков, называемой иерархией Хомского.

каждом уровне повторяется множество раз. В некоторых случаях это простые повторы, а иногда — те же образы, но в другом ракурсе. Именно поэтому мы можем узнать знакомое лицо в разной ориентации и при разной освещенности. Избыточность характерна для каждой ступени иерархической лестницы, что позволяет распознавать различные вариации образов и понятий.

Итак, если бы мы захотели изучить процессы в нашей новой коре, например, в тот момент, когда смотрим на любимого человека, мы бы обнаружили множество возбужденных аксонов распознающих модулей всех уровней — от модулей, распознающих самые простые чувственные образы, до многих других, распознающих изображение близкого человека. Мы бы также обнаружили возбуждение многих других клеток, реагирующих на различные аспекты ситуации: движения человека, его слова и т. д. Таким образом, картина гораздо шире, чем просто упорядоченное перемещение по иерархической лестнице.



*Компьютерная симуляция одновременного возбуждения множества распознающих модулей новой коры.*

Но общий механизм иерархического сопряжения распознающих модулей, при котором каждый более высокий

концептуальный уровень отвечает за более абстрактное и интегрированное понятие, все же справедлив. Еще активнее осуществляется передача информации в обратном направлении, поскольку активированные распознающие модули каждого уровня посылают предсказательные сигналы нижестоящим модулям, сообщая о приближающемся событии. Кажущееся богатство человеческого опыта объясняется тем, что все сотни миллионов распознающих модулей нашей новой коры одновременно обрабатывают поступающие к ним сигналы.

В пятой главе мы поговорим о том, как тактильная, зрительная, звуковая и другая информация, полученная с помощью наших органов чувств, передается в новую кору. Эти исходные сигналы обрабатываются отделами коры, специально предназначенными для восприятия соответствующих импульсов (хотя функциональность различных областей мозга чрезвычайно пластична, что связано с универсальностью механизмов действия новой коры). Понятийная иерархия продолжается и за пределами наивысших понятий в каждой сенсорной зоне новой коры. В ассоциативных зонах коры происходит интеграция сигналов от различных чувствительных органов. Когда человек слышит что-то, напоминающее голос его жены, и видит что-то, что может говорить о ее присутствии, он не прибегает к сложным логическим умозаключениям, а на основании сочетания этих чувственных образов тотчас понимает, что жена где-то рядом. Он интегрирует все сенсорные и перцептивные доказательства — возможно, даже запах духов — в виде единого многоуровневого ощущения.

На уровне выше ассоциативных зон коры мы можем воспринимать, запоминать и осмысливать еще более сложные абстрактные понятия. На высшем понятийном уровне мы воспринимаем такие образы, как «это хорошо», «она привлекательна», «это забавно» и т. д. Наша память сохраняет эти абстрактные образы наравне с другими. Например, мы можем вспомнить, что говорили с какой-то женщиной, она сказала что-то забавное

и мы смеялись, хотя не можем вспомнить, в чем именно заключалась шутка. Память сохранила наше ощущение смешного, но не содержание разговора.

В предыдущей главе я обратил внимание на то, что часто нам удастся распознать образ, даже если мы не в состоянии его описать. Например, мне кажется, я смог бы выбрать из набора фотографий разных женщин фотографию той женщины с коляской, которую встретил на прогулке, хотя и не могу ее как следует себе представить или описать. В данном случае моя память о ней представляет собой список некоторых образов высокого уровня иерархии. Эти образы не имеют никаких речевых или зрительных пометок и не выражаются пикселями, поэтому я могу думать о ней, но не могу ее описать. Однако, если мне показать фотографии, мысленная обработка визуальной картинки приведет к узнаванию тех образов высокого уровня иерархии, которые были зарегистрированы при первой встрече. Так я смогу найти совпадение и выбрать ее фотографию среди других.

Даже если я встретил эту женщину на прогулке всего один раз, вполне возможно, что в моей новой коре сохранилось несколько копий ее образа. Однако, если я не думаю о ней какое-то время, соответствующие распознающие модули займутся обработкой других образов. Вот почему со временем воспоминания стираются: избыточность сокращается, и в какой-то момент некоторые воспоминания исчезают. Тот факт, что я написал об этой женщине в книге, скорее всего, позволит мне сохранить более устойчивое воспоминание о ней.

### **Самоассоциация и инвариантность**

В первой главе я уже писал о том, что нам удастся распознать образ, даже если он представлен не целиком или в искаженном виде. Первая способность, называемая самоассоциацией, заключается в умении ассоциировать образ и его часть. Структура каждого распознающего модуля позволяет реализовать эту функцию.

В процессе распознавания каждый сигнал поднимается от распознающего модуля более низкого уровня к модулю более высокого уровня, и эта связь имеет определенный «вес», указывающий на важность этого конкретного элемента в общем образе. Более важные элементы имеют больший вес при распознавании образа. Таким образом, бородака Линкольна, бачки Пресли и знаменитый высунутый язык Эйнштейна, по-видимому, имеют большой вес в тех образах, с которыми у нас ассоциируются эти знаменитости. Распознающие модули рассчитывают фактор вероятности, учитывающий параметр значимости элемента. Таким образом, общая вероятность распознавания снижается, если один или несколько элементов отсутствуют, но порог узнавания все же может быть преодолен. Как я уже говорил, расчет общей вероятности присутствия образа сложнее, чем расчет просто взвешенной суммы, в которой величина параметров тоже учитывается.

Если распознающий модуль получил сигнал от модуля более высокого уровня об «ожидании» образа, пороговое значение понижается (то есть его легче достичь). В альтернативном варианте такой сигнал может просто прибавляться к остальным входным сигналам, компенсируя недостающий элемент. Это происходит на всех уровнях, так что такой образ, как лицо, находящийся на несколько уровней выше базового, может быть распознан даже при отсутствии нескольких элементов.

Способность распознавать трансформированные образы называется инвариантностью и реализуется четырьмя способами. Во-первых, важные превращения образа происходят до того, как сигналы попадают в новую кору. О передаче сигнала от кожи, глаз и ушей мы поговорим в разделе «Сенсорное восприятие».

Второй способ основан на избыточности информации, хранящейся в новой коре. Особенно если речь идет о важных или распространенных вещах, мы многократно встречаемся

с разными вариантами и ракурсами одного и того же образа. Поэтому в новой коре независимо хранится и обрабатывается множество вариантов таких образов.

Третий и самый мощный способ заключается в способности объединять два списка. В одном может содержаться набор трансформаций, которые способны происходить с определенной категорией образов. Новая кора применит этот список возможных трансформаций к другому образу. Именно так мы понимаем речевые приемы типа метафор и сравнений.

Например, мы постепенно узнаём, что некоторые фонемы в разговорной речи могут выпадать (например, «когда» вместо «когда»). Если затем мы слышим другое слово (например, «всегда»), то узнаём его, даже если в нем отсутствует одна из фонем («всегда»), поскольку уже знакомы с явлением выпадения некоторых фонем. Другой пример. Мы знаем, что какой-то актер любит подчеркивать (увеличивать) размеры части лица (например, носа). Это позволяет нам узнать знакомое лицо, с которым проделан такой же трюк, даже если раньше с этим лицом подобные модификации не производились. Некоторые варианты актерского грима подчеркивают те самые черты, которые распознаются распознающими модулями новой коры. На этом же строится прием карикатуры.

Четвертый метод основан на величине параметров и позволяет одному модулю кодировать множество примеров образа. Например, мы много раз слышали слово *steep*. Модуль, ответственный за распознавание этого слова в речи, может кодировать множество примеров его произношения с разной длительностью [E]. Если все модули, распознающие слова с [E], обладают этим общим свойством, такая варибельность должна быть зарегистрирована в самом образе [E]. Однако разные слова с этим звуком (а также со многими другими фонемами) могут различаться по параметру ожидаемой варибельности. Например, слово *peak* (читается «пик») не содержит фонемы [E], как в слове *steep*.

## Обучение

Не сами ли мы создаем себе преемников в земном господстве, ежедневно совершенствуя их красоту и утонченность, ежедневно обучая их новым навыкам и постоянно снабжая новыми способностями саморегулирования и автоматизма, которые будут лучше любого интеллекта?

*Сэмюэл Батлер\**

Главная работа мозга заключается в изменении самого себя.

*Марвин Минский\*\**. *Общество разума*. 1986

До сих пор мы говорили о том, как мы распознаём (сенсорно-перцептивными методами) образы и вызываем в памяти последовательности образов (наши воспоминания о вещах, людях или событиях). Однако, когда мы родились, ничего этого в нашей новой коре не было. При формировании мозга новая кора девственно чиста. Она обладает способностью учиться и, следовательно, создавать связи между распознающими модулями, но все эти связи появляются только с опытом.

Процесс обучения начинается еще до нашего рождения — одновременно с биологическими процессами в развивающемся мозге. У месячного зародыша уже есть головной мозг, хотя он пока еще напоминает мозг рептилии, поскольку в процессе внутриутробного развития зародыш в ускоренном темпе проходит все этапы эволюции. К третьему триместру беременности мозг плода — уже очевидно человеческий мозг с человеческой новой корой. В этот период плод набирается определенного опыта, и его новая кора обучается. Он слышит звуки, особенно

---

\* Сэмюэл Батлер (1835–1902) — английский писатель, художник и переводчик.

\*\* Марвин Ли Минский (род. в 1927) — американский ученый в области искусственного интеллекта, один из основателей лаборатории искусственного интеллекта в Массачусетском технологическом институте.

сердцебиение матери, кстати, это одна из возможных причин, объясняющих ритмичность музыки во всех человеческих культурах. Музыка является составной частью культуры всех известных на сегодня человеческих цивилизаций, чего нельзя сказать о других видах искусства, например о живописи. Кроме того, музыкальные ритмы сравнимы с ритмом нашего сердца. Конечно, музыкальные ритмы сильно различаются, иначе было бы неинтересно, но и ритмы сердца тоже различаются. Между прочим, абсолютно регулярное биение сердца является симптомом заболевания. Через 26 недель после зачатия глаза плода начинают открываться, а через 28 недель — практически постоянно открыты. Внутри матки мало интересного, но новая кора плода начинает распознавать образы света и темноты.

Таким образом, опыт новорожденного ребенка весьма ограничен. Новая кора может учиться и у старого мозга (мы поговорим об этом в четвертой главе), но, в общем, ребенку предстоит многое узнать, начиная от распознавания самых примитивных звуков и форм и заканчивая метафорами и сарказмом.

Обучение играет важнейшую роль в формировании человеческого разума. Если бы мы смогли создать идеальную модель человеческой новой коры (что и планируется сделать в рамках проекта Blue Brain) и всех других отделов мозга, которые нужны для его функционирования (таких как гиппокамп и таламус), такой мозг мало на что был бы способен — как и мозг новорожденного ребенка (за исключением того, чтобы делать этого ребенка милым, поскольку это важнейшее условие выживания).

Обучение и распознавание происходят одновременно. Мы начинаем обучаться сразу, и как только мы изучили образ, мы сразу же начинаем его узнавать. Новая кора постоянно пытается определить смысл поступающих в нее сигналов. Если какой-то иерархический уровень не справляется со своей задачей и не может полностью идентифицировать образ, сигналы направляются на уровень более высокого порядка. Если же распознавание

не происходит ни на одном из уровней, образ воспринимается как новый. Отнесение образа к категории новых образов не обязательно означает, что новыми являются все его элементы. Если мы смотрим на картину какого-то художника и видим на ней кошачью морду со слоновьим хоботом, мы узнаем все элементы, но понимаем, что этот комбинированный образ является новым, и запоминаем его. На более высоких понятийных уровнях иерархии новой коры, воспринимающих контекст (например, что данная картина является плодом творчества конкретного художника и что мы пришли на новую выставку его работ), будет зарегистрирован необычный комбинированный образ слонкота, а все детали контекста сохранятся в виде дополнительных воспоминаний.

Новые воспоминания вроде морды слонкота записываются в доступном распознающем модуле. В этом процессе задействован гиппокамп, и о том, как все происходит на самом деле, мы поговорим в следующей главе. В рамках обсуждаемой модели новой коры достаточно сказать, что неопознанные образы сохраняются в виде новых образов и связаны с составляющими их образами более низкого уровня. В частности, морда слонкота записывается несколькими способами: регистрируется новое расположение черт, а также имя художника, ситуация, возможно, наше воспоминание о том, как мы смеялись, впервые увидев картину.

Успешно распознанные воспоминания также могут сформировать новый образ, позволяющий дополнительно увеличить избыточность информации. Если же образы распознаны недостаточно хорошо, они, скорее всего, сохраняются как варианты распознанного образа в новом ракурсе.

Каким же методом в целом можно определить, какие образы сохраняются? На языке математики проблема формулируется так: как оптимальным образом представить входные сигналы, которые уже встречались ранее, чтобы не превышать доступной емкости новой коры? Некоторая избыточность необходима, но

было бы нерационально заполнять все отпущенное для хранения информации пространство (то есть всю новую кору) повторяющимися образами, поскольку в таком случае сокращается разнообразие запоминаемых образов. Например, образ звука [и] мы встречали бесконечное число раз. Это простой образ звуковой частоты, который, несомненно, представлен в коре с очень большой избыточностью. Мы можем заполнить всю кору повторами образа звука [и], однако полезная избыточность имеет пределы, которые в случае такого распространенного образа, конечно же, были достигнуты.

Подобная математическая проблема оптимизации решается с помощью так называемого линейного программирования, позволяющего наилучшим образом использовать ограниченные ресурсы (в нашем случае — ограниченное число распознающих модулей) для представления всех примеров, на которых отработывалась модель. Метод линейного программирования был разработан для систем с одномерными входными параметрами, и это еще одна причина, по которой она оптимально подходит для описания линейной последовательности входных сигналов. Мы можем использовать этот математический подход для создания компьютерных программ, и хотя реальный мозг ограничен имеющимися физическими связями, которые он распределяет между распознающими модулями, метод тем не менее очень похожий.

Важным результатом подобной оптимизации является то, что постоянно встречающиеся образы распознаются, но не приводят к возникновению стойких воспоминаний. На утренней прогулке я пережил множество впечатлений, распознаваемых на всех иерархических уровнях, — от простых видимых углов и теней до более сложных объектов, таких как фонарные столбы, почтовые ящики, люди, животные и растения. Скорее всего, ни один из них не был уникален, и распознанные мной образы уже давно достигли оптимального уровня избыточности. В результате у меня не осталось практически никаких воспоминаний

об этой прогулке. Те немногие детали, которые я запомнил, по-видимому, были вытеснены записанными поверх них новыми образами, зарегистрированными мозгом во время следующих прогулок, а эту конкретную прогулку я запомнил лишь по той причине, что написал о ней в книге.

Один важный аспект касательно функционирования нашей новой коры и попыток ее моделирования заключается в трудности одновременного постижения образов, относящихся сразу ко многим понятийным уровням. Вообще говоря, мы можем одновременно осваивать один, максимум два уровня. Если процесс обучения проходит стабильно, мы можем переходить на следующий уровень. Мы можем продолжать более тонкую настройку нижних уровней, но сфокусированы на следующем более высоком уровне абстракции. Это справедливо как в самом начале жизни, когда новорожденный ребенок осваивает основные формы и предметы, так и впоследствии, когда мы осваиваем новые материи — один уровень сложности за раз. То же самое можно сказать и о компьютерном моделировании новой коры. Если машине представлять материал в порядке увеличения абстрактности — за один раз один уровень, машины способны обучаться точно так же, как обучается человек (хотя пока не воспринимают такого множества понятийных уровней).

Выходные сигналы от каких-то распознающих модулей могут идти обратно к образам более низкого уровня или возвращаться к тем же самым образам, что объясняет мощную рекурсивную способность человеческого мозга. Отдельный элемент образа может быть точкой принятия решений для другого образа. Это особенно полезно для списков, в состав которых входят действия: например, нужно взять новый тюбик зубной пасты, если закончился старый. Такие условные зависимости существуют на всех уровнях. Каждый, кто пытался создать компьютерную программу, знает, что условные зависимости — обязательный элемент для описания последовательности действий.

## Язык мысли

Сны являются предохранительными клапанами перегруженного мозга.

*Зигмунд Фрейд. Толкование сновидений. 1911*

Головной мозг: мы думаем, что это аппарат, с помощью которого мы думаем.

*Амброз Бирс\*. Словарь Сатаны*

Чтобы суммировать все то, что мы узнали о функционировании новой коры, вернитесь, пожалуйста, к диаграмме «Распознающий модуль новой коры» в начале данной главы (см. рис. 8).

А. Дендриты входят в модуль, распознающий конкретный образ. Даже если образы кажутся двумерными или трехмерными, они представляются в виде одномерных последовательностей сигналов. Чтобы произошло распознавание, образ должен быть представлен в виде такой последовательности. Каждый дендрит связывается с одним или несколькими аксонами распознающих модулей более низкого понятийного уровня, распознавших более простые образы, являющиеся частью данного образа. Для каждого входного сигнала может существовать несколько модулей более низкого уровня, способных генерировать сигналы. Необходимое для распознавания образа пороговое значение вероятности может быть достигнуто даже в том случае, если получены не все входные сигналы. Модуль рассчитывает вероятность присутствия образа, за распознавание которого он отвечает. Этот расчет основан на параметрах «значимости» и «величины» сигналов (см. ниже).

Обратите внимание, что одни дендриты передают сигналы в модуль, а другие — из модуля. Если все дендриты, входящие в распознающий модуль, сигнализируют о том, что распознаны все образы более низкого уровня, за исключением одного

---

\* Амброз Бирс (1842–1913/1914) — американский писатель и журналист, составитель сатирического сборника афоризмов «Словарь Сатаны».

или двух, распознающий модуль отправит нисходящие сигналы в модули, ответственные за распознавание этих оставшихся неузнанными образов, с указанием высокой вероятности того, что образ скоро будет распознан и что эти модули более низкого порядка должны быть к этому готовы.

Б. Когда распознающий модуль узнаёт соответствующий образ (на основании активации всех или большинства входных дендритов), активируется выходной аксон этого распознающего модуля. Этот аксон, в свою очередь, может сообщаться со всей сетью дендритов, ведущих ко многим распознающим модулям более высокого порядка, для которых данный образ является входным сигналом. Сигнал и его величина учитываются распознающими модулями следующего уровня.

В. Если распознающий модуль более высокого уровня получает положительные сигналы ото всех или от большинства составляющих образов, за исключением сигнала от данного конкретного распознающего модуля, распознающий модуль более высокого уровня может послать нисходящий сигнал этому модулю, указывая на большую вероятность «ожидания» соответствующего образа. Этот сигнал заставляет распознающий модуль понизить пороговые параметры, в результате чего он с большей вероятностью передает по своему аксону положительный сигнал (что означает, что соответствующий образ распознан), даже если некоторые из входных сигналов отсутствуют или сомнительны.

Г. Ингибирующие сигналы снизу понижают вероятность распознавания модулем соответствующего образа. Это может происходить в результате распознавания на более низком уровне образов, несовместимых с образом, за который отвечает данный распознающий модуль (например, распознавание усов снижает вероятность того, что увиденное мной лицо принадлежит «моей жене»).

Д. Ингибирующие сигналы сверху тоже могут снижать вероятность распознавания образа соответствующим модулем. Это

может происходить при распознавании образа более высокого порядка сложности, который несовместим с образом, распознанным данным модулем.

Е. Каждому сигналу соответствует ранее определенный набор параметров, таких как значимость, ожидаемая величина и ожидаемая вариабельность. Распознающий модуль рассчитывает общую вероятность присутствия образа на основании значений всех этих параметров для входных сигналов. Оптимальный математический метод решения данной задачи называется методом скрытых моделей Маркова. Если такие модели организованы иерархическим образом (как в новой коре или в математических моделях, воспроизводящих ее функции), мы называем их иерархическими скрытыми моделями Маркова.

Распознавание корой одних образов запускает распознавание следующих. Частично распознанные образы посылают нисходящие сигналы, полностью распознанные образы — восходящие сигналы. Эти распознанные новой корой образы являются языком мышления. Как и разговорный язык, язык мышления имеет иерархическую структуру, но не является «языком» в обычном понимании. Поначалу наши мысли не выражены словами, однако, поскольку речь тоже формируется в новой коре в виде иерархических образов, у нас бывают и мысли, основанные на языковых образах. Однако по большей части мысли выражаются неязыковыми образами.

Как я уже отмечал ранее, даже если бы мы умели регистрировать активацию образов в новой коре живого человека, мы все равно не смогли бы определить значение этих образов, если бы не имели доступа к полной иерархической системе вышестоящих и нижестоящих образов. Это сильно затрудняет доступ к мыслительному процессу другого человека. Достаточно сложно понять содержание наших собственных мыслей, а для понимания мыслей другого человека требуется овладеть его новой корой. Конечно же, пока нам это недоступно. Пока нам приходится использовать способность другого человека выражать

свои мысли с помощью речи (а также других средств, включая жесты). Учитывая недостаточную способность людей справляться с этой задачей, нет ничего удивительного в том, что мы зачастую плохо понимаем друг друга.

У нас два способа мышления. При ненаправленном способе мышления мысли подталкивают друг друга нелогичным образом. Иногда, когда мы заняты чем-то отвлеченным, например сгребая опавшие листья или просто идем по улице, у нас неожиданно возникают воспоминания о том, что мы делали несколько лет или десятилетий назад. Мы не можем немедленно восстановить в памяти те события, не прибегая к множеству других воспоминаний, которые помогают нам воссоздать более слаженную картину. Если нам удастся воссоздать визуальное изображение давно забытой сцены, это происходит в настоящий момент, поскольку память не хранит визуальные образы или изображения. Как я уже говорил, не всегда очевидно, какие стимулы вызывают у нас в голове появление подобных мыслей. Последовательность мыслей, вызвавших воспоминание, может быть немедленно забыта. Но если даже нам удастся ее восстановить, это окажется нелинейная, круглая последовательность ассоциаций.

Второй способ — направленное мышление, которое мы используем для решения задач или формулировки логичного ответа. Например, так мы готовимся кому-то что-то сказать или формулируем в уме фразу, которую собираемся написать (например, в книге о разуме). Когда мы обдумываем подобную проблему, мы расчлняем ее на иерархические составляющие. Например, чтобы написать книгу, нужно написать отдельные главы. Каждая глава строится из разделов, разделы состоят из параграфов, параграфы — из предложений, в которых сформулированы идеи. Каждая идея построена из нескольких элементов в определенной конфигурации. Все элементы идеи и все связи между элементами тоже должны быть произнесены. Кроме того, наша новая кора действует по определенным правилам. Если мы

выполняем письменную работу, нужно постараться избегать ненужных повторений, обеспечить читателю возможность следить за ходом изложения, использовать грамматические и стилистические правила и т. д. Следовательно, писатель должен создать в своей голове образ читателя, и этот образ также является иерархическим. При направленном мышлении мы продвигаемся по спискам новой коры, каждый из которых, в свою очередь, состоит из подписков, а те — из еще более простых наблюдений. Помним также, что элементы списков могут содержать условные зависимости, так что наши мысли и действия зависят от оценок, сделанных в ходе мыслительного процесса.

Кроме того, каждая направленная мысль запускает иерархический путь ненаправленного мышления. Происходит постоянный наплыв воспоминаний, связанных как с нашим чувственным опытом, так и с нашими попытками направленного мышления. Наш мыслительный опыт сложен и запутан и состоит из наплывов образов, изменяющихся сотни раз в секунду.

## **Язык снов**

Сны — пример ненаправленных мыслей. В них есть некий смысл, поскольку одни мысли запускают другие в соответствии с реальными связями между образами, хранящимися в новой коре. Неосмысленные элементы снов мы пытаемся зафиксировать с помощью нашей способности к конфабуляции (фантазии). Как я расскажу в девятой главе, больные с расщепленным мозгом (при поражении мозолистого тела, связывающего между собой полушария мозга) прибегают к конфабуляциям (с помощью левого полушария, контролирующего речевой центр) для объяснения того, что делает с полученными сигналами правое полушарие, к которому у левого полушария нет доступа. Мы постоянно прибегаем к конфабуляциям для объяснения исхода различных событий. Хотите яркий пример — просто прослушайте дневные комментарии по поводу состояния финансовых рынков. Вне зависимости от того, как изменяется ситуация на

рынках, всегда можно найти причины этих изменений. И примеров подобных комментариев можно привести множество. Если бы комментаторы действительно понимали ситуацию, они не тратили бы время на составление подобных комментариев.

Конечно же, за конфабуляции тоже отвечает новая кора; они возникают в тех случаях, когда мы встречаемся с некоторыми трудностями при пересказе каких-то историй или объяснении каких-то фактов. Мы заполняем свой рассказ недостающими или забытыми подробностями, и он становится более осмысленным. Вот почему истории в устах разных рассказчиков, возможно, имеющих разные намерения, изменяются со временем. Однако с появлением письменного языка у нас появилась возможность фиксировать окончательную версию истории и предотвращать подобные вариации.

Истинное содержание сна в том виде, в каком мы его запоминаем, опять-таки представляет собой последовательность образов. Образы составляют основу истории, а мы дополняем ее конфабуляциями, ограниченными существующими образами. Эта фантазия и есть та версия сна, которую мы можем пересказать или вспомнить. Пересказывая сон, мы запускаем каскад образов, дополняющих наш сон в том виде, в каком мы увидели его изначально.

В этом заключается одно из ключевых различий между нашими мыслями во сне и при бодрствовании. В процессе жизни мы узнаём, что существуют действия и даже мысли, недопустимые в реальном мире. Например, мы узнаём, что не можем немедленно осуществить свои желания. Закон запрещает вытаскивать деньги из кассы в магазине или вступать в связь с любым человеком, который просто показался физически привлекательным. Мы узнаём, что какие-то мысли недопустимы ввиду культурных ограничений. Когда мы приобретаем профессиональные навыки, мы учимся мыслить таким способом, который признаётся и вознаграждается в нашей профессии, и привыкаем избегать того образа мыслей, что может нарушать

нормы и методы профессии. Многие из этих табу имеют смысл, поскольку поддерживают социальный порядок и способствуют прогрессу. Но они могут и тормозить прогресс, усиливая непродуктивный консерватизм. Именно такой консерватизм преодолел Эйнштейн, когда в своих мысленных экспериментах попытался прокатиться на световом луче.

Культурные правила закрепляются в новой коре с помощью старого мозга, особенно миндалевидной железы. Каждая наша мысль запускает другие мысли, и некоторые из них связаны с определенной опасностью. Например, мы узнаём, что нарушение культурных норм даже в нашей частной жизни, может привести к остракизму, и новая кора понимает, что это опасно для нашего благополучия. Если у нас рождается подобная мысль, начинает работать миндалевидная железа, вызывая страх, и мысль прерывается.

Однако во сне табу ослаблены, и нам часто снится нечто запрещенное в культурном, сексуальном или профессиональном плане. Как будто наш мозг понимает, что во сне мы не являемся реальными действующими лицами. Об этом писал Фрейд, который также отмечал, что мы прячем опасные мысли, по крайней мере, когда пытаемся их вспомнить, так что проснувшийся мозг продолжает быть защищенным.

Ослабление профессиональных табу может быть полезным для творческого решения проблем. Каждый вечер перед тем, как идти спать, я думаю о какой-либо специфической задаче. Это запускает последовательность мыслей, которая продолжается во сне. Во сне я могу обдумывать — *видеть во сне* — решение задачи, не ограничивая себя теми профессиональными табу, что действуют днем. Утром между сном и пробуждением я могу вспомнить эти мысли; такое состояние иногда называют «осознанными сновидениями»<sup>5</sup>.

Всем известно, что Фрейд писал о возможности понять психологию человека путем интерпретации его снов. О различных аспектах этой теории существует обширная литература, но

основное положение о возможности изучения самих себя путем анализа собственных мыслей вполне осмысленно. Наши сны создаются в новой коре, поэтому отражают ее содержимое и имеющиеся в ней связи. Ослабление мыслительных запретов, действующих при бодрствовании, тоже полезно для анализа содержимого коры, к которому в противном случае мы не имели бы прямого доступа. Резонно предположить также, что образы, формирующие наши сны, важны для нас и поэтому являются ключом к пониманию наших скрытых страхов и желаний.

### **Суть модели**

Как я уже говорил ранее, в 1980–1990-х гг. я руководил группой исследователей, пытавшихся применить метод иерархических скрытых моделей Маркова для распознавания и понимания человеческой речи. Эта работа предшествовала появлению широко распространенных сегодня коммерческих систем, распознающих и понимающих нашу речь (автомобильных навигационных систем, Сири в айфонах и многих других). Созданная нами технология строилась практически по тем же принципам, что реализуются при мысленном распознавании образов. Она также была основана на иерархии образов, при которой каждый более высокий уровень отличался от нижестоящего уровня большей абстрактностью. Например, в системе распознавания речи основные образы звуковой частоты формировали нижние уровни, за ними следовали фонемы, затем слова и фразы (часто распознававшиеся так, как будто были словами). Некоторые системы распознавания речи могут понимать речевые команды и в таком случае включают в себя еще более высокие иерархические уровни, содержащие такие структуры, как именные и глагольные группы. Каждый распознающий модуль умеет распознавать линейную последовательность образов с нижестоящего понятийного уровня. Каждый входной сигнал характеризуется значимостью, величиной и вариабельностью величины. Существуют также и нисходящие сигналы, указывающие на ожидание

образа низшего уровня. Подробнее я расскажу об этих исследованиях в седьмой главе.

В 2003 и 2004 гг. создатели карманного персонального компьютера Джеф Хокинс и Дайлип Джордж сформировали иерархическую модель коры, названную временной иерархической памятью. В соавторстве с писательницей Сандрой Блейкли Хокинс описал эту модель в книге «Об интеллекте» (On Intelligence). Хокинс доказывает универсальность алгоритма действия коры и ее иерархической организации в виде списков. Между моделью Хокинса и моделью, которую предлагаю вам я в своей книге, есть несколько существенных различий. Как следует из названия модели, Хокинс настаивает на временной (основанной на времени) природе списков. Другими словами, списки состояются всегда в одном и том же направлении — вперед во времени. Наличие временного направления у элементов двумерных образов, таких как печатная буква «А», он объясняет движением глаз. Он объясняет, что мы визуализируем изображения при помощи саккад — очень быстрых произвольных движений глаз. Таким образом, информация поступает в новую кору не в виде двумерных наборов данных, а скорее в виде списков организованных во времени элементов. Это правда, что наши глаза совершают очень быстрые движения, однако они не всегда видят элементы образов (таких как буква «А») в последовательном временном порядке. Например, глаза не обязательно сначала регистрируют верхний угол буквы «А», а затем ее нижнюю часть. Кроме того, мы способны распознать зрительный образ, видимый лишь на протяжении нескольких десятых миллисекунды, а за это время саккадные движения глаз просто не успевают его просканировать. Верно, что распознающий модуль новой коры сохраняет образ в виде списка и что список этот хранится в соответствии с определенным порядком, только порядок этот не обязательно временной. Он действительно может быть временным, но также может быть пространственным или понятийным, как я объяснял выше.

Но самое главное различие между моделями заключается в наборе параметров, с помощью которых я характеризую каждый сигнал, входящий в распознающий модуль, в частности параметров величины и ее вариабельности. В 1980-х гг. мы пытались распознавать человеческую речь без учета информации подобного рода. Тогда лингвисты уверяли нас, что информация о длительности сигнала не играет решающей роли. Такой подход используется в словарях, в которых произношение каждого слова передается в виде последовательности фонем. Например, слово *steep* записывается как последовательность звуков [s], [t], [E] и [p] без указания ожидаемой длительности каждой фонемы. В результате, если созданная нами программа, способная распознавать фонемы, встречает в речи эту специфическую последовательность четырех фонем, она сможет распознать слово. Созданная по такому принципу программа работала, но недостаточно хорошо, чтобы справляться с большим набором слов, распознавать речь нескольких говорящих или слова, произнесенные без пауз. Качество программы выросло в значительной степени, когда мы с помощью иерархических скрытых моделей Маркова ввели для каждого входного сигнала параметр распределения величины.



# Глава четвертая

## БИОЛОГИЧЕСКАЯ НОВАЯ КОРА

Поскольку дело серьезное, вам понадобится череп для вашего мозга, пластиковый футляр для расчески и бумажник для денег.

*Джордж Костанца, персонаж американского телесериала «Сайнфилд», 1998*

Теперь мы впервые наблюдаем за работой мозга в глобальном плане с такой ясностью, что сможем найти общие программы его удивительных способностей.

*Дж. Тейлор, Б. Горовиц, К. Фристон\**

Работу мозга над полученной информацией можно в определенной степени сравнить с работой скульптора над каменной глыбой. В том смысле, что статуя всегда находилась внутри камня. Однако кроме нее в камне хранились еще тысячи других, и только благодаря скульптору из него была извлечена именно эта. Так же и мир каждого из нас, как бы по-разному мы его ни воспринимали, погружен в изначальный хаос ощущений, который всем нам дает одинаковую *материю* для размышлений. Если захотим, мы можем мысленно повернуть время вспять к той черной и единой бесконечности пространства и движущихся и роящихся атомов, какую наука называет единственным реальным миром. Но все-таки мир, который *мы* воспринимаем и в котором живем, такой, какой наши предки и мы сами постепенно высекали из этого первозданного пространства, просто отделяя определенные фрагменты. Другой скульптор — другая статуя из того же камня! Другой разум — другой мир из того же самого монотонного и невыразительного хаоса! Мой мир — лишь один из похожих вкрапленных миров, столь же реальных для тех, кто умеет абстрагироваться. Сколь другим, должно быть, является мир в сознании муравья, каракатицы или краба!

*Уильям Джемс\*\**

---

\* J.G. Taylor, B. Horwitz, K.J. Friston, Neural modeling and functional brain imaging: an overview, *Neural Networks*, Volume 13, Nov. 2000, 829–846.

\*\* Уильям Джемс (или Джеймс; 1842–1910) — американский философ и психолог, один из основателей прагматизма.

**Я**вляется ли разум главной целью или лишь одной из целей биологической эволюции? Стивен Пинкер\* пишет: «Мы рассматриваем свой мозг в качестве вершины эволюции<sup>1</sup>... но это [мнение] неоправданно... поскольку естественный отбор не имеет никакого отношения к развитию интеллекта. Этот процесс определяется различием в выживаемости и воспроизводстве организмов в специфических условиях окружающей среды. Со временем организмы приобретают черты, позволяющие им адаптироваться к жизни и воспроизводству в данных конкретных условиях; ничто, кроме их успешного процветания здесь и сейчас, не движет их развитием ни в каком направлении». Далее Пинкер заключает: «Жизнь похожа на густой и ветвистый куст, а не на линейку или лестницу, и живые организмы располагаются на концах ветвей, а не у их основания».

В отношении человеческого мозга он задается вопросом, насколько «преимущества превышают затраты». К затратам он относит «большой объем мозга. Таз женщины едва вмещает в себя голову ребенка. Такое компромиссное строение является причиной гибели множества женщин в момент деторождения и объясняет раскачивающуюся походку женщин, из-за которой по биомеханическим причинам женщины перемещаются менее эффективно, чем мужчины. Кроме того, болтающаяся на шее тяжелая голова повышает вероятность смертельных повреждений при таких несчастных случаях, как падения». Он продолжает список недостатков, среди которых большой расход энергии мозгом, медленные реакции и медленное обучение.

В принципе, все эти утверждения точны (хотя многие знакомые мне женщины ходят гораздо легче, чем я), однако Пинкеру не хватает обобщений. Это верно, что биологическая эволюция не имеет специфического направления. Скорее она ищет пути среди «густых и ветвистых кустов» природы. Так же верно, что

---

\* Стивен Пинкер — канадско-американский ученый, психолог, популяризатор науки.

эволюционные изменения *не обязательно* происходят в направлении повышения интеллекта — они идут *по всем* направлениям. Можно привести множество примеров успешных организмов, которые практически не изменились за миллионы лет (скажем, аллигаторы появились 200 млн лет назад, а многие микроорганизмы — еще раньше). Однако в процессе последовательного развития myriad эволюционных ветвей одно из направлений *действительно* ведет к повышению интеллекта. И именно этот факт является предметом нашей дискуссии.

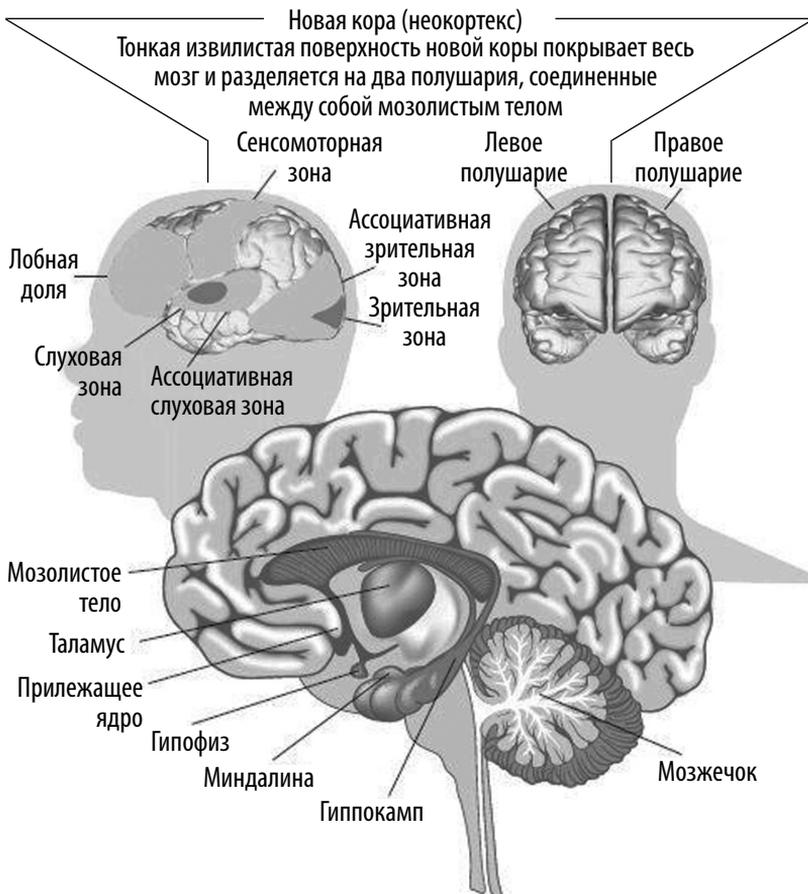
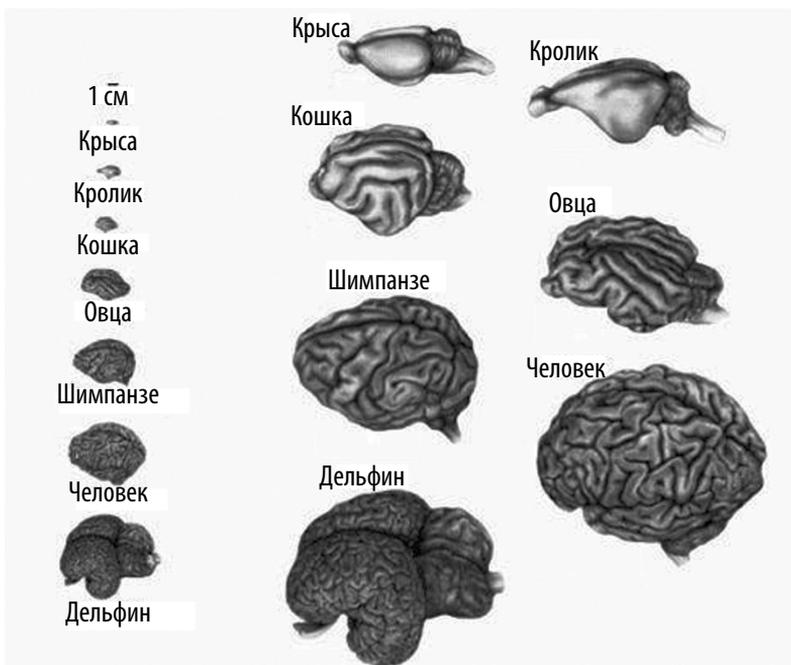


Схема расположения ключевых отделов мозга человека.



*Новая кора некоторых млекопитающих*

Предположим, у нас есть банка с синим газом. Когда мы открываем крышку, к молекулам газа не поступает никакого сигнала типа «эй, ребята, крышка открыта, путь свободен!» Молекулы просто продолжают делать то же, что обычно, то есть движутся во всех направлениях без какого-либо предпочтения. Но при этом те из них, которые находятся в верхней части банки, действительно выходят наружу, и со временем большинство молекул покинет банку. Однажды биологическая эволюция создала механизм иерархического обучения, и он оказался чрезвычайно полезным для реализации одной из ее задач — выживания видов. Преимущество обладания новой корой становится совершенно очевидным в тех ситуациях, когда быстро меняющиеся обстоятельства благоприятствуют быстрому обучению. Самые разные виды организмов — и растения, и животные — учатся адаптироваться к изменяющимся условиям, однако за неимением новой

кору им приходится использовать процесс генетической эволюции. Для обучения новым стратегиям поведения (или другим адаптационным стратегиям в случае растений) организмам, не имеющим новой коры, требуются тысячи лет (множество поколений). Выдающимся преимуществом организмов с новой корой является возможность обучения на протяжении нескольких дней. Если вид попадает в ситуацию, характеризующуюся драматическими изменениями внешних условий и один из представителей вида изобретает, открывает или просто случайно обнаруживает (все эти три метода являются вариантами инноваций) путь адаптации к этим изменениям, другие представители вида замечают, усваивают и копируют этот метод, который в результате быстро, как вирус, распространяется во всей популяции. Мел-палеогеновое вымирание, произошедшее около 65 млн лет назад, стало причиной исчезновения многих видов организмов, которые не имели новой коры и не смогли достаточно быстро адаптироваться к резким изменениям окружающей среды. Это был поворотный момент в развитии млекопитающих, занявших освободившиеся экологические ниши. Так биологическая эволюция «обнаружила», что способность новой коры к иерархическому обучению настолько ценное приобретение, что данная область головного мозга продолжала увеличиваться и теперь составляет основную часть мозга *Homo sapiens*.

Открытия в нейробиологии подтвердили ключевую роль иерархических способностей новой коры, а также представили доказательства теории мысленного распознавания образов (ТРО). Эти доказательства складываются из множества наблюдений и анализов, часть которых я здесь представлю. Канадский психолог Дональд Хебб (1904–1985) первым предпринял попытку идентифицировать неврологические основы процесса обучения. В 1949 г. он описал механизм физиологического изменения нейронов в зависимости от пережитого опыта, что может быть основой обучаемости и пластичности головного мозга. Он писал: «Предположим, что сохранение или повторение

реверберационной активности способствует длительным клеточным изменениям... Если аксон клетки *A* находится достаточно близко, чтобы возбуждать клетку *B*, и возбуждает ее непрерывно или многократно, происходит некий рост или определенные метаболические изменения в одной или в обеих клетках, в результате чего возрастает эффективность возбуждения клеткой *A* клетки *B*<sup>2</sup>. Эта теория, выраженная в тезисе *cells that fire together wire together* («между клетками, которые возбуждаются одновременно, возникает прочная связь»), стала известна как «правило Хебба». Идея Хебба получила подтверждение, и теперь известно, что клетки мозга могут образовывать новые связи и усиливать их благодаря собственной активности. Теперь сканирование головного мозга позволяет наблюдать за развитием подобных связей. Строение искусственных сетей нейронов основано на теории Хебба об обучении нейронов.

Теория Хебба строится на предположении, что основной обучающейся единицей новой коры является нейрон. Теория распознавания образов, которую предлагаю вам я в данной книге, строится на другой основополагающей единице — не на отдельном нейроне, а скорее на совокупности нейронов, число которых в этой совокупности, по моим оценкам, близко к сотне. Связи и синаптические взаимодействия *внутри* каждой такой единицы достаточно прочные и определены генетически: организация каждого распознающего модуля генетически предопределена. Обучение происходит в процессе образования связей *между* этими единицами, а не внутри них и, возможно, связано с усилением синаптической передачи в этих связях.

Недавно новые подтверждения того, что основной единицей обучения является модуль из нескольких десятков нейронов, были получены швейцарским нейрофизиологом Генри Марккромом (род. в 1962 г.), чей амбициозный проект «Голубой мозг» (*Blue Brain*) я представляю в седьмой главе. В статье 2011 г. он описывал свои эксперименты по сканированию и анализу нейронов новой коры млекопитающих для «поиска доказательств существования

ансамблей Хебба на самом примитивном уровне организации коры». Однако вместо них он обнаружил «трудноуловимые ансамбли [клеток], связь и синаптический вес которых вполне предсказуемы и ограничены». На основании этих результатов Маркрам пришел к заключению, что «опыт не может просто моделировать синаптические связи этих ансамблей» и что «они служат врожденными строительными блоками знаний для восприятия типа кубиков лего, а формирование памяти включает в себя соединение этих кубиков в сложные конструкции». Он продолжает: «О существовании функциональных ансамблей нейронов было известно уже несколько десятилетий, однако не существовало прямого доказательства наличия кластеров синаптически связанных нейронов... Поскольку все эти ансамбли похожи между собой по топологии и синаптическому весу и не сформированы под действием какого-либо специфического опыта, можно считать, что это врожденные ансамбли... а опыт играет лишь минорную роль в формировании синаптических связей и их весов... Наши исследования подтвердили существование врожденных ансамблей типа кубиков лего, состоящих из нескольких десятков нейронов... Благодаря связям между ансамблями в слое новой коры могут возникать более сложные структуры, затем ансамбли еще более высокого порядка в кортикальной колонке, еще более сложные ансамбли отделов мозга, и, наконец, самый сложный ансамбль соответствует всему головному мозгу в целом... Накопление воспоминаний сильно напоминает строительство из кубиков лего. Каждый ансамбль эквивалентен блоку лего, содержащему некоторое количество элементарных врожденных знаний о том, как воспринимать внешний мир и реагировать на него... При соединении разных блоков образуется уникальная комбинация этих врожденных воспринимаемых образов, составляющих специфические знания и опыт индивидуума»<sup>3</sup>.

Предлагаемая Маркрамом модель «кубиков лего» полностью совместима с описанной мной теорией распознающих модулей. В электронном письме Маркрам назвал эти «кубики лего»

«общим и врожденным знанием»<sup>4</sup>. Я считаю, что задача модулей заключается в распознавании и запоминании образов, а также их предсказании на основании частичных образов. Заметим, что предположение Маркрама о том, что в состав каждого модуля входит «несколько дюжин нейронов», основано лишь на анализе пятого слоя новой коры. Действительно, пятый слой очень богат нейронами, но если учитывать содержание нейронов в шести уровнях, мы получим значение около 100 нейронов на модуль, что соответствует моим оценкам.

Последовательное возбуждение и модульное строение новой коры были известны уже на протяжении многих лет, но в данном исследовании впервые говорится о стабильности этих модулей при динамических процессах, происходящих в мозге.

Результаты другого нового исследования, проведенного в Массачусетском госпитале при финансовой поддержке Национального института здоровья, опубликованные в мартовском номере журнала *Science* за 2012 г., также демонстрируют регулярную структуру связей в новой коре<sup>5</sup>. В статье описано возбуждение новой коры в соответствии с некоей координатной сеткой, как на карте упорядоченных городских улиц: «В общем, структура головного мозга напоминает Манхэттен с двумерным планом улиц и лифтом, движущимся по третьей, вертикальной оси», — пишет руководитель данного исследования нейрофизиолог из Гарварда Ван Дж. Виден.

В электронной версии статьи Ван Виден рассказывает о значении данного исследования: «Это было исследование трехмерной структуры связей в головном мозге. На протяжении последних ста лет ученые представляли себе головной мозг в виде порции спагетти — набор отдельных путей, не имеющих выраженной пространственной связи. Используя метод магнитного резонанса, мы смогли провести экспериментальное исследование. И вместо независимых или связанных случайным образом путей мы обнаружили, что все пути в головном мозге образуют единую, чрезвычайно простую структуру. В общем

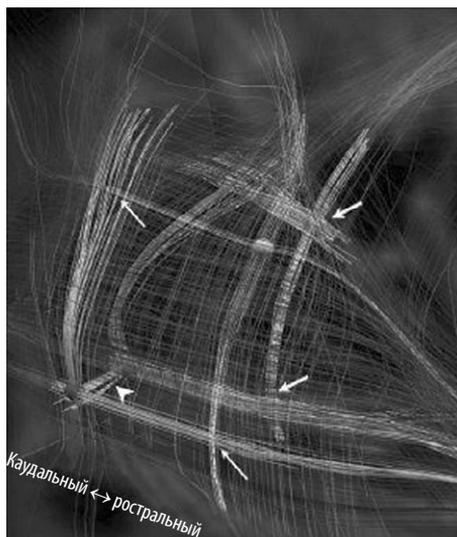
приближении они образуют кубик. Все пути идут в трех перпендикулярных направлениях, и на каждом из этих направлений пути организованы и строго параллельны. Таким образом, вместо несвязанных спагетти мы видим общую структуру мозговых связей как единую слаженную структуру».

Итак, исследования Маркрама демонстрируют наличие в новой коре повторяющихся модулей, состоящих из нейронов, а исследования Ван Видена — удивительную упорядоченность связей между этими модулями. Мозг представляет собой огромное множество «ожидающих связей», к которым могут подключаться распознающие модули. Таким образом, если один какой-то модуль хочет подключиться к другому, одному из них не нужно растить аксон, а другому дендрит, чтобы преодолеть физическое расстояние между ними. Они просто используют одну из аксональных ожидающих связей и подключаются к концу нити. Как пишут Ван Виден и его коллеги, «пути головного мозга следуют основному плану, сформированному... [в процессе] раннего эмбриогенеза. Таким образом, пути зрелого мозга отражают картину этих первозданных переходов, физически сформированных в ходе развития». Другими словами, когда мы учимся и приобретаем какой-либо опыт, распознающие модули новой коры подключаются к этим уже существующим связям, возникшим еще на стадии эмбрионального развития.

По такому же принципу устроен электронный чип, называемый «программируемой пользователем вентиляционной матрицей» (ППВМ, *англ.* Field-Programmable Gate Array, FPGA). Этот чип содержит миллионы модулей, использующих логическую функцию «ожидающих связей». При реализации определенной функции эти связи либо активируются, либо инактивируются (посредством электронных сигналов).

Те протяженные связи в новой коре, которые не применяются, в итоге отключаются, что отчасти объясняет, почему адаптация соседних участков новой коры для компенсации поврежденных участков не является столь же эффективной, как использование

исходного участка. В соответствии с результатами Ван Видена исходные связи в высшей степени упорядочены и повторяемы, как и сами модули, а их сеть применяется для «направления проведения [сигналов]» в новой коре. Такая картина обнаружена в головном мозге человека и всех приматов и распространяется на все зоны новой коры, начиная от участков, ответственных за распознавание ранних (наиболее простых) сенсорных образов, и заканчивая уровнем высших эмоциональных образов. В заключительной части статьи Ван Видена говорится, что «всеобъемлющая, единая и последовательная сетчатая структура церебральных путей соответствует трем основным осям развития». Это опять же говорит о едином алгоритме функционирования новой коры.



*Высокоупорядоченная структура решетки начальных связей нейронов новой коры (по данным исследований, проведенных Национальным институтом здоровья США).*

Уже давно известно, что как минимум некоторые отделы новой коры построены по иерархическому принципу. Лучшее исследование зрительный отдел коры, подразделяемый на зоны V1, V2 и MT (она же зона V5).



*Подпись: Еще одно изображение регулярных связей нейронов в новой коре.*

При перемещении в более высокие отделы этой области коры («высокие» в смысле понятийной иерархии, а не в физическом смысле, поскольку новая кора по толщине всегда соответствует лишь одному распознающему модулю) происходит распознавание все более сложных (более абстрактных) понятий.



*Структура решетки, обнаруженная в новой коре, удивительным образом напоминает структуру так называемой координатной коммутации, используемой в интегральных схемах и монтажных платах.*

Зона V1 отвечает за самые примитивные формы, V2 может распознавать контуры, соответствие изображений, полученных от каждого глаза, пространственную ориентацию, а также является ли данный образ частью какого-то объекта или фоном<sup>6</sup>. Наивысшие зоны коры распознают такие понятия, как идентичность объектов и лиц и их движения. Также давно известно, что связь между этими иерархическими зонами осуществляется как снизу вверх, так и сверху вниз и что сигналы могут быть как возбуждающими, так и тормозящими. Нейрофизиолог из Массачусетского технологического института Томазо Поджо (род. в 1947 г.) активно занимался изучением зрительных отделов коры, и полученные им за последние 35 лет результаты были очень полезны для понимания иерархических принципов обучения и распознавания образов в «ранней» (находящейся на нижнем понятийном уровне) зрительной области коры<sup>7</sup>.

Наши представления о нижних иерархических уровнях зрительной коры вполне соответствуют теории мысленного распознавания образов, о которой я говорил в предыдущей главе, а недавно удалось продемонстрировать иерархическую природу процессов далеко за пределами этих уровней новой коры. Профессор нейробиологии Техасского университета Даниель Дж. Феллман и его коллеги проследили «иерархическую организацию коры головного мозга...[в] 25 областях новой коры», в числе которых как зрительные зоны, так и участки высокого понятийного уровня, объединяющие образы, поступающие от нескольких органов чувств. Они обнаружили, что в соответствии с иерархическим строением коры обработка образов усложняется, затрагиваются все более обширные участки мозга и процесс занимает все более длительное время. И каждая проанализированная ими связь служит для проведения сигнала в обоих направлениях<sup>8</sup>.

Недавние исследования позволили нам значительно расширить эти наблюдения за пределы зрительной коры и связанных с ней областей, принимающих сигналы от различных органов

чувств. В 2007 г. профессор психологии Ури Хассон и его коллеги из Принстонского университета показали, что процессы, наблюдаемые в зрительной коре, имеют место и во многих других отделах коры: «Доказано, что нейроны зрительной коры имеют очень широкие рецептивные поля. В этом состоит основной принцип организации системы зрительного восприятия... Реальные события разворачиваются не только в пространственном, но и во временном измерении. Поэтому мы предполагаем, что иерархия, аналогичная иерархии пространственных рецептивных полей, должна существовать в различных отделах мозга и для реакций во временном измерении». Именно это они и обнаружили и пришли к заключению, что «существует иерархия прогрессивно увеличивающихся временных рецептивных полей, аналогичная известной иерархии пространственных рецептивных полей новой коры»<sup>9</sup>.

Самым мощным аргументом в пользу универсального характера обработки информации в новой коре является доказательство пластичности этого процесса (не только обучаемость, но и взаимозаменяемость). Иными словами, один отдел мозга может выполнять работу других отделов, пользуясь общим для всей новой коры алгоритмом. Многочисленные нейрофизиологические исследования были посвящены идентификации участков новой коры, ответственных за обработку различных типов образов. Классический подход к исследованию данного вопроса состоит в проведении корреляций между повреждениями головного мозга (в результате травмы или инсульта) и потерей соответствующей функции мозга. Например, если выясняется, что в результате повреждения веретенообразной извилины человек перестает распознавать лица, но все еще способен узнавать людей по голосам и манере говорить, можно сделать вывод, что поврежденный участок связан с распознаванием лиц. При этом предполагается, что каждый такой отдел мозга предназначен для распознавания и обработки образов определенного типа. В результате отдельные физические участки мозга стали связывать

с восприятием определенных типов образов, поскольку в норме именно в этих участках происходит обработка соответствующей информации. Однако выясняется, что при нарушении по какой-либо причине нормального потока информации эту функцию может взять на себя другой отдел новой коры.

Неврологам хорошо известна эта пластичность мозга, благодаря которой пациенты с повреждениями мозга в результате травмы или инсульта могут вернуть утерянные функции за счет других участков новой коры. Возможно, один из самых удивительных примеров пластичности мозга был продемонстрирован в 2011 г. американским нейрофизиологом Мариной Бедни и ее коллегами, которые изучали зрительную кору пациентов с врожденной слепотой. В соответствии с классическими представлениями, первичные уровни зрительной коры, такие как V1 и V2, воспринимают и обрабатывают самые примитивные образы (углы и линии), тогда как фронтальная кора (новый в эволюционном плане отдел мозга, заключенный в нашей необычайно развитой лобной части черепа) обрабатывает гораздо более сложные и тонкие речевые образы и другие абстрактные понятия. Однако Бедни и ее коллеги обнаружили следующее: «Считается, что левый лобный и височный отделы человеческого головного мозга предназначены для обработки речевых образов. Однако у слепых от рождения людей при решении некоторых речевых задач происходит активация зрительной коры. Мы показали, что эта активность зрительной коры на самом деле отражает обработку речевых образов. Мы обнаружили, что у людей с врожденной слепотой левая часть зрительной коры ведет себя аналогично классическим участкам, отвечающим за восприятие речи... Мы пришли к заключению, что отделы мозга, которые, как считалось, эволюционировали для восприятия зрительных образов, в результате раннего опыта могут участвовать в обработке речевых образов»<sup>10</sup>.

Подумайте, что это означает. Получается, отделы новой коры, которые физически удалены друг от друга и, как считалось,

отвечают за совершенно разные функции (восприятие примитивных зрительных образов и сложных речевых построений), фактически используют один и тот же алгоритм. Эти отделы, предназначенные для восприятия совершенно разных типов образов, могут заменять друг друга.

Нейрофизиолог Даниэль Фельдман из Университета Калифорнии в Беркли в 2009 г. опубликовал обширный обзор о том, что мы называем «механизмами синаптической пластичности новой коры», и привел доказательства подобной пластичности в разных отделах коры. Он пишет, что «пластичность позволяет мозгу узнавать и запоминать образы чувственного мира, совершенствовать движения... и восстанавливать функции после повреждений». Он добавляет, что эта пластичность достигается за счет «структурных изменений, включающих образование, удаление и морфологические перестройки кортикальных синапсов и дендритных шипиков»<sup>11</sup>.

Ученые из Университета Калифорнии в Беркли недавно продемонстрировали еще один удивительный пример пластичности коры (и, следовательно, универсальности алгоритма функции новой коры). Они использовали имплантированные микроэлектроды для получения сигналов из участков моторной коры мыши, которые контролируют движение усов. Эксперимент был поставлен таким образом, что мыши получали вознаграждение, если активировали соответствующие нейроны в определенных ситуациях, но при этом не шевелили усами. Для получения награды мышам приходилось решать задачу, в обычных условиях не связанную с активацией нейронов лобной доли. Тем не менее мыши оказались способны решить задачу, в основном используя свои моторные нейроны, при этом отключив их от выполнения двигательной функции<sup>12</sup>. Это означает, что моторная кора, ответственная за координацию движений мышц, тоже использует стандартный алгоритм новой коры.

Однако по некоторым причинам опыт или знания, полученные с помощью новой области коры, заменившей поврежденную

область, не всегда столь же полноценны, как исходные. В-первых, обучение какому-то навыку и его совершенствование продолжаются всю жизнь, поэтому переобучение за счет подключения другого участка новой коры не может немедленно привести к тем же результатам. Но еще важнее, что новый участок коры не просто находился в ожидании этой своей новой функции. Он участвовал в реализации каких-то других жизненно важных задач и поэтому вынужден отказаться от какой-то содержащейся в нем информации, чтобы компенсировать функцию поврежденного участка. Прежде всего, он уничтожает некоторые избыточные копии хранящихся в нем образов, что в какой-то степени ухудшает его существующие способности, но при этом не освобождает столько же пространства, сколько раньше занимали выучиваемые заново образы.

Существует и третья причина ограниченности пластичности. Поскольку у большинства людей образы определенного типа воспринимаются специфическими участками коры (например, лица — веретенообразной извилиной), биологическая эволюция привела к оптимизации восприятия этих образов данными участками. В седьмой главе мы поговорим о том, что такая же ситуация имеет место и в цифровой новой коре. Мы можем распознавать речь с помощью системы распознавания знаков и наоборот, но все-таки система распознавания речи оптимизирована для восприятия речи, а система распознавания знаков оптимизирована для распознавания печатных символов, так что при замене одной системы другой эффективность распознавания уменьшается. Для оптимизации мы используем эволюционные (генетические) алгоритмы, имитирующие нормальный биологический процесс. Учитывая, что большинство людей на протяжении сотен тысяч лет распознаёт лица с помощью веретенообразной извилины, биологическая эволюция усовершенствовала возможности обработки таких образов именно в этом отделе мозга. Здесь работает тот же самый базовый алгоритм, но он настроен на восприятие лиц. Как пишет голландский

невролог Рэндал Коэн, «[новая] кора очень однородна, каждая колонка или мини-колонка может, в принципе, делать то же, что делает любая другая»<sup>13</sup>.

Недавние исследования подтверждают наблюдение, что распознающие модули связываются в зависимости от того, какие образы им приходится распознавать. Например, нейробиолог Йи Зуо и ее коллеги следили за тем, как новые дендритные шипики образовывали связи между нервными клетками по мере того, как мыши обучались новому навыку (пролезать в щель, чтобы достать корм)<sup>14</sup>. Исследователи из Института Солка в Калифорнии выяснили, что это самовозбуждение модулей новой коры, по-видимому, контролируется лишь несколькими генами. И эти гены, и сам способ самовозбуждения тоже одинаковы во всей новой коре<sup>15</sup>.

Эти свойства новой коры описаны и во многих других исследованиях, так что давайте суммируем то, что мы узнали из литературы и из наших собственных мысленных экспериментов. Основная функциональная единица новой коры — ансамбль (модуль) нейронов, которых, по моим оценкам, в каждом модуле содержится около сотни. Модули сплетены между собой и образуют колонки, так что отдельные модули практически неразличимы. Направление связей и сила синаптических взаимодействий в каждом модуле достаточно стабильны. Обучение происходит при изменении направления и силы взаимодействий *между* модулями.

В новой коре насчитывается порядка квадрильона ( $10^{15}$ ) связей, но при этом в геноме содержится лишь около 25 млн байт соответствующей структурной информации (после сжатия без потерь), так что эти связи не могут быть предопределены генетически. Возможно, до некоторой степени обучение является результатом «общения» новой коры со старым мозгом, но и это позволяет объяснить происхождение лишь части информации. В целом связи между модулями формируются на основании опыта («воспитание сильнее природы»).

Головной мозг не обладает достаточной гибкостью для того, чтобы каждый распознающий модуль мог просто связываться с любым другим модулем (в отличие от легко программируемого компьютера или Интернета), — требуется осуществить реальное физическое соединение с участием аксона и дендрита. Как показало исследование Ван Видена, эти связи организованы повторяющимся и в высшей степени упорядоченным образом. Окончательное связывание с ожидающими аксонами происходит в результате распознавания образов каждым соответствующим распознающим модулем. Неиспользуемые связи в итоге удаляются. Все эти связи организованы иерархическим образом, что отражает естественное иерархическое устройство реальности. В этом и заключается ключевая способность новой коры.

Основной алгоритм действия распознающих модулей одинаков как для модулей «низшего порядка», имеющих дело с самыми примитивными сенсорными образами, так и для модулей «высшего порядка», распознающих более сложные понятия. Многочисленные доказательства пластичности и взаимозаменяемости разных отделов новой коры являются лучшим подтверждением этого важного наблюдения. Участки коры в некоторой степени оптимизированы для обработки образов определенного типа, но это вторичное явление; основной алгоритм универсален.

Передача сигналов по иерархической лестнице происходит как снизу вверх, так и сверху вниз. Восходящий сигнал означает: «Я обнаружил образ». Нисходящий сигнал гласит: «Я ожидаю ваш образ» и имеет предсказательное значение. И восходящий, и нисходящий сигналы могут быть как возбуждающими, так и ингибирующими.

Каждый образ воспринимается в определенном порядке, и обратить его непросто. Даже если образ многомерный, он представляется в виде последовательности одномерных образов более низкого порядка. Образ — это упорядоченная последовательность других образов, так что процесс распознавания

по своей природе является рекурсивным. Эта иерархия может иметь множество уровней.

Образы, которые мы учимся распознавать, особенно наиболее важные из них, отличаются высокой степенью избыточности. Распознавание образов (таких как обычные предметы и лица) происходит по тому же механизму, что и воспоминание, которое как раз и представляет собой воспроизведение уже знакомого образа. Воспоминания тоже хранятся в виде последовательностей образов, обычно в виде текстов. Тот же механизм используется и для обучения движениям и их воспроизведению. Избыточность образов позволяет нам распознавать предметы, людей и идеи даже в измененном виде и в разном контексте. Величина и вариабельность величины поступающих сигналов также позволяют новой коре кодировать изменения амплитуды в разных измерениях (длительность в случае звука). Один из способов кодирования этих параметров состоит в записи множества образов с различающимися входными сигналами. Например, это могут быть образы произнесенного слова *steep* с разными вариантами длительности звука [E], каждый из которых характеризуется параметрами значимости, указывающими, что длительность этого [E] может быть различной. Этот подход не является математическим эквивалентом точному указанию параметров величины сигнала и на практике реализуется гораздо хуже, тем не менее это одна из возможностей записи амплитуды параметра. Самое весомое доказательство важности данных параметров заключается в том, что без них системы искусственного интеллекта не могут достичь такой же эффективности распознавания, как живой человек.

Все эти выводы следуют из приведенных мной выше примеров исследований и мысленных экспериментов. Я повторяю, что представленная мной модель является единственной, которая удовлетворительным образом соответствует всем условиям, определяемым данными практическими и мысленными экспериментами.

## Глава четвертая

Наконец, существует еще одно доказательство справедливости описанной модели. Созданные нами за последние десятилетия системы искусственного интеллекта для распознавания и обработки проявлений реальной жизни (таких как человеческая устная и письменная речь), а также для понимания природы языка, как выясняется, в математическом плане очень похожи на описанную мной модель. Они также являются примером теории мысленного распознавания образов. Создатели искусственного интеллекта не пытаются воспроизвести человеческий мозг в буквальном смысле, однако неизбежно приходят к тем же самым принципам.

# Глава пятая

## СТАРЫЙ МОЗГ

У меня старый мозг, но великолепная память.

*Искусственный интеллект Льюис*

И теперь мы стоим посреди этого нового мира с нашим примитивным мозгом, настроенным на простую пещерную жизнь, и имеем в своем распоряжении невероятные возможности, которые вполне способны реализовать, однако не можем предугадать последствий.

*Альберт Сент-Дьердьи\**

---

\* Альберт Сент-Дьердьи (1893–1986) — американский биохимик венгерского происхождения, лауреат Нобелевской премии «За исследования биологического окисления».

**С**тарый мозг — тот, который имели наши предки еще до того, как стали млекопитающими, — никуда не исчез. Он по-прежнему решает многие задачи, такие как стремление к удовольствию и уход от опасности. Однако эти задачи модулируются новой корой, которая у человека является доминирующей частью мозга как по массе, так и по активности.

Животные жили и выживали, не имея новой коры, причем не млекопитающие и сегодня без нее обходятся. Мы можем воспринимать новую кору в качестве мощного сублиматора: наше примитивное желание избежать встречи с опасным хищником сегодня трансформируется новой корой в стремление решить поставленную задачу так, чтобы угодить начальству. Охота, возможно, подменяется написанием книг (например, о мозге), а задача воспроизводства — стремлением к популярности или страстью к оформлению домашнего интерьера (впрочем, последнее обычно не скрывают).

Новая кора способна помочь нам решить эти проблемы по той причине, что прекрасно моделирует окружающий мир, отражая его истинно иерархическую природу. Но задачи перед нами ставит старый мозг. Конечно, как любой умный чиновник, новая кора часто переформулирует проблемы, прежде чем их решать. Учитывая это, давайте все же рассмотрим процессы обработки информации, происходящие в старом мозге.

## **Сенсорное восприятие**

Эти изображения, распространяясь при помощи движения вдоль волокон оптических нервов в мозг, являются причиной зрения\*.

*Исаак Ньютон (1642–1727)*

Каждый из нас живет во вселенной — или в тюрьме — собственного мозга. От него отходят миллионы хрупких чувствительных нервных волокон, которые образуют группы, уникальным образом

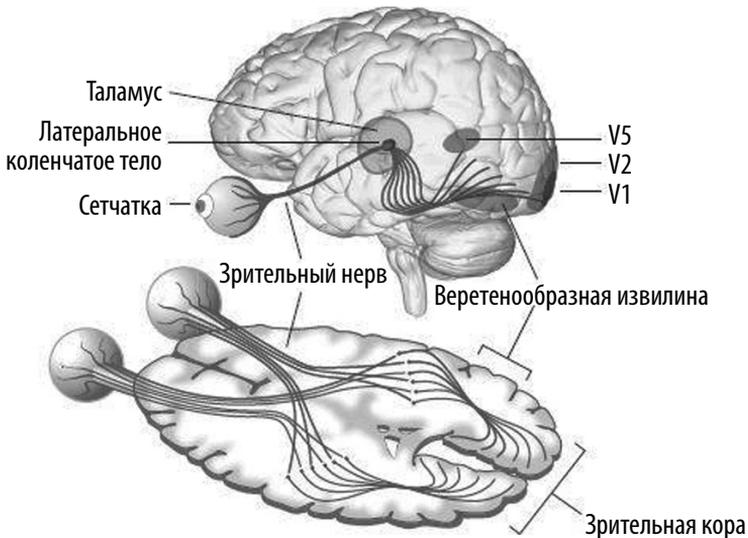
---

\* *Ньютон И.* Оптика, или Трактат об отражениях, преломлениях, изгибаниях и цветах света. М. : Техтеоргиз, 1954.

адаптированные для определения энергетического состояния окружающего нас мира: тепла, света, силы и химического состава. Только это нам и известно из непосредственного опыта; все остальное — логические измышления.

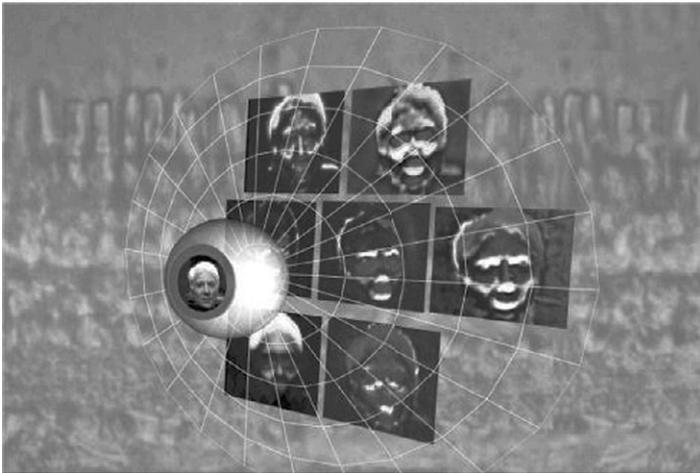
Вернон Б. Маунткасл<sup>1</sup>

Хотя нам кажется, что глаза воспринимают изображение с высоким разрешением, на самом деле зрительный нерв посылает в мозг лишь набор очертаний и указаний относительно объектов, попадающих в поле нашего зрения. А затем мы воображаем себе окружающий мир на основании воспоминаний, хранящихся в новой коре, которая с очень низкой скоростью интерпретирует серию образов, поступающих по параллельным каналам. Профессор молекулярной и клеточной биологии из Калифорнийского университета в Беркли Фрэнк С. Верблин и докторант Ботон Роска опубликовали в журнале *Nature* статью, где показали, что зрительный нерв содержит от 10 до 12 выходных каналов, через каждый из которых проходит лишь небольшое количество информации о конкретной ситуации<sup>2</sup>.



Зрительная система человека

Так называемые ганглионарные клетки посылают информацию лишь о контурах предметов (контраст). Другая группа клеток детектирует только протяженные области одинакового цвета, а третья группа воспринимает исключительно фон за интересующим нас предметом.

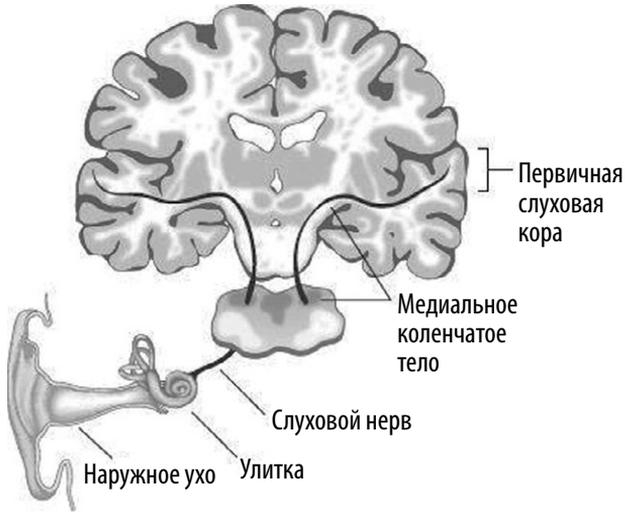


*Семь из двенадцати изображений низкого порядка, которые зрительный нерв направляет в мозг*

«Хотя нам кажется, что мы видим мир во всем объеме, на самом деле мы получаем лишь подсказки, контуры в пространстве и времени, — говорит Верблин. — Эти 12 изображений мира составляют всю информацию о том, что нас окружает, и на основании этих 12 изображений, что чрезвычайно мало, мы воссоздаем все богатство видимого мира. Мне любопытно, почему природа отобрала эти 12 простых картинок и как их может быть достаточно, чтобы снабдить нас всей информацией, в которой мы нуждаемся».

Такой способ передачи информации в области искусственного интеллекта называют разреженным кодированием. При создании искусственных систем мы обнаружили, что наилучшие результаты получаются, если отбросить большую часть

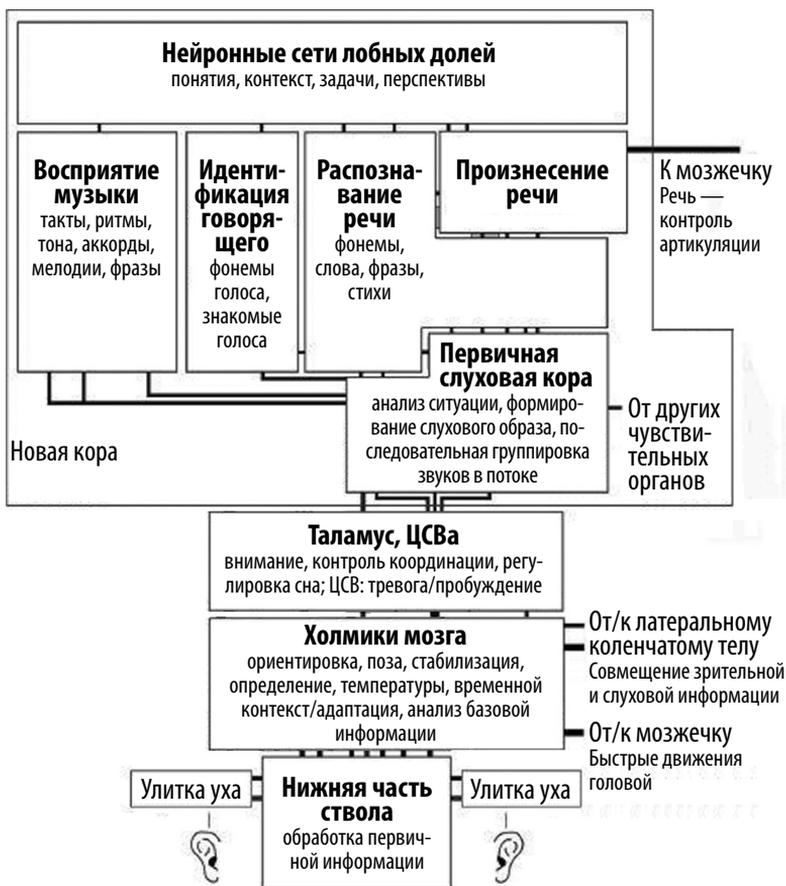
поступающей информации и оставить лишь самые заметные детали. В противном случае новая кора (биологическая или искусственная) не справляется с обработкой информации.



*Слуховая система человека*

Ллойд Уатт и его исследовательская группа из компании Audience Inc.<sup>3</sup> тщательно смоделировали процессы обработки звуковой информации, поступающей от улитки человеческого внутреннего уха через подкорку в отделы новой коры. Ученые разработали технологию, позволяющую разложить звук на 600 частотных полос (60 на октаву). Это уже близко к значению 3000 полос — именно так оценивается способность человеческого уха (для сравнения: коммерческие звукораспознающие системы используют от 16 до 32 полос). Применяя два микрофона, а также детальную (и обладающую высоким спектральным разрешением) модель обработки звуковой информации, ученые создали коммерческую технологию (с чуть более низким спектральным разрешением, чем в лабораторной модели), которая эффективно удаляет фоновые шумы при разговоре. Эта система теперь широко применяется во многих мобильных телефонах и является

впечатляющим примером коммерческого продукта, основанного на понимании способности человеческого слухового аппарата фокусироваться на одном конкретном источнике звука.



Упрощенная схема обработки слуховой информации в подкорке и новой коре (модель AUudience, Inc., рисунок (с модификациями) взят из статьи L. Watts, Reverse-Engineering the Human Auditory Pathway, in J.Liu et al. (eds.), WCCI 2012; Berlin: Springer-Verlag, 2012, p.49)

Входные сигналы от собственного тела человека (по оценкам — сотни мегабит в секунду), включая сигналы от нервных клеток кожи, мышц, внутренних органов и других зон,

направляются в верхнюю часть спинного мозга. Это не только тактильная информация, но и данные о температуре, уровне кислотности (например, о содержании молочной кислоты в мышцах), перемещении пищи по пищеварительному тракту и о многом другом. Эта информация обрабатывается стволом мозга и средним мозгом. Важнейшие клетки, называемые нейронами первого слоя, создают карту тела, отражающую его текущее состояние; ее можно сравнить с дисплеем, на котором авиадиспетчеры отслеживают движение самолетов. Отсюда информация поступает в загадочный отдел мозга, называемый таламусом. И это тема нашего следующего раздела.

## Таламус

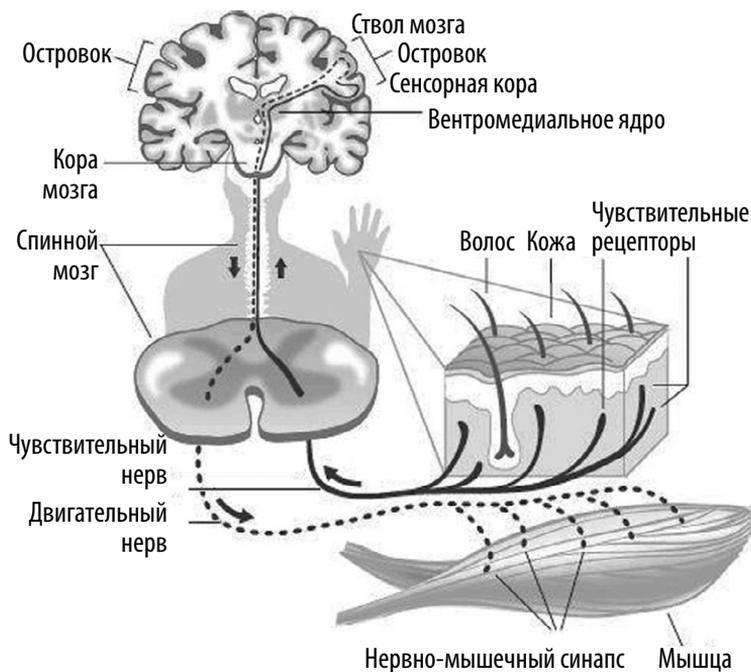
Каждый знает, что такое внимание. Это сосредоточение мыслительного процесса, в четкой и яркой форме, на одном из нескольких одновременно существующих объектов или цепочек мыслей. Ключевые элементы этого процесса — локализация, концентрация и осознание. Процесс состоит в удалении одних вещей для более эффективного обдумывания других...

*Уильям Джемс*

От среднего мозга сенсорная информация далее проходит через заднее вентромедиальное ядро таламуса (ВМЯ; участок мозга размером с лесной орех), в котором формируются сложные реакции организма, такие как «какой ужасный вкус», «что за вонь» или «это прикосновение возбуждает». Постепенно обработанная информация собирается в двух отделах новой коры, называемых островками мозга. Эти структуры размером с небольшой палец локализованы на правой и левой стороне коры. Артур Крейг из Неврологического института Барроу в Финиксе (Аризона) определяет ВМЯ и островки как «систему, представляющую меня в материальном мире»<sup>4</sup>.

Таламус выполняет множество функций, в частности считается местом, откуда предварительно обработанная сенсорная

информация направляется в новую кору. Кроме тактильной информации, поступающей через ВМЯ, обработанная информация от зрительного нерва (которая, как было сказано выше, уже подверглась существенной трансформации) направляется в участок таламуса, называемый латеральным коленчатым ядром, откуда поступает в зону VI новой коры. Звуковая информация идет через медиальное коленчатое ядро таламуса, а затем достигает первичной слуховой коры. Вся сенсорная информация (возможно, за исключением обонятельной информации, которая передается через обонятельные луковицы) проходит через специфические участки таламуса.



*Проведение тактильного сигнала в теле человека*

Однако наиболее важная функция таламуса заключается в непрерывном контакте с новой корой. Распознающие модули новой коры посылают в таламус предварительные данные

и получают ответы главным образом в виде возбуждающих и ингибирующих сигналов от шестого уровня каждого модуля. Мы помним, что это не беспроводная система коммуникаций и между всеми участками коры и таламусом действительно возбуждается огромное число аксонов. Если учесть широту поля деятельности (в виде физической массы необходимых связей), то становится понятно, что таламус постоянно проверяет информацию от сотен миллионов распознающих модулей новой коры<sup>5</sup>.

О чем же эти сотни миллионов модулей сообщают таламусу? Судя по всему, это важная информация, поскольку серьезные двусторонние повреждения основных участков таламуса могут приводить к длительной потере сознания. У человека с повреждением таламуса новая кора по-прежнему может функционировать, и самовозбуждающаяся система ассоциативного мышления продолжает работать. Но направленное мышление (заставляющее нас вылезать из постели, садиться в машину или за рабочий стол) без таламуса не работает. Известен случай 21-летней Карен Энн Куинлан, которая после перенесенного сердечного приступа и остановки дыхания на протяжении десяти лет находилась в вегетативном состоянии, не реагируя на окружающий мир. Произведенное после смерти вскрытие показало, что ее новая кора работала нормально, но таламус был разрушен.

Ключевая роль таламуса в сосредоточении внимания связана с использованием данных, структурированных в новой коре. Таламус применяет хранящийся в новой коре список, позволяющий нам думать в определенном направлении или следовать определенному плану действий. Как показали исследования нейробиологов из Института обучения и памяти имени Пиковеера при Массачусетском технологическом институте, наша рабочая память способна одновременно удерживать четыре вопроса — по два в каждой полусфере мозга. Пока еще неясно, руководит ли таламус новой корой или

наоборот, но для нормального функционирования нужны оба отдела.

### **Гиппокамп**

В каждом полушарии мозга содержится гиппокамп — маленькая структура в форме подковы, расположенная в медиальных височных отделах мозга. Его основная функция состоит в запоминании новых событий. Поскольку сенсорная информация проходит через новую кору, именно она решает, что данная информация является новой и ее нужно представить гиппокампу. Это происходит в том случае, если новая кора не может распознать определенный набор черт (например, новое лицо) или если уже знакомая ситуация приобрела новые характеристики (например, если на лице вашей жены появились накладные усы).

Гиппокамп запоминает эти ситуации, хотя, по-видимому, делает это с помощью отсылок к новой коре. Таким образом, сохраняемые в гиппокампе воспоминания также фиксируются в новой коре в виде образов низкого порядка. У животных, не имеющих новой коры, при модуляции чувственного опыта гиппокамп просто запоминает поступающую сенсорную информацию, которая, однако, подвергается предварительной обработке (например, в зрительном нерве).

Хотя гиппокамп использует новую кору (если она есть) в качестве сверхоперативной памяти, его собственная память (отсылки к новой коре) не имеет иерархического строения. Таким образом, животные, не имеющие новой коры, также могут запоминать информацию с помощью гиппокампа, но их воспоминания не являются иерархическими.

Емкость гиппокампа ограничена, поэтому хранящиеся в нем воспоминания недолговечны. Гиппокамп отправляет определенные последовательности образов на долгосрочное хранение, вновь и вновь переправляя их в новую кору. Таким образом, гиппокамп нужен нам для запоминания новых данных и нового

опыта (хотя исключительно моторные функции, по-видимому, усваиваются другим путем). Человек с двусторонними повреждениями гиппокампа сохранит имеющиеся воспоминания, но не сможет усваивать новый материал.

Нейробиолог Теодор Бергер и его коллеги из Университета Южной Калифорнии создали модель гиппокампа крысы и успешно провели эксперименты по имплантации искусственного гиппокампа. В статье, опубликованной в 2011 г., они описали блокирование определенных навыков крыс с помощью лекарственных препаратов. Используя искусственный гиппокамп, крысы быстро вновь обучались утерянным навыкам. Описывая возможность контроля имплантированного органа, Бергер писал: «Когда переключатель включен, крысы вспоминают; когда он выключен, они забывают». В другом эксперименте искусственный гиппокамп работал одновременно с настоящим. В результате повысилась способность крыс воспринимать новую информацию. Бергер писал: «Эти комплексные экспериментальные модели впервые показали, что... нейронные имплантаты способны в реальном времени идентифицировать процесс кодирования и манипулировать им, восстанавливая и даже улучшая когнитивные мнемонические процессы»<sup>7</sup>. Гиппокамп — из тех отделов мозга, которые в первую очередь повреждаются при болезни Альцгеймера, поэтому одна из целей данного исследования заключается в создании нейронных имплантатов, которые могли бы сглаживать первую фазу повреждений в ходе развития заболевания.

## **Мозжечок**

Есть два способа поймать летящий мяч. Можно одновременно решить комплекс уравнений, описывающих поведение мяча, наклон вашего тела в соответствии с положением мяча и движение вашего тела, плеча и руки в пространстве и во времени.

Головной мозг использует другой подход. Он упрощает проблему, сводя множество уравнений к простой эксперименталь-

ной модели, исследуя лишь пространство, в котором мяч оказывается в поле нашего зрения, и скорость его перемещения. Тем же способом он контролирует и движение руки, делая в основном линейные предсказания относительно положения мяча и руки. Цель заключается в том, чтобы рука и мяч встретились в какое-то время в каком-то месте. Если кажется, что мяч движется слишком быстро, а рука чересчур медленно, мозг заставит руку двигаться быстрее навстречу мячу. Такое разрешение Гордиева узла сложных математических проблем называют «основными функциями» мозга, и происходит этот процесс в мозжечке — области мозга размером с кулак, расположенной в основании ствола мозга<sup>8</sup>.

Мозжечок — отдел старого мозга, который когда-то контролировал практически все движения гоминидов. Он по-прежнему содержит половину всех нейронов мозга, но поскольку основная часть из них — сравнительно небольшие нейроны, на мозжечок приходится лишь около 10% всей массы мозга. Это еще один отдел, характеризующийся повторяемостью своих структур. Мы мало знаем о том, как строение мозжечка закодировано в геноме, поскольку он представляет собой сочетание нескольких нейронов, повторяющихся миллиарды раз. Подобно новой коре, мозжечок имеет равномерную структуру<sup>9</sup>.

Со временем новая кора взяла под свой контроль большую часть нашей мышечной деятельности, используя тот же алгоритм распознавания образов, который она применяет для узнавания и обучения. Функцию новой коры при выполнении движений мы вправе называть реализацией образов. Новая кора использует хранящиеся в мозжечке воспоминания для записи тонких набросков движений, таких как подпись или характерные элементы художественной манеры в музыке или танце. Изучение роли мозжечка в обучении детей письму показало, что клетки Пуркиньи в мозжечке контролируют последовательность движений, причем каждая клетка чувствительна к определенной последовательности<sup>10</sup>. Поскольку большая часть наших

движений теперь контролируется новой корой, даже довольно серьезные повреждения мозжечка приводят лишь к незначительной неловкости или неточности движений.

Новая кора также может обращаться к мозжечку за помощью для расчета основных функций в реальном времени, чтобы предсказать результат действий, которые мы никогда не делали, но собираемся совершить, а также реальных или возможных действий других людей. Это еще один пример врожденной предсказательной способности мозга.

Ученые достигли значительных успехов в моделировании поведения мозжечка, а именно его способности динамическим образом реагировать на сенсорные стимулы, используя те основные функции, о которых я говорил выше, как в восходящих (биохимических) моделях, так и в нисходящих моделях, основанных на математических расчетах функции всех повторяющихся единиц мозжечка<sup>11</sup>.

## **Удовольствие и страх**

Страх — главный источник суеверий и один из основных источников жестокости. Преодоление страха — начало премудрости.

*Бертран Рассел*

Бойся... но действуй!

*Сьюзен Джефферс\**

Итак, новая кора хороша для решения различных проблем, но что же является нашей главной проблемой? Проблема, которую всегда приходилось решать в ходе эволюции, заключается в выживании вида. А выживание вида подразумевает выживание каждого из нас, и каждый из нас использует свою новую

---

\* Сьюзен Джефферс — современный американский психолог, автор нескольких бестселлеров; ее книга «Бойся... но действуй!» на русском языке вышла в издательстве «София» в 2008 г.

кору, чтобы интерпретировать эту задачу миллиардами разных способов. Животным для выживания нужно добывать себе пищу и при этом не стать пищей для кого-нибудь другого. А еще им необходимо воспроизводиться. Уже на самых ранних этапах эволюции мозга возникли ощущения удовольствия и страха, которые способствовали реализации этих важнейших целей и сопутствующих им задач. Постепенное изменение среды и конкурирующих видов приводило к соответствующим модификациям. С появлением иерархического мышления удовлетворение насущных потребностей стало более сложным процессом, поскольку должно было соответствовать широкому кругу идей внутри других идей. Однако, несмотря на значительное влияние новой коры, старый мозг все еще жив и по-прежнему использует такие мотивации, как удовольствие и страх.

С ощущением удовольствия связан участок мозга, называемый прилежащим ядром. В знаменитых экспериментах 1950-х гг. было показано, что крысы, способные напрямую стимулировать этот небольшой участок мозга (нажимая на рычаг, активировавший имплантированные электроды), предпочитали это занятие всем другим, включая секс и еду, доводя себя до истощения и смерти<sup>12</sup>. У человека ощущение удовольствия связано и с другими отделами мозга, такими как вентральный паллидум (бледный шар) и, конечно же, сама новая кора.

Удовольствие регулируется такими химическими веществами, как дофамин и серотонин. Я не буду подробно рассказывать об их действии, скажу только, что мы унаследовали их от наших древнейших предков, существовавших еще до появления млекопитающих. Задача новой коры заключается в том, чтобы сделать нас хозяевами страха и удовольствия, а не их рабами. Однако, учитывая широкое распространение различных зависимостей, приходится признать, что новая кора не всегда справляется с этой задачей. В частности, дофамин является нейромедиатором, участвующим в реакции удовольствия. Когда мы переживаем приятный момент — выигрываем в лотерею, получаем

одобрение коллег, обнимаем любимого человека или просто заставляем друга улыбнуться новой шутке, — у нас в организме происходит выделение дофамина. Иногда мы, как те крысы, что умирали от голода, стимулируя свои центры удовольствия, тоже стремимся прийти к удовольствию по короткому пути, что далеко не всегда хорошо.

Например, азартные игры могут стимулировать выброс дофамина, особенно при выигрыше, который, в свою очередь, зависит от удачи. В принципе, азартные игры могут служить для повышения уровня дофамина, но, учитывая, что шансы исходно против вас (иначе модель игорного бизнеса не могла бы работать), регулярная игра вас разорит. Все другие зависимости чреваты аналогичными опасностями. Одна врожденная мутация гена дофамина D2-рецептора связана с очень сильным ощущением удовольствия при первом контакте с вызывающим привыкание веществом или опытом, однако, как многие знают (но не всегда помнят), способность этих веществ вызывать удовольствие при продолжительном использовании постепенно ослабевает. Другая генетическая мутация приводит к тому, что в организме человека не происходит нормального выброса дофамина, что также может привести к поиску других источников удовольствия и развитию зависимостей. Эта незначительная часть населения с подобными генетическими нарушениями создает огромную социальную и медицинскую проблему для общества в целом. Даже те, кто избегает серьезной зависимости, вынуждены балансировать между дофаминовым вознаграждением и желанием избежать его негативных последствий.

Серотонин — еще один нейромедиатор, который играет важнейшую роль в регуляции настроения. При высоком уровне серотонина человек ощущает моральный подъем и удовлетворение. Но серотонин выполняет и другие функции, включая модуляцию синаптического напряжения, аппетита, сна, сексуального влечения и пищеварения. Такие антидепрессанты, как селективные ингибиторы обратного захвата серотонина (которые

повышают уровень доступного для рецепторов серотонина), порой оказывают серьезное побочное действие (например, подавление либидо). В отличие от нормальных процессов в новой коре, в ходе которых распознавание образов и активация аксонов влияют лишь на небольшую часть нейронных цепей новой коры, эти вещества воздействуют на обширные области мозга или даже на всю нервную систему в целом.

Каждое полушарие мозга человека содержит миндалину (миндалевидное тело), состоящую из нескольких небольших долей. Это еще одна часть старого мозга, принимающая участие в различных эмоциональных реакциях, главной из которых является страх. У животных, стоящих на эволюционной лестнице ниже млекопитающих, некоторые предпрограммированные сигналы опасности поступают непосредственно на миндалины, запускающие реакции типа «борьба или бегство». У человека функция миндалин зависит от ощущения опасности, которое должно быть передано новой корой. Например, неудовольствие вашего начальника может стартовать подобную реакцию, вызывая у вас страх возможной потери работы (а может быть, и нет, если у вас есть запасной вариант). Если миндалины «решают», что вам грозит опасность, реализуется древний механизм: миндалины дают сигнал гипофизу выделить адренкортикотропный гормон, который, в свою очередь, стимулирует высвобождение гормона стресса кортизола из надпочечников, что приводит к приливу энергии к мышцам и нервной системе. Надпочечники также производят адреналин и норадреналин, которые подавляют функцию пищеварительной, иммунной и репродуктивной системы (поскольку в экстренной ситуации другие приоритеты). Повышается кровяное давление, а также уровень сахара, холестерина и фибриногена (который способствует свертыванию крови) в крови. Учащаются дыхание и сердцебиение. Даже зрачки расширяются и повышается острота зрения — так легко обнаружить врага или найти путь к отступлению. Все это чрезвычайно полезно, если вы сталкиваетесь с реальной опасностью,

например с крупным хищником. Однако хорошо известно, что в современном мире частая активация механизма «борьба или бегство» может приводить к хроническим проблемам со здоровьем в форме повышенного артериального давления, высокого уровня холестерина и т. д.

Система регуляции уровня таких нейромедиаторов, как серотонин, и таких гормонов, как дофамин, весьма сложная и запутанная, и мы могли бы посвятить ей всю оставшуюся часть книги (замечу, что этой теме посвящено множество книг), но следует заметить, что скорость обработки информации в этой системе гораздо ниже, чем в новой коре. В функционировании этой системы участвует лишь несколько веществ, а их концентрация в головном мозге изменяется очень медленно и достаточно равномерно, тогда как новая кора состоит из сотен триллионов быстро изменяющихся связей.

Нужно сказать, что наши эмоциональные реакции происходят и в старой, и в новой части мозга. Мыслительные процессы осуществляются в новой коре, но чувства — продукт действия обеих частей мозга. Поэтому любая эмуляция человеческого поведения требует моделирования функции обеих частей. Но если мы хотим исследовать лишь познавательные способности человека, нам достаточно рассмотреть только новую кору. В таком случае мы можем заменить старый мозг прямой мотивацией небиологической новой коры. Например, в случае с Ватсоном была сформулирована задача: получить правильные ответы на вопросы «Джеопарди!» (однако для выбора стратегии использовалась другая программа, которая понимала суть игры). Задача новой версии Ватсона, созданной для медицинских целей совместными усилиями Nuance и IBM, заключается в том, чтобы помочь лечить людей. Перед будущими системами могут быть поставлены такие задачи, как полностью искоренить болезни или бороться с бедностью. В значительной мере система таких функций, как удовольствие и страх, для человека уже устарела, поскольку старый мозг эволюционировал задолго до того, как

## Глава пятая

возникло даже самое примитивное человеческое общество. Во многом наш старый мозг аналогичен мозгу рептилий.

Между новым и старым мозгом человека идет постоянная борьба. Старый мозг пытается установить свой порядок, оперируя в рамках системы удовольствия и страха, тогда как новая кора пытается понять сравнительно простые алгоритмы старого мозга и научиться манипулировать ими для собственных нужд. Вспомним, что миндалины самостоятельно не могут оценить опасность — в мозге человека они опираются на сигналы новой коры. Этот человек друг или враг? Сторонник или предатель? Решить это может только новая кора.

Поскольку мы уже не участвуем в смертельных схватках с врагами и не охотимся, чтобы прокормиться, наша древнейшая мотивация, по крайней мере отчасти, уступила место творческим задачам. В следующей главе мы как раз и обратимся к теме любви и творчества.

# Глава шестая

## ТРАНСЦЕНДЕНТНЫЕ СПОСОБНОСТИ

Знаешь, я не могу спать, я не могу перестать думать.

*Битлз. I'm So Tired*

Вот моя простая религия. Нет нужды в храмах, нет нужды в сложной философии. Наш мозг и наше сердце — вот наш храм. Философия — в доброте.

*Далай-лама*

Моя рука движется, потому что ее приводят в действие определенные усилия — электрические, магнетические и любые другие, которые могут быть созданы «нервной энергией», — порожденные моим мозгом. Эта нервная энергия, порождаемая мозгом, со временем, когда наука достигнет более высокого развития, может быть сведена к действию химических веществ, поступающих в мозг вместе с кровью и зависящих исключительно от пищи, которой я питаюсь, и воздуха, которым я дышу.

*Льюис Кэрролл\**

---

\* Кэрролл Л. Сильвия и Бруно / Пер. А. Голова. М. : Альфа-книга, 2010.

**Н**аши эмоции также рождаются в новой коре, но формируются под влиянием разных отделов мозга — от миндалин в старом мозге до таких новых в эволюционном плане структур, как веретенообразные нейроны, которые, по-видимому, играют ключевую роль в формировании сложных эмоций. В отличие от структур новой коры с регулярным и рекурсивным строением, веретенообразные нейроны характеризуются чрезвычайно нерегулярной формой и связями. Это самые крупные нейроны головного мозга человека, пронизывающие мозг на всем его протяжении. Они плотно переплетены между собой за счет сотен тысяч связей в разных отделах новой коры.

Как было сказано выше, островки мозга помогают обрабатывать сенсорные сигналы, но они также играют важную роль в формировании сложных эмоций. Именно отсюда выходят веретенообразные клетки. Исследования с помощью функциональной магнитно-резонансной томографии показывают, что эти клетки наиболее активны тогда, когда человек испытывает такие эмоции, как любовь, гнев, печаль и сексуальное влечение. Например, в ситуации, когда кто-то другой разглядывает вашего партнера или когда вы слышите плач своего ребенка.

Веретенообразные клетки имеют длинные выросты, называемые апикальными дендритами и способные связывать между собой отдаленные отделы новой коры. Такое «глубокое» связывание, посредством которого некоторые нейроны соединяют несколько отделов мозга, усиливается по мере продвижения по эволюционной лестнице. Нет ничего удивительного в том, что веретенообразные клетки, участвующие в регуляции эмоций и моральных суждений, обладают такой способностью образовывать связи, поскольку эмоциональные реакции возникают у нас при осмыслении самых разнообразных тем и событий. Наши сложные эмоции испытывают влияние всех перцептивных и когнитивных областей мозга, с которыми связаны веретенообразные клетки. Важно подчеркнуть, что эти клетки не занимаются рациональным решением проблем, и поэтому мы

не можем рациональным образом контролировать свою реакцию на музыку или состояние влюбленности. Однако остальная часть мозга усиленно пытается придать смысл нашим загадочным сложным эмоциям.

Веретенообразных клеток в мозге не так много — около 80 тыс., из них примерно 45 тыс. сосредоточено в правом полушарии и 35 тыс. — в левом. Это небольшое неравенство отчасти объясняет, почему эмоциональное восприятие является прерогативой правого полушария мозга.

У горилл примерно 16 тыс. веретенообразных клеток, у карликовых шимпанзе бонобо — 2100, а у обыкновенных шимпанзе — 1800. У других млекопитающих этих клеток нет совсем.

Антропологи полагают, что веретенообразные клетки впервые появились от 10 до 15 млн лет назад у какого-то еще неизвестного предка гоминидов (предшественников человека) и быстро размножились примерно сотню тысяч лет назад. Интересно, что эти клетки отсутствуют у новорожденных детей, но начинают появляться в возрасте около четырех месяцев, и численность их значительно увеличивается от одного года до трех. Способность детей понимать моральные аспекты вопросов и переживать такие сложные эмоции, как любовь, развивается в этот же период.

## **Талант**

Вольфганг Амадей Моцарт (1756–1791) в пять лет сочинил менуэт, а в шесть лет играл для императрицы Марии Терезии в императорском дворце в Вене. Он умер в возрасте 35 лет, но успел создать 600 музыкальных произведений, включая 41 симфонию, и считается величайшим европейским композитором-классиком. Смело можно сказать, что у него был музыкальный талант.

Что это означает в контексте нашей беседы о теории мысленного распознавания образов? Понятно, что часть того, что мы называем талантом, является продуктом влияния среды и окружающих людей. Моцарт родился в музыкальной семье. Его отец, Леопольд,

служил композитором и капельмейстером в придворной капелле князя-архиепископа Зальцбурга. Юный Моцарт с самого детства был погружен в музыку, и отец начал учить его игре на скрипке и клавишных инструментах всего в трехлетнем возрасте.

Однако только влияние среды не может полностью объяснить гениальность Моцарта. Понятно, что какой-то вклад вносит сама природа. В какой же форме это проявляется? Как мы обсуждали в четвертой главе, в результате биологической эволюции отдельные участки новой коры были оптимизированы для обработки определенных типов образов. Безусловно, основной алгоритм распознавания образов один и тот же во всех модулях новой коры, но, поскольку некоторые типы образов проходят через определенные участки коры (например, лица — через веретенообразную извилину), эти участки лучше справляются с обработкой соответствующих образов. Существует множество параметров, определяющих реализацию общего алгоритма в каждом модуле. Например, насколько точное совпадение требуется для распознавания конкретного образа? Как этот порог изменяется, если модуль более высокого порядка посылает сигнал о том, что данный образ «ожидается»? Как учитываются параметры величины сигнала? Эти и другие факторы по-разному определяются в разных отделах коры для преимущественной обработки определенного типа образов. В нашей работе, связанной с созданием искусственного интеллекта, мы обратили внимание на это явление и имитировали эволюцию, направленную на оптимизацию этих параметров.

Если отдельные участки мозга могут быть оптимизированы для обработки разных типов образов, это означает, что мозг разных людей может различаться по способности заучивать, распознавать и создавать определенные типы образов. Например, мозг одного человека может иметь врожденную способность к музыке за счет лучшего распознавания ритмических образов или лучшего понимания геометрического строения гармонии. По-видимому, абсолютный слух (способность распознать

и воспроизвести звук без помощи эталона), связанный с музыкальными способностями, имеет генетическую природу, однако его нужно развивать. Таким образом, абсолютный слух — продукт и природы, и воспитания. Генетические основы абсолютного слуха, по-видимому, связаны не с новой корой, а с процессом предварительной обработки слуховой информации, а вот развитие абсолютного слуха происходит в новой коре.

Степень человеческого таланта зависит от многих других навыков вне зависимости от того, идет ли речь о гениях или о вариабельности в общей популяции. Важную роль играют многие способности новой коры — например, способность контролировать сигналы страха, направляемые миндалинами, а также такие черты, как уверенность в себе, организаторские способности и умение убеждать других людей. Еще одна очень важная черта, о которой я уже говорил выше, заключается в смелости следовать идее, идущей вразрез с общепринятыми теориями. Все те, кого мы считаем гениями, проводили собственные мысленные эксперименты таким путем, который исходно не был понят или оценен их современниками. Талант Моцарта получил признание при жизни, но истинное почитание пришло позже. Он умер нищим, был зарыт в общей могиле, а его гроб провожали лишь двое музыкантов.

## Творчество

Творчество — это лекарство, без которого я не могу жить.

*Сесил Б. Демилль\**

Проблема состоит не в том, чтобы обзавестись новыми мыслями, а в том, чтобы избавиться от старых. Мозг любого человека — это здание, заполненное старьем. Освободите уголок своего мозга, и в нем немедленно поселится творчество.

*Ди Хок\*\**

---

\* Сесил Блаунт Демилль (1881–1959) — американский режиссер и продюсер.

\*\* Ди Хок (род. в 1929) — основатель и бывший глава компании Visa Inc.

Человечество может достаточно прохладно относиться к тем, кто смотрит на мир иначе.

*Эрик А. Бернс\**

Творчество может решить практически любую проблему. Творческий акт, победа новизны над привычкой, преодолевает все.

*Джордж Лоис\*\**

Важнейший аспект творчества — поиск хороших метафор. Новая кора — великолепный аппарат для поиска метафор, который наделяет человека уникальными творческими способностями. Каждый из 300 млн распознающих модулей новой коры распознаёт и определяет образ, а также дает ему «имя» — в данном случае это просто определенный аксон, выходящий из данного распознающего модуля и возбуждающийся при обнаружении данного образа. Этот образ (или символ) становится частью другого образа. И каждый из этих образов по сути является метафорой. Модули могут активироваться до 100 раз в секунду, так что мы теоретически можем распознавать до 30 млрд метафор в секунду. Конечно, в каждом цикле активируются не все модули, но все же мы действительно способны распознавать миллионы метафор за одну секунду.

Естественно, одни метафоры лучше, другие хуже. Дарвин понял, что догадка Чарлза Лайеля о роли постепенных изменений ландшафта под действием струйки воды была прекрасной метафорой для описания того, как малые эволюционные изменения могут приводить к значительным различиям, определяющим разнообразие видов. Мысленные эксперименты, подобные экспериментам Эйнштейна, с помощью которых он пытался понять истинное значение экспериментов Майкельсона—Морли, также являются метафорами в том смысле, что «символизируют что-то другое», если следовать определению из «Американского словаря

---

\* Эрик Бернс — современный американский писатель, драматург и журналист.

\*\* Джордж Лоис (род. в 1931) — американский фотограф, дизайнер и писатель.

английского языка» издательства Мерриам—Уэбстер. Вы можете найти какую-нибудь метафору у Шекспира в Сонете 73?

То время года видишь ты во мне,  
 Когда один-другой багряный лист  
 От холода трепещет в вышине  
 На хорах, где умолк веселый свист.  
 Во мне ты видишь тот вечерний час,  
 Когда поблек на западе закат  
 И купол неба, отнятый у нас,  
 Подобьем смерти — сумраком объят.  
 Во мне ты видишь блеск того огня,  
 Который гаснет в пепле прошлых дней,  
 И то, что жизнью было для меня,  
 Могилою становится моей.  
 Ты видишь все. Но близостью конца  
 Теснее наши связаны сердца!\*

Для описания происходящих с ним изменений поэт широко использует метафоры. Его возраст — это осень, когда на деревьях остался всего «один-другой багряный лист», когда стихли птицы и «умолк веселый свист». Его возраст подобен сумеркам, когда «поблек на западе закат и купол неба, отнятый у нас, подобьем смерти — сумраком объят». Он — лишь отблеск того огня, который «гаснет в пепле прошлых дней». Весь сонет — метафора, хотя одни фразы являются более яркими метафорами, чем другие.

Поиск метафор заключается в распознавании образов при изменении деталей и контекста — то, чем мы занимаемся постоянно в нашей каждодневной жизни. Значащие метафоры часто обнаруживаются на стыках разных дисциплин. Следовательно, деятельность, направленная против этой важнейшей способности к творчеству, является движением к еще большей специализации в рамках отдельных наук (и не только наук). Вот что писал

---

\* Перевод С. Маршака.

американский математик Норберт Виннер (1894–1964) в своей знаменитой книге «Кибернетика», опубликованной в год моего рождения (1948):

«Ниже мы увидим, что существуют области научной работы, исследуемые с разных сторон чистой математикой, статистикой, электротехникой и нейрофизиологией. В этих областях каждое понятие получает особое название у каждой группы специалистов и многие важные исследования проделываются трижды или четырежды. В то же время другие важные исследования задерживаются из-за того, что в одной области неизвестны результаты, уже давно ставшие классическими в смежной области. Именно такие пограничные области науки открывают перед надлежаще подготовленным исследователем богатейшие возможности. Но изучение таких областей представляет и наибольшие трудности для обычного метода массового наступления с разделением труда»\*.

Техника борьбы с возрастающей специализацией, которую сам я использовал в своей работе, состояла в том, чтобы соединить усилия многих экспертов (например, в исследованиях по распознаванию речи участвовали лингвисты, психоакустики и специалисты по распознаванию образов, а также, естественно, компьютерщики) и стимулировать каждого из них объяснять всем другим членам коллектива свои методы и свою терминологию. Затем мы отбросили всю эту терминологию и разработали общую. Нам всегда удавалось найти метафоры из одной дисциплины, которые разрешали проблемы в другой дисциплине.

Можно сказать, что мышь, пытающаяся найти путь к побегу, спасаясь от домашнего кота в какой-то новой ситуации, занимается творчеством. Наш с вами творческий потенциал несоизмеримо больше, чем у мыши, и происходит на гораздо более высоком уровне абстракции, поскольку мы являемся обладателями гораздо более развитой новой коры, способной к более

---

\* *Винер Н.* Кибернетика, или Управление и связь в животном и машине / Пер. с англ. И. В. Соловьева и Г. Н. Поварова. М. : Наука, 1983.

сложным иерархическим построениям. Таким образом, один из способов повышения творческого потенциала заключается в эффективной сборке более объемной новой коры.

Одна из возможностей — совместная деятельность многих людей. Это рутинная практика, заключающаяся в общении между людьми, занятыми решением общей задачи. Недавно для усиления этой практики в области математики и других дисциплин стали применяться возможности работы в интерактивном режиме в реальном времени<sup>1</sup>.

Следующим шагом, конечно же, будет увеличение объема самой новой коры за счет ее небиологического эквивалента. Это будет высшее достижение творчества: создание способности создавать! В итоге небиологическая новая кора сможет быстрее работать и быстрее находить метафоры, вдохновлявшие Дарвина и Эйнштейна. Она будет систематически эксплуатировать все пограничные зоны между экспоненциально расширяющимися областями знаний.

Некоторые беспокоятся о том, что станет с людьми, которые не примут участия в этом расширении разума. Я хочу заметить, что этот дополнительный разум будет главным образом находиться в облаке (в экспоненциально расширяющейся компьютерной инфраструктуре, к которой мы подключаемся для интерактивного общения), где уже теперь сосредоточена большая часть имеющегося в нашем распоряжении «машинного разума». Когда вы используете поисковую систему, узнаете голос в телефоне, консультируетесь с виртуальным помощником типа Сири или используете телефон для перевода информации на другой язык, «разум» заключен не в том устройстве, с которым вы имеете дело, а в облаке. И наша расширенная новая кора тоже будет там. Возможно, мы станем обращаться к этому разуму путем прямого контакта через нервные клетки, а может быть, так, как мы это делаем сейчас, — через вспомогательные устройства. Мне кажется, что все мы сможем расширить свои творческие способности вне зависимости от того, решимся мы напрямую

подсоединиться к этому расширенному разуму или нет. Мы уже передали облаку значительную часть нашей персональной, социальной, исторической и культурной памяти и в итоге поступим так же и с нашим иерархическим мышлением.

Осуществленный Эйнштейном прорыв в науке объясняется не только тем, что он использовал метафоры в своих мысленных экспериментах, но и тем, что у него хватило смелости верить в эти метафоры. Он смог отступить от традиционных теорий, не сумевших объяснить его эксперименты, и выдержать насмешки современников за те нелепые объяснения, которые он выдвигал на основе своих метафор. Эти качества — вера в метафору и убежденность — мы тоже сможем программировать в нашей небιологической новой коре.

## Любовь

Ясность ума влечет за собой чистоту страсти, поэтому мужчина, наделенный глубоким и ясным умом, способен горячо любить и всегда отдает себе отчет в том, что он любит.

*Блез Паскаль*

Любовь — это безумие, но и в безумии есть разум.

*Фридрих Ницше*

Когда вы узнаете жизнь так же, как я, вы не будете недооценивать власть одержимой любви...

*Дж. К. Роулинг. Гарри Поттер и принц-полукровка*

Мне всегда нравятся хорошие математические решения для любой любовной проблемы.

*Майкл Патрик Кинг\* (телесериал «Секс в большом городе», серия «Отведи меня на бейсбол», 1999)*

Если вам лично не довелось пережить любовного экстаза, вы наверняка о нем слышали. Значительная часть всего ми-

---

\* Майкл Патрик Кинг (род. в 1954) — американский сценарист, режиссер и продюсер.

рового искусства — книг, музыки, танца, картин, телевизионных передач и фильмов — вдохновляется именно историями заставляющей любви.

Недавно к изучению этого вопроса подключилась и наука, и теперь мы можем идентифицировать биохимические изменения в организме влюбленного человека. Уровень дофамина повышается, что создает ощущение счастья и восхищения. Уровень норадреналина тоже взлетает, что повышает частоту сердечных сокращений и вызывает общее возбуждение. Эти вещества, а также обладающий стимулирующим действием фенилэтиламин вызывают эйфорию, прилив энергии, способствуют концентрации внимания, потере аппетита и непреодолимой тяге к предмету обожания. Проведенные недавно в Университетском колледже Лондона исследования показали, что при влюбленности падает уровень серотонина, как при обсессивно-компульсивном расстройстве, что вполне объясняет маниакальное состояние человека в начале влюбленности<sup>2</sup>. Высокий уровень дофамина и норадреналина объясняют краткосрочную сосредоточенность, эйфорию и страстное стремление к любимому человеку.

Как видно, биохимические процессы при влюбленности очень напоминают таковые при реакции борьбы или бегства. Вообще говоря, это практически одно и то же состояние, за тем исключением, что здесь мы стремимся к кому-то или к чему-то. Циник сказал бы, что мы бежим к опасности, а не убегаем от нее. Такие же изменения наблюдаются и в ранней фазе развития зависимости. Очень точно это состояние описано в песне группы Roxу Music под названием Love is the Drug («Любовь — это наркотик»; герой песни хочет получить следующую «дозу» любви). Исследования религиозного экстаза демонстрируют те же самые явления; можно сказать, что человек влюбляется в Бога или в какой-то иной объект поклонения.

В начале романтической любви эстроген и тестостерон, безусловно, играют роль в сексуальном поведении людей, но если бы

единственной целью любви в эволюционном плане было воспроизведение, романтический аспект вовсе не понадобился бы. Как писал психолог Джон Уильям Мани (1921–2006), «страсть похотлива, любовь лирична». Фаза любовного экстаза сменяется фазой привыкания, которая затем переходит в долгосрочную связь. Все эти процессы тоже стимулируются химическими веществами, включая окситоцин и вазопрессин. Сравним, к примеру, два родственных вида мышей-полевков — степную и горную. Они очень похожи, вот только у степных полевков есть рецепторы окситоцина и вазопрессина, а у горных полевков их нет. Степные полевки заводят моногамные отношения, выбирая себе партнера на всю жизнь, а горные вступают в однократные, «случайные» связи. Ясно, что для полевков рецепторы окситоцина и вазопрессина играют решающую роль в характере их любовных отношений.

Для людей эти гормоны тоже важны, но ведущую роль, как и во всей нашей жизни, играет новая кора. У полевков есть новая кора, но эта плоская структура размером с почтовую марку может помочь им только в поиске партнера на всю жизнь (или на одну ночь) и руководить обычными жизненно важными функциями. У нас кора достаточно велика, чтобы снабдить нас «лирическими» эмоциями, о которых говорит Мани.

В эволюционной перспективе любовь существует для того, чтобы удовлетворять нужды новой коры. Если бы у нас не было коры, сексуального влечения оказалось бы вполне достаточно для обеспечения воспроизводства. Любовный экстаз превращается в привыкание, в зрелую любовь и перерастает в долговую связь. Это, в свою очередь, необходимо хотя бы для того, чтобы обеспечить благоприятную среду для подрастающих детей в тот период, когда их собственная новая кора обучается и «взрослеет». Обучение в богатой среде — важнейший принцип действия новой коры. Кроме того, окситоцин и вазопрессин играют ключевую роль в установлении связи между родителями (главным образом матерью) и ребенком.

В ходе развития любовной истории объект любви начинает занимать значительную часть нашей новой коры. После десятилетий совместной жизни в новой коре существует виртуальный образ любимого человека, так что мы всегда можем предсказать, что он сейчас скажет или сделает. Образы в нашей новой коре состоят из мыслей и картинок, отражающих сущность этого человека. Когда мы его теряем, мы буквально теряем часть самого себя. И это *не просто* метафора — все распознающие модули, заполненные образами любимого человека, внезапно меняют свое содержимое. Казалось бы, распознающие модули можно считать ценнейшим хранилищем виртуального образа любимого человека внутри нас самих, но дело в том, что эти образы вдруг перестают вызывать радость, а начинают порождать грусть.

Эволюционные основы любви и фаз ее развития еще не до конца понятны. Мы давно отделили занятия сексом от их биологической функции: мы можем иметь детей, не занимаясь сексом, и уж точно можем заниматься сексом, не производя детей. Занятия сексом выполняют плотскую и в каком-то смысле коммуникационную функцию. Но мы по-прежнему влюбляемся, и совсем не по той причине, что хотим иметь детей.

Аналогичным образом, с античных времен существовала потребность изображать любовь и ее всевозможные формы с помощью самых разных видов искусства. Наша способность воспроизводить этот извечный человеческий опыт — о любви и о чем-то еще — как раз и определяет уникальность нашего вида.

Новая кора — величайшее творение природы. А стихи о любви и все другие плоды нашего творчества — величайшие достижения новой коры.



# Глава седьмая

## ЦИФРОВАЯ КОРА — МОДЕЛЬ БИОЛОГИЧЕСКОЙ КОРЫ

Не верь тому, что может думать, но не видно, где у него мозги.

*Артур Уизли, персонаж книги  
«Гарри Поттер и узник Азкабана» Дж. К. Роулинг*

Нет, меня не интересует создание мощного разума. Все, что меня интересует, — это создание заурядного ума, что-то вроде ума президента Американской телефонной и телеграфной компании.

*Алан Тьюринг\**

Можно считать, что компьютер «мыслит», если человек не может отличить его от другого человека.

*Алан Тьюринг*

Я полагаю, что к концу столетия использование слов и общий уровень образования изменятся настолько, что каждый будет способен говорить о машинном мышлении, не ожидая, что ему начнут противоречить.

*Алан Тьюринг*

---

\* Алан Мэтисон Тьюринг (1912–1954) — английский математик и логик, один из основоположников современной информатики; предложенная им абстрактная вычислительная «машина Тьюринга» считается прототипом современного компьютера.

**К**рыса строит гнездо для своего потомства, даже если никогда в жизни не видела других крыс<sup>1</sup>. Паук плетет паутину, гусеница вьет кокон, а бобер строит плотину, даже если никто из сородичей никогда не показывал им, как это делается. Но это не означает, что этим навыкам они не учились. Это означает, что животные учились этому не на протяжении одной жизни, а на протяжении тысяч жизней. Эволюция поведения животных представляет собой процесс обучения, но это обучение происходит не на уровне отдельных особей, а на уровне целых видов, а его результаты записываются в ДНК.

Чтобы оценить эволюцию новой коры, представьте себе, что процесс ее обучения (иерархическим знаниям) сократился от тысячелетий до месяцев (и даже меньше). Даже если миллионам животных какого-то вида млекопитающих не удавалось решить определенную проблему (требующую иерархического подхода), одно из них вдруг неожиданно находило такое решение. И постепенно этот новый метод копировался и распространялся во всей популяции.

Теперь, осуществляя переход от биологического к небιологическому разуму, мы ускоряем процесс обучения еще в тысячи или миллионы раз. После того как цифровая новая кора усваивает какую-то информацию, она может передать ее через минуты или даже секунды. Вот лишь один пример. В моей первой компании Kurzweil Computer Products (KCP, теперь это Nuance Speech Technologies), основанной в 1973 г., мы потратили несколько лет на то, чтобы обучить компьютеры распознавать печатные буквы на отсканированных документах; эта технология называлась омнифонтовым распознаванием знаков (распознаванием знаков любого типа). Эта технология продолжала развиваться на протяжении 40 с лишним лет, и сегодня Nuance производит программу под названием OmniPage. И теперь, если нужно, чтобы ваш компьютер смог распознавать печатные буквы, не нужно учить его этому несколько лет, как это делали мы. Вы просто загружаете выученную компьютером информацию

в виде компьютерной программы. В 1980-х гг. мы начали заниматься распознаванием речи, и эта технология, которая также развивалась на протяжении нескольких десятилетий, теперь лежит в основе функционирования Сири (виртуальный помощник в вашем айфоне). И вновь, если нужно, вы за секунды можете включить те знания, что исследовательские компьютеры осваивали на протяжении десятилетий.

Когда-нибудь мы создадим искусственную новую кору, которая будет обладать всеми возможностями биологического аналога. Зачем это нужно? Электронные схемы действуют в миллионы раз быстрее биологических. В первую очередь мы используем этот выигрыш в скорости, чтобы компенсировать некоторый недостаток параллельности в работе современных компьютеров, но в конце концов искусственная новая кора сможет действовать намного быстрее своего биологического аналога, и скорость ее работы будет возрастать и дальше.

При усилении биологической новой коры искусственным аналогом не нужно беспокоиться о том, сколько дополнительных единиц новой коры может физически поместиться в нашем теле или мозге, поскольку большая часть этой структуры будет находиться в облаке, как и большая часть используемых сегодня компьютерных систем. Я уже писал, что, по моим оценкам, в нашей биологической новой коре содержится около 300 млн распознающих модулей. Больше в наш череп вместить невозможно — даже после появления у нас в ходе эволюции крупного лба и с учетом того, что новая кора занимает до 80% всего объема мозга. Но, когда наше мышление будет происходить в облаке, исчезнут природные ограничения, и мы сможем использовать миллиарды или триллионы распознающих модулей — столько, сколько нужно и сколько будет позволять закон ускорения отдачи.

Чтобы обучить искусственную новую кору новым навыкам, понадобится еще множество итераций, как и в случае биологической коры, но, когда одна какая-то цифровая кора в какой-то

момент и в каком-то месте усваивает знания, она способна немедленно обменяться этими знаниями с любой другой искусственной корой. И каждый из нас сможет иметь персональное расширение своей новой коры в облаке, как сегодня имеет хранилище персональных данных.

Наконец, и это очень важно, мы сумеем сохранять цифровую часть нашего разума. Как мы уже поняли, в нашей новой коре содержится информация, которая, страшно подумать, никак не сохраняется. Конечно, эту информацию можно записать. Возможность передать хотя бы часть нашего разума некоему носителю, который переживет наше биологическое тело, безусловно, является важным шагом вперед, но значительная часть содержащейся в нашем мозге информации по-прежнему теряется.

### **Имитация мозга**

Один из подходов к созданию искусственного мозга состоит в точной имитации биологического мозга. Например, докторант из Гарварда Дэвид Делримпл (род. в 1991 г.) планирует воспроизвести головной мозг нематоды (круглого червя)<sup>2</sup>. Делримпл выбрал нематоду по той причине, что у нее достаточно простая нервная система, состоящая примерно из 300 нейронов, которую он планирует воспроизвести на молекулярном уровне. Он также собирается создать компьютерную модель тела нематоды в соответствующей реальности окружающей среде, так что эта виртуальная нематода сможет (виртуально) добывать себе пищу и делать всякие другие вещи, которые обычно делают нематоды. Делримпл считает, что эта работа станет первым полным воссозданием головного мозга биологического животного в форме виртуального мозга виртуального животного. Являются ли такие виртуальные нематоды (да и биологические нематоды) существами сознательными — вопрос спорный, однако в своей борьбе за пропитание, в процессах пищеварения, попытках избежать хищников и в воспроизведении они должны опираться на определенный осознанный опыт.

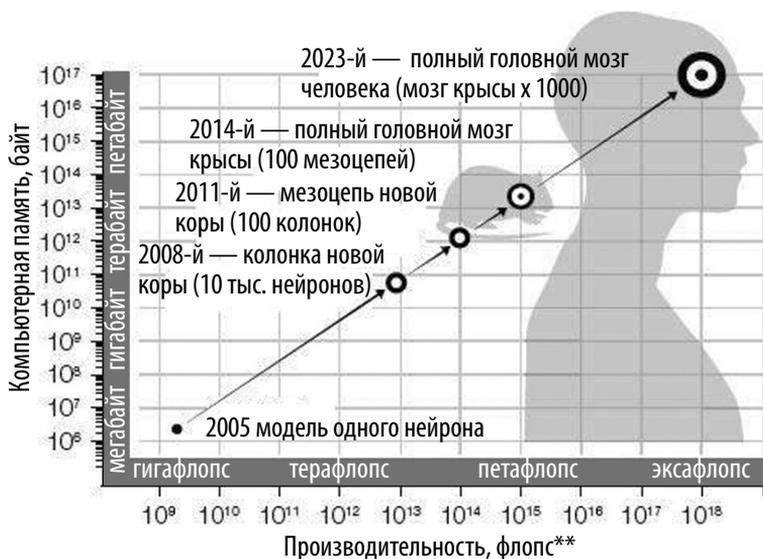
На другой границе спектра исследований, в рамках проекта «Голубой мозг» Генри Маркрама, планируется проектирование человеческого мозга, включая всю новую кору, а также такие отделы старого мозга, как гиппокамп, миндалины и мозжечок. Имитация будет осуществляться на разных уровнях детализации, вплоть до полной имитации на молекулярном уровне. Как я рассказывал в четвертой главе, Маркрам открыл существование ключевого модуля из нескольких десятков нейронов, многократно повторяющегося в структурах новой коры, и показал, что именно эти модули, а не отдельные нейроны отвечают за процесс обучения.

Развитие идей Маркрама происходило по экспоненциальному закону. В 2005 г. — в первый год реализации проекта — было осуществлено моделирование действия одного нейрона, а в 2008 г. его научная группа моделировала активность целой колонки новой коры головного мозга крысы, состоящей из 10 тыс. нейронов. К 2010 г. исследователи смогли осуществить моделирование 100 колонок, что в сумме составляло около миллиона клеток (Маркрам назвал эту группу клеток мезоцепью). Одна из проблем заключается в том, что мы не имеем возможности проверить точность этого моделирования. Чтобы это сделать, с помощью данной модели нужно продемонстрировать процесс обучения, о чем я расскажу ниже.

К концу 2014 г. Маркрам планировал создать модель целого мозга крысы, состоящую из 100 мезоцепей; эта система включает 100 млн нейронов и примерно 1 трлн синапсов. В своем выступлении на конференции TED (Technology Entertainment Design) в 2009 г. в Оксфорде Маркрам заявил: «Нет ничего невозможного в том, чтобы создать головной мозг человека, и мы сделаем это через 10 лет»<sup>3</sup>. Но теперь он планирует решить эту задачу к 2023 г.<sup>4</sup>

Маркрам и его коллеги строят свою модель на детальном анатомическом и электрохимическом анализе нейронов. С помощью созданного ими автоматического устройства, которое

они назвали «пэтч-кламп-роботом»\*, они анализируют состояние специфических ионных каналов, нейромедиаторов и ферментов, отвечающих за электрохимическую активность каждого нейрона. По утверждению Маркрама, эта автоматическая система проделала тридцатилетний объем работы за шесть месяцев. Именно на основании этого анализа они сформулировали идею «кубиков лего», составляющих основную функциональную единицу новой коры.



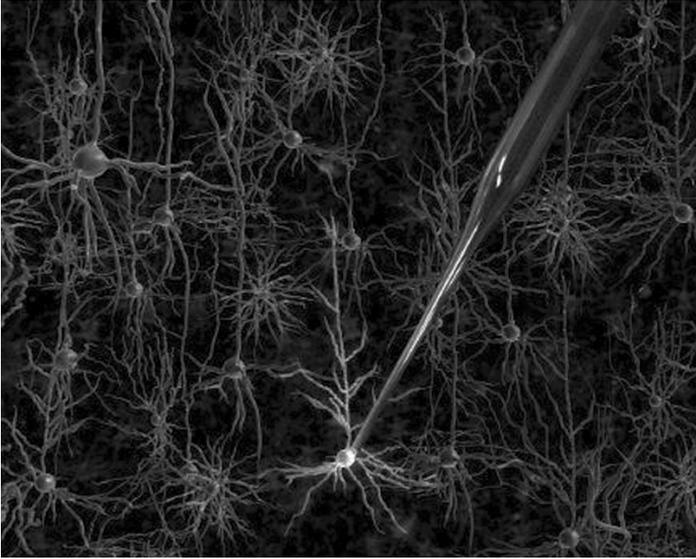
Реальный и прогнозируемый прогресс в развитии проекта «Голубой мозг»

Важный вклад в развитие метода внесли нейрофизиолог Эд Бойден из Массачусетского технологического института, профессор Крейг Форест из отдела механики Технологического университета Джорджии и его студент Сухаса Кодандарамаях.

\* Метод пэтч-кламп (от англ. *patch* — фрагмент, *clamp* — фиксация) — электрофизиологический метод изучения свойств ионных каналов.

\*\* Флопс (от англ. *floating-point operations per second*) — внесистемная единица измерения производительности компьютеров, показывающая, сколько операций в секунду выполняет вычислительная система.

Они создали автоматическую систему, способную сканировать нервную ткань с точностью до одного микрометра, не повреждая тонкие мембраны нейронов. «Это то, что может делать робот, но не может человек», — прокомментировал Бойден.



*Наконечник «пэтч-кламп робота», созданного в Массачусетском технологическом институте, сканирует нервную ткань.*

Говорят, что после моделирования поведения одной колонки новой коры Марккрам заявил: «Теперь нам остается только масштабировать процесс»<sup>5</sup>. Масштабирование, безусловно, важный фактор, но есть и еще один — обучение. Если «Голубой мозг» должен «говорить, мыслить и вести себя во многом как человек», о чем Марккрам говорил в интервью на BBC в 2009 г., его новая кора должна иметь значительное содержание<sup>6</sup>. Но, как может подтвердить всякий, кто пытался беседовать с новорожденным ребенком, чтобы достичь этой цели, мозг должен очень многому научиться.

Для решения подобной задачи в такой модели, как «Голубой мозг», существуют две очевидные возможности. Первая

состоит в обучении искусственного мозга таким же способом, как обучается мозг человека. Вначале вы имеете систему, соответствующую мозгу новорожденного ребенка, обладающую способностью усваивать иерархические знания и заранее заданной способностью трансформировать сенсорные сигналы. Однако для обучения небиологического мозга нужен такой же контакт, как между новорожденным ребенком и развивающим его взрослым человеком. Проблема заключается в том, что создание такого мозга ожидается не ранее начала 2020-х гг. Однако функционирование в реальном времени будет слишком медленным, поскольку исследователям придется подождать десять или двадцать лет, пока мозг достигнет интеллектуального уровня взрослого человека, хотя, конечно, производительность компьютеров продолжает неуклонно возрастать.

Второй подход заключается в том, чтобы скопировать образы, содержащиеся в новой коре одного или нескольких взрослых людей, обладающих достаточными знаниями, чтобы нормально пользоваться речью и вообще вести себя как развитый взрослый человек. В данном случае проблема заключается в том, чтобы найти неинвазивный и неdestructивный метод сканирования с достаточным временным и пространственным разрешением и скоростью, чтобы провести эту процедуру быстро и без потерь. Мне кажется, этот метод «загрузки информации» не будет реализован до 2040-х гг.

Но есть еще и третья возможность, которую, как мне кажется, и следует использовать при развитии таких моделей, как «Голубой мозг». Можно упростить молекулярные модели путем создания функциональных эквивалентов на разных уровнях специфичности, используя различные методы — от моего собственного функционального алгоритмического метода (описанного в этой книге) до практически полных молекулярных моделей. При этом в зависимости от степени упрощения можно повысить скорость обучения в несколько сотен

или даже тысяч раз. Для такой модели искусственного мозга можно разработать обучающую программу, которая будет осваиваться достаточно быстро. А затем упрощенную модель можно заменить полной молекулярной моделью, сохраняя накопленные знания. После этого можно начать моделирование процесса обучения полной молекулярной модели на гораздо более низкой скорости.

Американский специалист в области информатики Джармендра Модха и его коллеги из корпорации IBM на клеточном уровне создали модель участка человеческой зрительной коры, состоящий из 1,6 млрд виртуальных нейронов и 9 трлн синапсов, что по размеру эквивалентно новой коре кота. Эта система работает в 100 раз медленнее, чем суперкомпьютер IBM BlueGene/P, состоящий из 147 456 процессоров. Эта работа была удостоена премии Гордона Белла, которая вручается Ассоциацией вычислительной техники (ACM).

Задача таких моделей, как «Голубой мозг» или новая кора Модха, заключается в уточнении и подтверждении функциональной модели. Модель человеческого искусственного разума в принципе будет использовать именно такие алгоритмы. Однако молекулярные модели помогают усовершенствовать эти алгоритмы и более полно изучить наиболее важные детали. В ходе наших экспериментов 1980-х и 1990-х гг., посвященных созданию технологии распознавания речи, нам удалось усовершенствовать наши алгоритмы, когда мы поняли суть реальных превращений, происходящих в слуховом нерве и некоторых участках новой коры. Даже если наша функциональная модель безупречна, понимание ее функционирования в биологическом мозге даст важную информацию о функции и дисфункции человеческого организма.

Для создания моделей мозга нам нужна очень подробная информация о функционировании настоящего мозга. Группа Маркрама собирает собственные данные. Сбору подобной информации и ее передаче широкой научной общественности

посвящено несколько крупномасштабных проектов. Например, в результате сканирования головного мозга мышей лаборатория в Колд-Спринг-Харбор получила 500 терабайт информации, которую выложила для общественного доступа в июне 2012 г. Этот проект позволяет пользователям изучать мозг таким же образом, как программа Google Earth позволяет изучать поверхность планеты. Вы можете перемещаться внутри мозга и при более сильном приближении рассматривать отдельные нейроны и их контакты. Вы можете выделить отдельную связь и следовать по ней вдоль всего мозга.

Шестнадцать отделов Национального института здоровья США объединились и выделили 38,5 млн долларов на реализацию проекта под названием «Коннектом человека»<sup>7</sup>. Проект выполняется под руководством Университета Вашингтона в Сент-Луисе, Университета Миннесоты, Гарвардского университета, Массачусетского генерального госпиталя и Университета Калифорнии в Лос-Анджелесе и направлен на создание трехмерной карты связей нейронов в головном мозге человека. С этой целью применяются различные неинвазивные технологии сканирования, включая новые варианты МРТ, магнитоэнцефалографию (МЭГ — измерение магнитных полей, образующихся в результате электрической активности мозга) и диффузионную трактографию (анализ пучков нервных волокон в головном мозге). Как я расскажу в десятой главе, пространственное разрешение неинвазивных методов сканирования мозга увеличивается с невероятной скоростью. Одними из первых результатов проекта были данные Ван Видена и его коллег из Массачусетского генерального госпиталя о том, что в головном мозге существует регулярная сеть нейронов, напоминающая структуру возбуждения нейронов новой коры, о которой я рассказывал в четвертой главе.

Специалист в области компьютерной нейробиологии из Оксфордского университета Андерс Сэндберг (род. в 1972 г.) и шведский философ Ник Бостром (род. в 1973 г.) разработали детальный план эмуляции головного мозга, в котором определены

подходы к моделированию головного мозга человека (и других существ) на разном уровне детализации — от сложных функциональных моделей до имитации действия отдельных молекул<sup>8</sup>.

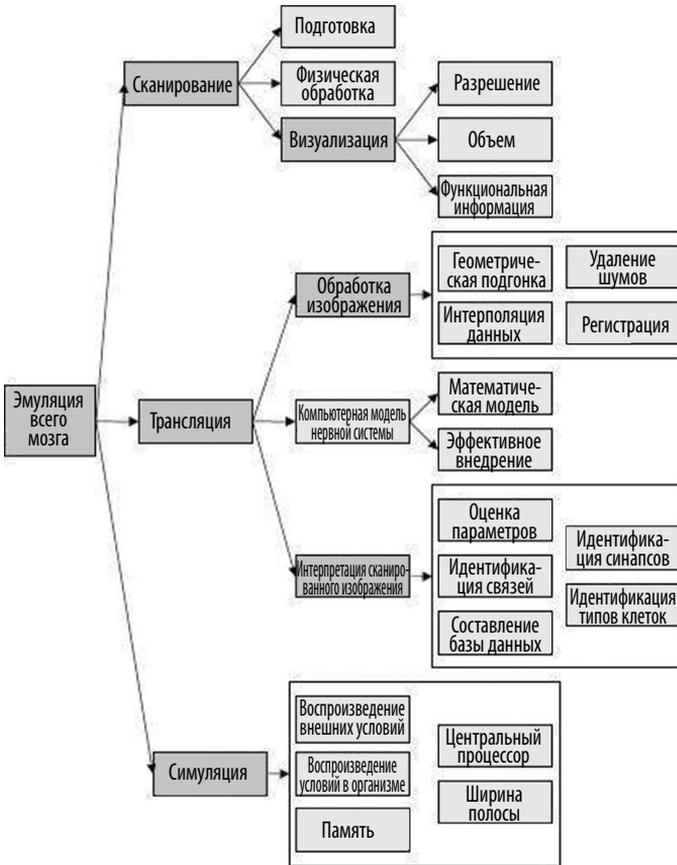
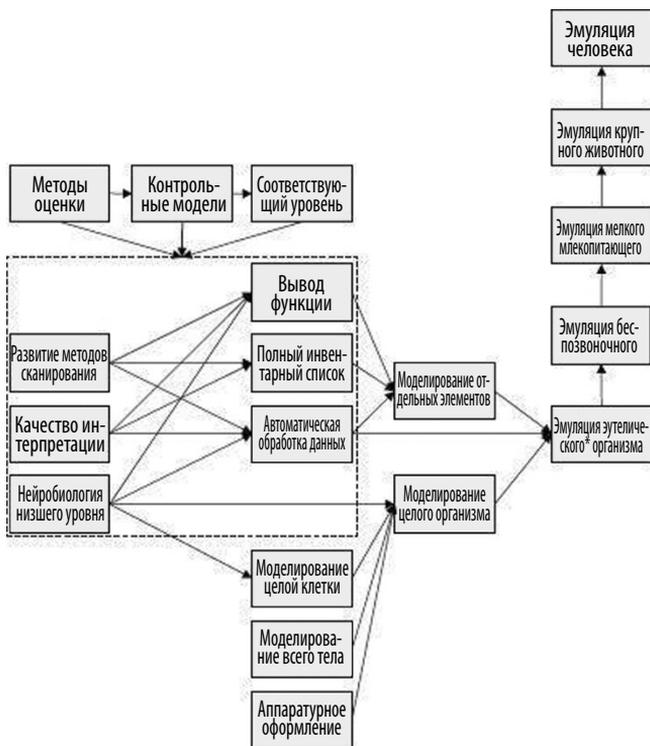


Схема эмуляции активности человеческого мозга  
(Andrs Sandberg, Nick Bostrom, *Whole Brain Emulation: A Roadmap*).

Этот план не дает временных ориентиров, но описывает требования, которые должны быть выполнены для воссоздания различных типов мозга с разной степенью детализации в плане сканирования, моделирования, хранения информации и вычислений. Авторы проекта указывают на неизбежный

экспоненциальный рост информации во всех этих направлениях, который позволит осуществить моделирование человеческого мозга с высокой степенью детализации.



## Нейронные сети

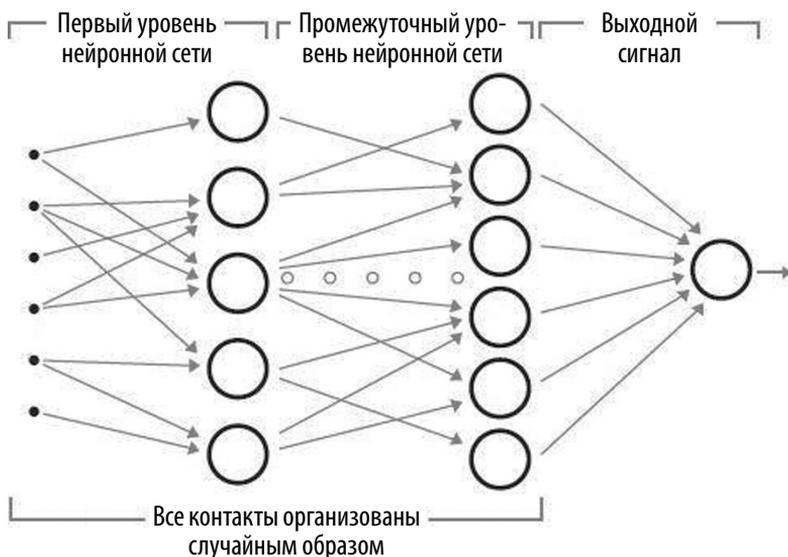
В 1964 г., когда мне было 16 лет, я написал письмо профессору Фрэнку Розенблатту из Корнельского университета с просьбой рассказать о машине под названием Mark 1 Perception. Профессор создал эту машину за четыре года до того, и она, как было заявлено, обладала свойствами человеческого мозга. Розенблатт пригласил меня к себе, чтобы опробовать машину.

\* Этулия — постоянство клеточного состава организма; такие организмы обладают определенным числом клеток, характерным для данного вида.

Система Perception была построена на основе электронных моделей нейронов. На вход в систему подавались двумерные значения. В случае речевых образов одно измерение соответствовало частоте, а второе — времени, так что каждое значение демонстрировало интенсивность частоты в данный момент. В случае зрительных образов каждая точка представляла собой пиксел в двумерном массиве. Входные сигналы случайным образом поступали на модельные нейроны первого уровня. Все связи характеризовались синаптическим потенциалом, отражающим их значимость, который исходно устанавливался случайным образом. Каждый нейрон суммировал поступающие к нему сигналы. Если суммарный сигнал превышал некое пороговое значение, нейрон возбуждался и посылал выходной сигнал на следующий контакт. Если суммарный сигнал был меньше порогового значения, нейрон не возбуждался и выходной сигнал равнялся нулю. Выходные сигналы всех нейронов случайным образом поступали на нейроны следующего уровня. Таких уровней в системе Mark 1 Perception существовало три, и каждый из них мог быть организован в различной конфигурации. Например, какой-то уровень мог иметь обратную связь с предыдущим. На высшем уровне в результате случайного выбора выходных сигналов одного или нескольких нейронов формировался ответ (алгоритм действия сетей нейронов подробно описан в приложении<sup>9</sup>).

Поскольку возбуждение нейронов и величина синаптического потенциала сначала выбирались случайным образом, ответы такой необученной сети нейронов тоже были случайными. Таким образом, важнейшим элементом в моделировании нейронной сети является обучение, как и в головном мозге млекопитающего, который она имитирует. Сначала нейронная сеть ничего не знает. Ее учитель (это может быть человек, компьютерная программа или, возможно, другая, более зрелая нейронная сеть, которая уже получила какие-то знания) вознаграждает обучающуюся сеть, когда та генерирует правильный ответ,

и наказывает, если ответ неправильный. Эта обратная связь используется обучающейся сетью для корректировки силы межнейронных контактов. Контакты, приводящие к правильным ответам, становятся прочнее, а те, что формируют неправильный ответ, напротив, ослабевают.



Со временем нейронная сеть организуется таким образом, что может выдавать правильные ответы безо всякой подсказки. Эксперименты показывают, что нейронные сети могут обучаться даже у ненадежных учителей. Если учитель вносит исправления лишь в 60% случаев, обучающаяся нейронная сеть все равно выучит урок так, что точность ответов приблизится к 100%.

Однако достаточно скоро стало очевидно, что Perception может усвоить лишь ограниченный объем материала. Когда я посетил профессора Розенблатта в 1964 г., я попытался слегка изменить входные данные. Система была натренирована на распознавание печатных букв и делала это достаточно точно. Она довольно хорошо выполняла задачу самоассоциации (узнавала букву, если я закрывал ее часть), но не справлялась

с инвариантностью (путалась при изменении размера буквы или шрифта).

Во второй половине 1960-х гг. нейронные сети стали невероятно популярными, и «коннекционизму» была посвящена как минимум половина исследований в области искусственного интеллекта. Между тем, более традиционный подход к созданию искусственного интеллекта состоял в попытках программирования решений специфических задач, таких как распознавание инвариантных свойств печатных букв.

В том же 1964 г. я посетил Марвина Минского, одного из основоположников исследований в области искусственного интеллекта. Хотя он сам выполнил несколько пионерских исследований по изучению нейронных сетей еще в 1950-х гг., он был недоволен активным развитием исследований в этом направлении. Считалось, что нейронные сети не нуждаются в программировании, поскольку обучаются решать задачи самостоятельно. В 1965 г. я поступил в Массачусетский технологический институт. Минский был моим научным руководителем, и я разделил его скептицизм в отношении идеи коннекционизма.

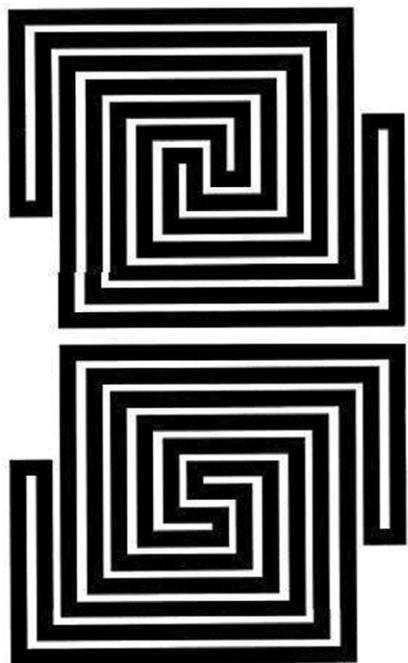
В 1969 г. Минский и Сеймур Пейперт — основатели лаборатории искусственного интеллекта при Массачусетском технологическом институте — написали книгу *Perceptrons\**, в которой излагалась проблема перцептрона, заключающаяся в том, что такие перцептроны, как Mark 1, не способны определить является ли изображение связанным. Эта книга вызвала бурю гнева. Человек может очень легко определить, является ли образ цельным (связанным), и для компьютера эта задача не представляет никакой сложности. Тот факт, что перцептроны не могут этого сделать, многими был воспринят как полный провал.

Суть книги, однако, интерпретировалась шире, чем подразумевали авторы. Теорема Минского и Пейперта применима

---

\* В русском переводе: *Минский М., Пейперт С. Перцептроны. М. : Мир, 1971.*

только к определенному типу нейросетей, называемых упрещающими сетями (к которым, в частности, относится перцептрон Розенблатта), но другие типы нейронных сетей не имеют этого ограничения.

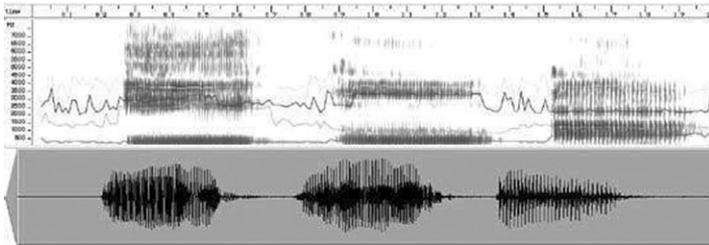


*Два изображения с обложки книги «Перцептроны» (авторы Марвин Мински и Сеймур Паперт). Верхнее изображение: неподключенный образ (темный участок состоит из двух разъединенных частей). Внизу: образ подключенный.*

И все же книга отчасти способствовала ослаблению интереса к нейронным сетям в 1970-х гг. Исследования возобновились в 1980-х гг. в попытках использовать «более реалистичные модели» биологических нейронов и избежать ограничений, накладываемых теоремой Минского—Пейперта. Тем не менее способность новой коры решать ключевую проблему инвариантности по-прежнему не поддавалась воспроизведению в рамках коннекционизма.

## Разреженное кодирование: векторное квантование

В начале 1980-х гг. я начал заниматься проектом, посвященным другой классической проблеме распознавания образов — пониманию человеческой речи. Сначала мы использовали классический подход, заключающийся в прямом программировании знаний о фундаментальных единицах речи, называемых фонемами, и лингвистических правилах связывания отдельных фонем в слова и фразы. Каждая фонема характеризуется определенными частотными параметрами. Например, мы знаем, что такие гласные, как «и» и «а», характеризуются определенными резонансными частотами, называемыми формантами, причем каждой фонеме соответствует определенное соотношение формант. Свистящие звуки, как [з] и [с], характеризуются шумовым импульсом, охватывающим несколько частот.

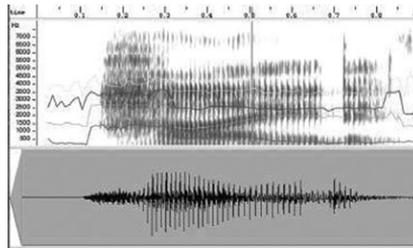


*Спектр трех гласных звуков. Слева направо: звук [i], как в слове «appreciate», [u], как в «acoustic», и [a], как в «ah». По оси Y отложены звуковые частоты. Чем темнее полосы, тем больше энергии соответствует этой частоте.*

Мы регистрировали речь в виде звуковых волн, которые затем конвертировали в частотные полосы, используя частотные фильтры. Результат такой трансформации (спектрограмма) выглядит следующим образом.

Фильтры имитируют функцию улитки человеческого уха на начальном этапе биологического процесса восприятия звука. Компьютерная программа сперва идентифицировала фонемы на основании распознанных частот, а затем на основании характерных последовательностей фонем идентифицировала слова.

Кое-что у нас получилось. Мы смогли натренировать нашу систему на распознавание речи конкретного человека, пользовавшегося небольшим набором слов (несколько тысяч). Далее мы попытались научить ее распознавать десятки тысяч слов, речь разных людей и непрерывную речь (без пауз между словами), но столкнулись с проблемой инвариантности. Разные люди произносят одну и ту же фонему по-разному: в устах одного человека фонема «и» может звучать, как «а» у другого человека. Даже один и тот же человек не отдает себе отчета в том, как он произносит отдельные фонемы. Характер звучания часто зависит от соседних фонем. Некоторые фонемы полностью исчезают. Произношение слов (то есть сочленение фонем) также весьма вариабельно и зависит от контекста. Те лингвистические правила, которые мы заложили в нашу программу, не смогли учесть чрезвычайную вариабельность разговорной речи.



*Спектр слова «hide» («хайд» — прятать). Горизонтальные линии соответствуют формантам — частотам с наиболее высокой энергией.*

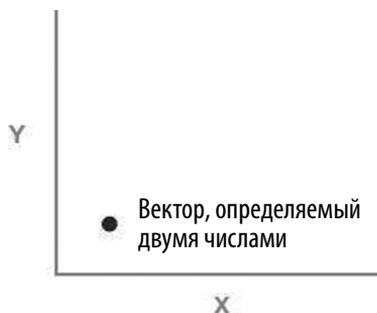
Тогда мне стало ясно, что распознавание образов и понятий человеческим мозгом имеет иерархическую природу. Это совершенно очевидно для человеческой речи, состоящей из сложных иерархических структур. Но что лежит в основе этой структуры? Это был первый вопрос, который я задал себе, собираясь заниматься созданием автоматических систем распознавания разговорной человеческой речи. Звук проникает в ухо как вибрация воздуха и под действием примерно 3000 внутренних волосковых клеток внутри ушной улитки превращается во

множество полос различной частоты. Каждая волосковая клетка настроена на восприятие определенной частоты (звуковые частоты мы воспринимаем как тона), и каждая действует как частотный фильтр, испуская сигнал, когда улавливает звук своей или близкой резонансной частоты. Таким образом, выходящий из улитки уха звук составлен примерно из 3000 отдельных сигналов, каждый из которых характеризуется изменяющейся во времени интенсивностью узкой звуковой полосы (конечно, эти полосы перекрываются).

Даже при условии множественных повторов структур мозга мне кажется совершенно невероятным, что мозг обрабатывает 3000 отдельных звуковых сигналов. У меня вызывала большое сомнение столь низкая эффективность эволюции. Теперь мы знаем, что еще до того, как звуковые сигналы попадают в новую кору, в слуховом нерве происходит очень значительное сокращение объема информации. В наших искусственных системах распознавания речи мы тоже применяли фильтры (сначала 16, а потом 32; больше, как оказалось, использовать бессмысленно). Таким образом, в нашей системе в каждый момент времени входной сигнал был представлен 16 числами. И нам предстояло свести эти 16 потоков данных к единому потоку, сохранив при этом информацию, необходимую для распознавания речи.

С этой целью мы использовали математический метод, называемый векторным квантованием. Итак, представьте себе, что в любой момент времени звук (выходящий из уха к мозгу) представлялся компьютерной программой в виде набора из 16 чисел — выходные сигналы, прошедшие 16 частотных фильтров (в человеческом ухе было бы 3000 чисел, соответствующих выходным сигналам из 3000 волосковых клеток). На языке математики каждый подобный набор чисел (3000 в биологической системе или 16 в компьютерной модели) называют вектором.

Для простоты рассмотрим процесс векторного квантования для вектора из двух чисел. Каждый вектор можно представить себе в виде точки в двумерном пространстве.

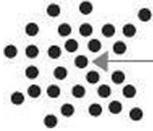


Если у нас имеется множество таких векторов, при нанесении их на график мы можем заметить формирование определенных кластеров.



Для идентификации кластеров нужно решить, сколько их будет. В нашей работе мы обычно допускали существование 1024 кластеров, которые могли пронумеровать и обозначить 10-разрядным двоичным числом (поскольку  $2^{10} = 1024$ ). Наша выборка векторов представляет необходимое нам разнообразие. Сначала мы предполагали, что первые 1024 вектора образуют одноточечные кластеры. Затем мы рассматривали 1025-й вектор и находили ближайшую к нему точку. Если это расстояние больше самого маленького расстояния между двумя любыми точками из 1024, мы считаем эту точку началом нового кластера. Тогда мы «схлопываем» два ближайших кластера в единый кластер (при этом у нас по-прежнему остается 1024 кластера).

После добавления 1025-го вектора один из наших кластеров содержит более одной точки. Мы продолжаем действовать таким же образом, все время сохраняя 1024 кластера. Когда мы добавим все точки, мы можем представить каждый многоточечный кластер в виде геометрического центра точек этого кластера.



Кластер точек; мы можем его изобразить в виде единственной точки, соответствующей геометрическому центру всей этой группы точек

Мы повторяем этот процесс до тех пор, пока не учтем все точки из нашей выборки. Обычно мы укладываем в 1024 ( $2^{10}$ ) кластера миллионы точек (мы также работали с 2048 ( $2^{11}$ ) и с 4096 ( $2^{12}$ ) кластерами). Каждый кластер можно представить в виде единственного вектора, расположенного в геометрическом центре всех точек кластера. Таким образом, сумма расстояний от всех точек кластера до центральной точки минимальна.

В результате применения данного метода вместо миллионов точек, с которыми мы начинали работать, у нас остается всего 1024 точки, оптимальным образом распределенные в пространстве. Неиспользованные части пространства не принадлежат ни одному кластеру.

Далее мы нумеруем наши кластеры (в данном случае приписываем им номера от 0 до 1023). Этот номер является сокращенным, «квантованным» представлением кластера, вот почему данный метод называется методом векторного квантования. Любой новый входящий вектор будет характеризоваться номером кластера, центр которого расположен ближе всего к этому новому вектору.

Теперь мы можем рассчитать расстояние от центра каждого кластера до центра всех других кластеров. Тем самым мы получаем расстояние от этого нового входящего вектора (представленного квантованным значением — номером кластера, к которому новый вектор расположен ближе всего) до всех других

кластеров. Поскольку все точки характеризуются только номерами ближайших кластеров, мы знаем расстояния между конкретной точкой и всеми другими возможными точками.

Я описал суть метода в применении к двумерным векторам, но работа с 16-мерными векторами осуществляется по совершенно аналогичной схеме. Мы выбираем векторы с 16 числами, соответствующими 16 различным частотным полосам, так что каждая точка в нашей системе располагается в 16-мерном пространстве. Нам трудно вообразить пространство, распространяющееся более чем в трех измерениях (ну максимум в четырех, если мы добавим шкалу времени), но у математиков нет таких ограничений.

Применение данного метода позволило нам решить четыре задачи. Во-первых, мы значительно понизили уровень сложности системы. Во-вторых, мы представили 16-мерные данные в виде одномерных. В-третьих, мы усовершенствовали свою способность находить инвариантные черты, поскольку выделяем те участки звукового пространства, которые содержат наибольшее количество информации. Большинство комбинаций частот является физически невозможным или очень маловероятным, поэтому нет смысла выделять одинаковое пространство для вероятных и маловероятных комбинаций входных сигналов. Этот метод позволяет ограничить набор данных равновероятными возможностями. Четвертое достижение заключается в том, что мы можем использовать одномерные распознающие модули, даже если исходные данные являются многомерными. Это самый эффективный подход к использованию имеющихся в настоящее время вычислительных ресурсов.

### **Чтение мыслей с помощью скрытых моделей Маркова**

Метод векторного квантования позволил нам упростить данные таким образом, чтобы выделить ключевые признаки, но нам по-прежнему нужно было научиться представлять иерархию инвариантных признаков, чтобы оценивать новую информацию.

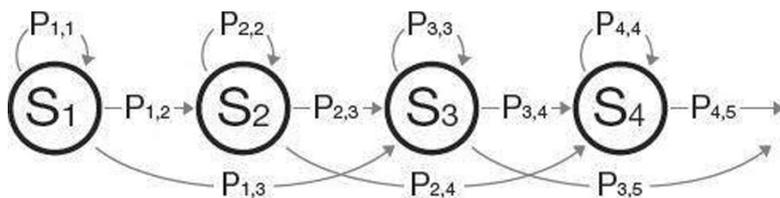
В начале 1980-х гг., когда я уже занимался проблемой распознавания образов на протяжении почти 20 лет, я знал, что одномерное представление данных — самый мощный, эффективный и надежный способ получения инвариантных результатов. В то время мы мало знали о функционировании новой коры, но на основании своих исследований в области распознавания образов я предположил, что мозг, возможно, редуцирует многомерные данные (поступающие от глаз, ушей или кожи) до одномерных, особенно когда речь идет об иерархии понятий в новой коре.

Организация информации в речевых сигналах, как оказалось, представляет собой иерархию образов: каждый образ складывается из линейной последовательности элементов. Каждый элемент образа может быть составлен из образов более низкого порядка или представлять собой базовую единицу сигнала (соответствующую нашим с вами квантованным векторам).

Эта ситуация полностью соответствует модели новой коры, о которой я рассказывал ранее. Таким образом, человеческая речь является продуктом иерархической организации линейных образов. Если бы мы умели анализировать эти образы в мозге говорящего человека, мы могли бы просто сопоставлять каждое новое высказывание с образами в мозге и понимать, что же человек говорит. К сожалению, у нас нет прямого доступа к головному мозгу — мы имеем лишь то, что слышим. Конечно же, в этом и заключается весь смысл разговорной речи — человек передает в словах часть своих мыслей.

Я задал себе вопрос: нет ли какого-то математического метода, который помог бы нам анализировать образы в мозге говорящего человека на основании произнесенных слов? Одной фразы, конечно, было бы недостаточно, но если бы у нас имелся большой набор образцов, возможно, мы смогли бы использовать эту информацию для «прочтения» образов в новой коре говорящего человека (или хотя бы сформулировать математический эквивалент, который позволил бы нам распознавать новые словосочетания).

Часто люди недооценивают мощь математического аппарата; подумайте, ведь поиск самой разной информации за считанные секунды с помощью поисковых программ основан на математическом подходе. Так вот, когда в начале 1980-х гг. я занимался проблемой распознавания речи, оказалось, что для решения стоявших перед нами задач практически идеально подходит так называемый метод скрытых моделей Маркова. Русский математик Андрей Андреевич Марков (1856–1922) создал математическую теорию иерархических последовательностей состояний. Модель основана на возможности перехода из одного состояния в другое по цепочке; если переход совершается успешно, следующее состояние характеризуется более высоким уровнем иерархии. Вам это ничего не напоминает?



*Простой пример одного слоя в скрытой модели Маркова. Символы от  $S_1$  до  $S_4$  обозначают «скрытые» состояния системы. Параметры переходов  $P_{i,j}$  соответствуют вероятности перехода из состояния  $S_i$  в состояние  $S_j$ . Эти вероятности определяются системой путем обучения на тренировочных данных и в рабочем порядке. Новая последовательность (например, новая последовательность звуков) соотносится со значениями  $P_{i,j}$  для определения вероятности воспроизведения моделью этой последовательности.*

В модели Маркова учитываются вероятности переходов из каждого состояния в следующее. В ней рассматривается ситуация, когда система следует такой иерархической линейной цепи состояний, но их невозможно проанализировать напрямую — отсюда название «*скрытые* модели Маркова». На самом нижнем иерархическом уровне подаются сигналы, и это все, что мы можем видеть. Марков предложил математический метод расчета, показывающий, что вероятность каждого перехода должна быть основана на значении выходного сигнала. Этот метод в 1923 г.

усовершенствовал знаменитый английский математик Норберт Винер (1894–1964). В результате стало возможным определять связь между звеньями в цепи Маркова; в частности, все связи, характеризующиеся очень низкой вероятностью, стали считаться несуществующими. Именно так действует и человеческая новая кора: если какие-то связи не используются или используются очень редко, они воспринимаются как маловероятные и со временем исчезают. В нашем случае наблюдаемым выходным сигналом является речевой сигнал, издаваемый говорящим человеком, а состояния в цепи Маркова и связи между ними описывают иерархическую организацию новой коры.

Я задумал систему, в которой мы могли бы собирать образцы человеческой речи, применять метод скрытых моделей Маркова для выявления иерархии состояний с их связями и вероятностями (модель новой коры), а затем использовать эту иерархическую сеть для распознавания новых словосочетаний. Для создания системы, способной распознавать речь разных людей, следовало натренировать скрытые модели Маркова на образцах речи многих людей. Добавляя элементы иерархии для воспроизведения иерархической природы языковой информации, мы получили бы в чистом виде иерархические скрытые модели Маркова (ИСММ).

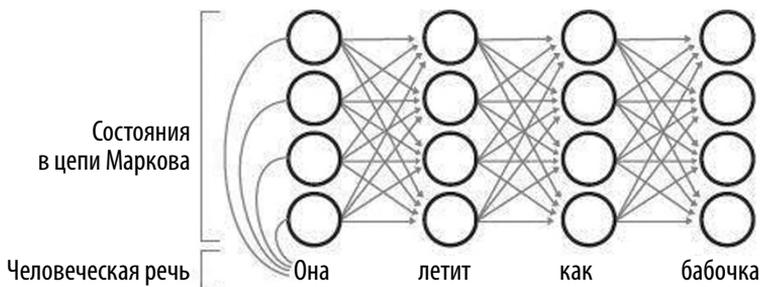
Мои коллеги в компании Kurzweil Applied Intelligence весьма скептически отнеслись к этому плану, учитывая сходство этой модели с самоорганизующимися нейронными сетями, которые уже не были столь популярны и не помогли нам серьезно продвинуться в наших предыдущих исследованиях. Я же указывал, что связи в системе нейронных сетей являются фиксированными и не адаптируются к входным сигналам — веса адаптируются, а связи нет. В системе скрытых моделей Маркова — при условии, что она правильно подготовлена, — исчезают неиспользуемые связи и происходит адаптация.

Я организовал небольшой изолированный проект, функционировавший вне общего направления исследований. Над проектом трудился я сам, один программист, работавший неполный

день, и один инженер-электрик (который должен был создать базу частотных фильтров). К огромному удивлению коллег, наш проект оказался весьма успешным и позволил создать систему, с высокой точностью распознающую речь из большого набора слов.

После этого все последующие системы распознавания речи строились на иерархических скрытых моделях Маркова. Другие исследовательские группы, по-видимому, обнаружили достоинства этой модели независимо, и начиная с середины 1980-х гг. именно этот подход лег в основу большинства автоматизированных систем распознавания речи. Скрытые модели Маркова используются и для создания речи: не забывайте, что иерархическая организация нашей биологической новой коры важна не только для распознавания входных сигналов, но и для создания выходных сигналов, таких как речь или движение.

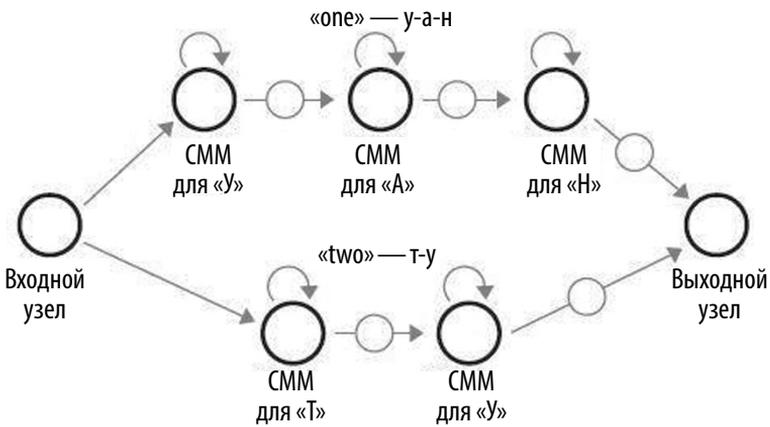
ИСММ также используются в системах, понимающих смысл предложений на более сложном понятийном уровне иерархии.



*Состояния в скрытой модели Маркова и возможные переходы, необходимые для создания последовательности слов в разговорной речи.*

Чтобы понять, как работает метод ИСММ, начнем с рассмотрения сети, состоящей из всех возможных состояний и переходов. Здесь важнейшую роль играет описанный выше метод векторного квантования, поскольку без него пришлось бы рассматривать слишком большое число вариантов.

Вот возможный вариант упрощения исходной топологии.



*Простая скрытая модель Маркова для распознавания двух слов разговорной речи.*

Образцы словосочетаний обрабатываются по очереди. В каждом случае мы многократно изменяем вероятность переходов, чтобы как можно лучше отразить только что обработанный входной сигнал. В системах распознавания речи модели Маркова описывают вероятность обнаружения специфического звукового образа в каждой фонеме, взаимное влияние фонем и вероятный порядок фонем. Кроме того, система может включать и такие более сложные языковые закономерности, как порядок слов, составление фраз и т. д.

Наши первые системы распознавания речи были основаны на специфических правилах, описывающих структуру фонем и предложений, составленных и закодированных лингвистами, но новым системам на основе ИСММ не нужно было объяснять, что в английском языке 44 фонемы, каковы последовательности вероятных векторов для каждой фонемы и что одни последовательности фонем встречаются чаще других. Мы предоставили системе возможность самостоятельно обнаружить эти «правила» в процессе тысяч часов обработки транскрибированной человеческой речи. Преимущество этого подхода, в отличие от

тренировки модели на заранее заданных правилах, заключается в том, что система находит вероятностные закономерности, о которых лингвисты порой даже не подозревают. Мы обратили внимание на то, что правила, усвоенные системой в автоматическом режиме при обработке предоставленных ей данных, не очень сильно отличаются от правил, сформулированных экспертами, но эти отличия могут быть чрезвычайно важными.

После тренировки системы мы приступаем к распознаванию речи, рассматривая альтернативные пути и выбирая наиболее вероятный путь с учетом реальной последовательности входных векторов. Другими словами, если мы видим последовательность состояний, которая с большой вероятностью соответствует данному словосочетанию, мы делаем вывод, что данное словосочетание задается этой кортикальной последовательностью. Такая основанная на ИСММ новая кора содержала метки слов и потому была способна предложить транскрипцию того, что слышала.

Мы смогли усовершенствовать систему путем дальнейших тренировок. Как говорилось выше, на всех иерархических уровнях нашей биологической новой коры процессы распознавания и обучения тоже происходят одновременно.

### **Эволюционные (генетические) алгоритмы**

Еще один важный аспект разработки программы заключается в том, чтобы найти способ установить все те многочисленные параметры, которые контролируют функционирование системы распознавания. Среди этих параметров — допустимое число векторов для этапа векторного квантования, исходная топология иерархических состояний (до того, как в процессе тренировки ИСММ ликвидирует неиспользуемые состояния), порог распознавания на каждом иерархическом уровне, параметры, определяющие величину сигналов, и многое другое. Мы можем использовать интуитивный подход, но результаты будут далеки от оптимальных.

Мы называем данные параметры «Божьими параметрами», поскольку они устанавливаются до включения самоорганизующегося метода определения топологии скрытых моделей Маркова (или, в случае биологической новой коры, до того как человек начинает заучивать урок путем создания связей между нейронами новой коры). Возможно, название неправильное, поскольку эти исходные данные, записанные в ДНК, определены ходом биологической эволюции, хотя некоторые усматривают в этом процессе божий промысел (я, безусловно, считаю эволюцию духовным процессом и обращаюсь к этой теме в девятой главе).

Когда мы начали устанавливать «Божьи параметры» при моделировании иерархических систем обучения и распознавания, мы опять-таки обратились к природе и попытались оценить их путем моделирования процесса эволюции. Мы использовали так называемые генетические (эволюционные) алгоритмы, основанные на моделировании полового размножения и мутаций.

Я опишу этот процесс в упрощенном виде. Сначала мы определяли способ кодирования возможных решений данной проблемы. Если проблема заключалась в оптимизации исходных параметров какой-то цепи, мы определяли список всех параметров (приписывая каждому специфический объем информации (число битов)) этой цепи. Этот список является своеобразным генетическим кодом генетического алгоритма. Мы случайным образом создавали тысячи и более таких генетических кодов. Каждый код (который представляет собой один набор исходных параметров) рассматривается в качестве модельного организма («решения»).

Далее мы оценивали каждый модельный организм в модельной среде, используя описанный метод определения каждого набора параметров. Эта оценка — ключ к успеху генетического алгоритма. В нашем примере мы проверяли все программы, созданные на основе наших параметров, и оценивали их по определенным критериям (решена ли задача, как быстро она решается

и т. д.). Лучшим организмам (лучшим «решениям») разрешалось выжить, а остальные были элиминированы.

Далее мы заставляли выжившие организмы размножаться до определенной численности «особей». Для этого мы имитировали половое размножение: иными словами, мы создавали потомство, в котором каждая особь наследовала одну часть генетического кода от одного родителя, а другую — от другого. Обычно в нашей системе нет никакой разницы между женскими и мужскими организмами, потомство может давать любая пара условных родителей, так что речь фактически идет об однополых браках. В этой связи данный процесс не столь интересен, как в природе. Для нас было важно только наличие двух родителей. По мере размножения наших модельных организмов мы допускали появление в их хромосомах случайных мутаций (изменений).

Итак, мы описали появление одного поколения модельных организмов. Далее мы повторяли все те же процессы для каждого следующего поколения: определяли, насколько улучшился дизайн (то есть рассчитывали среднее улучшение оценки для всех выживших организмов). Когда результаты оценки организмов из двух последовательных поколений практически переставали улучшаться, мы останавливали цикл и использовали лучшие образцы из последнего поколения (подробнее о генетическом алгоритме см. комментарий<sup>10</sup>).

Суть генетического алгоритма (ГА) заключается в том, что создатели программы не программируют решение, а позволяют ему сформироваться в результате многократных имитаций конкуренции и улучшений. Биологическая эволюция — процесс умный, но медленный, и мы попытались этот медленный процесс ускорить. Компьютер действует достаточно быстро, чтобы имитировать смену множества поколений за несколько часов или дней, а наши эксперименты длились несколько недель, так что мы могли наблюдать смену сотен тысяч поколений. Но этот многократно повторяющийся процесс мы должны

были воспроизвести лишь однажды. Как только мы запустили процесс искусственной эволюции, достаточно быстро смогли применить полученные усовершенствованные правила к решению реальных проблем. В нашей системе распознавания речи мы использовали их для обнаружения исходной топологии сети и других критических параметров. Итак, мы использовали две самоорганизующиеся системы: ГА для симуляции биологической эволюции, что позволило определить специфический дизайн коры, и ИСММ для симуляции организации коры, определяющей процесс обучения у человека.

Еще одно важное требование для успешного применения ГА заключается в обеспечении надежного метода оценки всех возможных решений. Эта оценка должна быть быстрой, поскольку в каждом поколении приходится анализировать тысячи возможных вариантов. Метод ГА прекрасно подходит для нахождения точного аналитического значения в задачах с очень большим числом переменных. Например, при проектировании моторов приходится учитывать более сотни переменных и удовлетворять десяткам различных условий. Использование инженерами General Electric метода ГА при создании реактивного двигателя позволило добиться более точного соответствия необходимым требованиям, чем применение традиционных методов.

При использовании ГА, однако, следует с особым вниманием относиться к тому, о чем вы собираетесь спросить. Этот алгоритм был разработан для решения статистической задачи о стопке брусков, для которой он позволяет найти идеальное решение, но за тысячи ходов. Программисты забыли включить в программу функцию минимизации операций.

Проект «Электрическая овца» Скотта Дрейва — основанная на генетическом алгоритме программа, которая умеет рисовать. Функция оценки использует в качестве оценщиков многие тысячи людей. Рисунки изменяются во времени — вы можете увидеть их на сайте [electricsheep.org](http://electricsheep.org).

В системах распознавания речи сочетание генетических алгоритмов и скрытых моделей Маркова дало возможность получить отличные результаты. Симуляция эволюции с помощью ГА в значительной степени улучшила производительность сетей ИСММ. Эволюция позволила достичь результата, который намного превзошел исходный вариант, основанный на нашей интуиции.

Далее мы экспериментировали, вводя в систему небольшие изменения. Например, мы создавали некоторые пертурбации (случайные небольшие изменения) входного сигнала. Другое изменение состояло в том, что мы заставляли соседние марковские цепи «протекать» друг в друга: результаты одной модели влияли на «соседние» модели. В то время мы не отдавали себе в этом отчета, однако производимые нами изменения действительно напоминали процессы, происходящие в структурах биологической коры.

В первую очередь такие изменения снижали эффективность системы (точность распознавания). Однако если мы повторяли эволюционный процесс (то есть перезапускали генетический алгоритм) при наличии подобных изменений, это позволяло адаптировать систему и оптимизировать ее к данным условиям. Обычно это приводило к восстановлению эффективности. Если затем мы удаляли внесенные изменения, эффективность вновь падала, поскольку система эволюционировала для их компенсации. Адаптированная система становилась зависимой от внесенных изменений.

В частности, небольшие случайные изменения на входе позволяли повысить эффективность системы (после повторного проведения ГА). Причина заключается в проблеме так называемой сверхподгонки, характерной для самоорганизующихся систем. Существует опасность, что такая система в основном будет настроена на специфические примеры, которые использовались для ее обучения. Внесение случайных изменений на входе позволяет добиться большей инвариантности данных для

обучения системы. Но этот метод работает только в том случае, если в новых условиях мы вновь запускаем генетический алгоритм.

Тут возникает дилемма в нашем понимании функционирования биологических кортикальных цепей. Например, было подмечено, что иногда происходит «протечка» кортикальных контактов, что объясняется самой природой биологических связей: электрохимические свойства аксонов и дендритов зависят от электромагнитного воздействия соседних контактов. Представьте, что нам удастся провести эксперимент с реальным мозгом, в котором подобная возможность устранена. Сделать это трудно, но не невозможно. Представьте, что в результате подобного эксперимента мы обнаружили, что кортикальные цепи работают менее эффективно. Нам пришлось бы заключить, что это явление — очень умное эволюционное приспособление, необходимое для достижения максимальной эффективности функции новой коры. А еще пришлось бы признать, что иерархические модели, описывающие восходящие потоки образов и нисходящие предсказательные сигналы, на самом деле гораздо сложнее, поскольку должны учитывать еще и взаимное влияние контактов.

Однако этот вывод не обязательно верен. Вернемся к нашему эксперименту с имитацией новой коры на основе иерархических скрытых моделей Маркова, в котором предусмотрим модификацию, напоминающую взаимное влияние нейронов. Если мы в этой системе повторим эволюционный процесс, эффективность системы восстановится (поскольку эволюция адаптируется к изменениям). Если же мы устраним взаимное влияние нейронов, эффективность вновь понизится. В биологической системе эволюция («настоящая» биологическая эволюция) действительно «прогонялась» при наличии подобного фактора. Таким образом, биологическая эволюция выбрала такие параметры системы, которые зависят от данных факторов, и их изменение негативно скажется на эффективности системы, если

только мы не «перезапустим» эволюционный процесс. В компьютерном мире сделать это несложно — здесь эволюционный процесс занимает дни или недели, но в природе для этого потребовались бы десятки тысячелетий.

Так каким же образом мы можем установить, является ли какой-то характерный признак строения биологической новой коры жизненно важной инновацией, внесенной биологической эволюцией (то есть этот признак важен для нашего уровня интеллекта), или это лишь артефакт строения, от которого система теперь зависит, но без которого вполне могла бы развиваться дальше? Чтобы ответить на этот вопрос, достаточно просто запустить имитацию эволюционного процесса при наличии и при отсутствии этих деталей строения (например, при наличии и при отсутствии протечки контактов). Более того, мы можем проследить и за биологической эволюцией, наблюдая за колонией микроорганизмов, у которых процесс воспроизводства измеряется часами, но со сложными организмами, такими как человек, проделать такой эксперимент не представляется возможным. Это еще один из многих недостатков биологии.

Вернемся к нашим экспериментам по распознаванию речи. Мы обнаружили, что результаты получаются лучше, если мы запускаем эволюцию (то есть ГА) *независимо* в исходной системе для 1) иерархических скрытых моделей Маркова для внутренней структуры фонем и для 2) ИСММ для структуры слов и фраз. На обоих уровнях использовались ИСММ, но метод ГА создавал вариации на стыке этих двух уровней. Метод позволил моделировать такие явления, как размывание фонем, что часто случается при сочленении некоторых слов.

Вероятно, подобные явления происходят в различных отделах биологической новой коры, и в зависимости от типов образов, которые они обрабатывают, в них эволюционировали небольшие различия. Хотя все эти отделы используют один и тот же базовый алгоритм, у биологической эволюции было достаточно времени для тонкой настройки структуры каждого

отдела, позволяющей оптимально обрабатывать образы соответствующего типа. Однако, как я уже отмечал ранее, нейробиологи и неврологи обнаружили в этих областях значительную пластичность, подтверждающую идею общего неокортикального алгоритма. Если бы фундаментальные принципы функционирования всех отделов коры кардинально различались, подобная взаимозаменяемость отделов коры была бы невозможна.

Используя описанную комбинацию методов самоорганизации системы, мы добились больших успехов. Наши системы распознавания речи впервые смогли разбирать непрерывную речь, состоящую из относительно неограниченного набора слов. Была достигнута высокая точность распознавания речи разных людей, говоривших с разным акцентом и на разных диалектах. Сейчас, когда я пишу эту книгу, последним достижением является программа под названием *Dragon Naturally Speaking* (версия 11.5) для PC компании Nuance (ранее Kurzweil Computer Products). Тем, кто сомневается в возможностях систем распознавания речи, я советую испытать этот продукт — через несколько минут настраивания на ваш голос при практически непрерывной речи и почти неограниченном наборе слов точность распознавания часто достигает 99%. *Dragon Dictation* — более простое, но тоже удивительное бесплатное приложение для айфона, не требующее голосового тренинга. Персональный помощник Сири, установленный на современных айфонах, использует ту же самую технологию распознавания речи, содержащую дополнительный модуль, воспринимающий разговорную речь.

Высокая эффективность подобных систем — заслуга математиков. Вместе с математиками мы моделируем процессы, происходящие в новой коре говорящего человека (хотя у нас нет прямого доступа к его головному мозгу), что является ключевым этапом в распознавании того, что произносит говорящий, и (в случае таких систем, как Сири) того, что эти выражения

означают. Возникает вопрос: если бы мы могли следить за тем, что происходит в новой коре говорящего, нашли бы мы корреляцию со связями и весами соответствующих иерархических скрытых моделей Маркова, рассчитанных нашей программой? Скорее всего, нам не удалось бы обнаружить точное соответствие — нейронные структуры в любом случае во многом отличаются от компьютерных моделей. Однако следует подчеркнуть, что между биологической системой и нашими попытками ее имитировать должны существовать важные математические соответствия — в противном случае эти модели не работали бы так хорошо.

## ЛИСП

ЛИСП (LISP, LISt Processor — процессор списков) — язык программирования, созданный пионером в области ИИ Джоном Маккарти в 1958 г. Как следует из названия, этот язык оперирует списками. Каждое утверждение в нем представляет собой список элементов, а каждый элемент — это либо еще один список, либо «атом» — неделимая единица, представляющая собой число или символ. Список внутри списка тоже может быть списком, так что ЛИСП способен на рекурсию. Кроме того, предложения являются рекурсивными, если один список содержит в себе другой, тот — третий и т. д. до тех пор, пока не будет определен исходный список. За счет такой структуры ЛИСП позволяет осуществлять иерархические построения. Список может быть условным, то есть «возбуждается» лишь в том случае, когда удовлетворяются все составляющие его элементы. Иерархию таких условий можно использовать для идентификации образов с возрастающим уровнем сложности.

ЛИСП был чрезвычайно популярен в среде специалистов в области ИИ в 1970-х и начале 1980-х гг. В начале данного периода поклонники этого языка считали, что он отражает способ работы человеческого мозга и что с его помощью можно легко и эффективно закодировать любой мыслительный процесс.

Возник даже «мини-бум» компаний по созданию «искусственного интеллекта», предлагавших программы-переводчики и другие программные продукты, однако в середине 1980-х гг. стало ясно, что сам по себе язык ЛИСП не может решить проблему создания разума, и инвестиционный пузырь лопнул.

Впрочем, нельзя сказать, что поклонники ЛИСП полностью заблуждались. На самом деле каждый распознающий модуль новой коры можно рассматривать как предложение на языке ЛИСП — в каждом модуле содержится список элементов, причем каждый элемент тоже может представлять собой список. Таким образом, новая кора действительно занимается обработкой списков символов, что очень сильно напоминает процесс, описываемый программой ЛИПС. Более того, новая кора одновременно обрабатывает 300 млн таких «предложений».

Однако программе ЛИПС не хватало двух очень важных элементов, одним из которых является возможность обучения. Программу строку за строкой кодировали программисты. Делались попытки автоматического кодирования программы с помощью нескольких разных методов, но этому подходу уделялось недостаточно внимания в рамках общей концепции языка. Напротив, новая кора программирует сама себя, наполняя свои «предложения» (списки) значимой информацией на основе собственного опыта и сигналов обратной связи. В этом и заключается основной принцип работы новой коры: каждый из ее распознающих модулей (которым соответствуют отдельные предложения на языке ЛИСП) способен составлять собственный список и связываться с «вышестоящим» и «нижестоящим» списком. Второй недостающий элемент — параметры величины сигнала. Можно создать вариант ЛИСП (на языке ЛИСП), который учитывал бы такие параметры, но в базовой структуре языка их нет.

ЛИПС соответствует исходной философии в сфере ИИ, которая состояла в поиске разумных решений проблем и их кодировании напрямую на языке программирования. Первая

попытка создания самоорганизующейся системы, способной обучаться на собственном опыте, — нейронной сети — оказалась неудачной, поскольку не давала возможности модифицировать топологию системы в процессе обучения. Иерархические модели Маркова обладают такой возможностью за счет механизма отсечения неиспользуемых связей. На сегодня модель ИСММ и ее математические аналоги составляют важнейшую часть систем ИИ.

Следствием наблюдаемого сходства ЛИПС и структуры списков в новой коре является аргумент, выдвигаемый теми, кто считает мозг слишком сложной структурой, не подвластной нашему пониманию. Эти люди подчеркивают, что в головном мозге имеются триллионы нервных контактов, и поскольку каждый из них спроектирован в специфическом месте, они эквивалентны триллионам строчек компьютерной программы. Как я уже писал, по оценкам, в новой коре содержится около 300 млн процессоров образов — или 300 млн списков, в которых каждый элемент списка связан с другим списком (или, на нижнем понятийном уровне, с поступающими извне базовыми образами). Но 300 млн строчек на языке ЛИПС — очень большое число, таких гигантских компьютерных программ пока не существует.

Однако не следует забывать, что эти списки на самом деле не определены в исходном плане нервной системы. Мозг сам их создает и автоматически осуществляет связи между ними на основе пережитого опыта. В этом основной секрет новой коры. Процессы, за счет которых происходит эта самоорганизация, значительно проще, чем 300 млн процессоров новой коры. Эти процессы определены в геноме. Как я расскажу в одиннадцатой главе, объем значимой информации (после обратимого сжатия) в геноме, относящейся к функции головного мозга, составляет примерно 25 млн байт, что эквивалентно менее чем миллиону строк программного кода. Причем в реальности алгоритм еще проще, поскольку бóльшая часть из этих 25 млн байт

генетической информации относится к биологическим нуждам нейронов, а не к их способности обрабатывать информацию. Но и с 25 млн байт информации мы в состоянии справиться.

## Иерархические системы памяти

Как я рассказывал в третьей главе, Джеф Хокинс и Дайлип Джордж в 2003–2004 гг. создали модель новой коры, которая включала в себя иерархические списки. Эта модель описана в книге Хокинса и Блексли *On Intelligence*, опубликованной в 2004 г.\* Более новую и очень элегантную версию теории иерархической временной памяти можно найти в докторской диссертации Дайлипа Джорджа (2008)<sup>11</sup>. Компания Numenta использовала эту модель в системе под названием NuPIC (Numenta Platform for Intelligent Computing) и включила в программы распознавания образов и извлечения данных для таких клиентов, как Forbes и Power Analytics Corporation. Проработав какое-то время с Numenta, Джордж занялся созданием новой компании, названной Vicarious Systems, при поддержке фонда Founder Fund (основанного Питером Тилем, также поддерживающим Facebook, и Шоном Паркером, первым президентом Facebook) и фонда Good Ventures, которым руководит сооснователь Facebook Дастин Москович. Джордж сообщает о значительных успехах автоматического моделирования, усвоения и распознавания информации с высокой степенью иерархической сложности. Он называет свою систему «рекурсивной кортикальной сетью» и планирует использовать ее, среди прочего, для медицинской диагностики и робототехники. Метод иерархических скрытых моделей Маркова по своей математической природе очень близок к этим иерархическим системам памяти, особенно если мы позволяем ИСММ самостоятельно образовывать связи между распознающими модулями. Как было отмечено в третьей главе, ИСММ позволяет учитывать ожидаемое

---

\* Книга вышла в русском переводе: Хокинс Д., Блексли С. Об интеллекте: М. : Вильямс, 2007.

распределение величин всех входных сигналов в расчете вероятности соответствующего образа. Не так давно я организовал новую компанию, Patterns, Inc., которая будет заниматься созданием самоорганизующихся иерархических моделей новой коры на основе ИСММ и родственных методов для распознавания человеческой речи. Система, над которой мы работаем, будет постоянно читать различные материалы, включая «Википедию» и другие источники информации, а также слушать все, что мы говорим, и следить за всем, что мы пишем (конечно, если мы позволим). Задача заключается в создании помощника, способного отвечать на ваши вопросы — еще *до того*, как вы их сформулируете, — снабжать вас полезной информацией и направлять вас в течение дня.

### **Раздвигаем границы ИИ. Повышение компетентности**

1. Длинная утомительная речь, как пенистое украшение пирога\*.
2. Предмет детской одежды, возможно, во время путешествия на корабле.
3. Виновный в поедании воинов короля Хротгара на протяжении двенадцати лет; дело поручено воину Беовульф<sup>\*\*</sup>.
4. Процесс постепенного развития в мозге или в организме при беременности<sup>\*\*\*</sup>.
5. Национальный День учителя и день дерби в Кентукки.
6. Он, по выражению Вордсворта, парит, но не улетает<sup>\*\*\*\*</sup>.

---

\* Игра слов: слово *harangue* («разглагольствование») созвучно слову *meringue* («безе») (*англ.*).

\*\* Сюжет англосаксонской эпической поэмы VII века о тролле-людоеде Гренделе и герое Беовульфе.

\*\*\* В английском языке «гестация» — не только беременность, но и созревание в более общем смысле слова, например созревание идеи.

\*\*\*\* Уильям Вордсворт (1770–1850), стихотворение «Жаворонку»: Так и мудрец парит, вдале не стремясь / И в небе с домом сохраняя связь! (перевод М. Зенкевича).

7. Слово из четырех букв, обозначающее железную накладку на копыте лошади и коробку для карт в казино.
8. В третьем акте оперы Верди, написанной в 1846 г., этот бич божий смертельно ранен своей любовницей Одабеллой.

*Примеры из викторины «Джеопарди!», на все вопросы которой Ватсон дал правильные ответы: разглагольствование, фартук, Грендель, гестация, май, жаворонок, shoe (башимак, подкова). В ответ на 8-й вопрос Ватсон ответил: «Это Аттила?» Его попросили уточнить, тогда он сказал: «Это Аттила, гунн?» — что сочли правильным ответом.*

Техника, которую компьютер применяет для ответов на вопросы «Джеопарди!», очень напоминает мою собственную. Машина начинает поиск от ключевого слова в вопросе, а затем прочесывает свою память (в случае Ватсона это 15-терабайтный банк человеческих знаний) для выявления кластеров ассоциаций с этим словом. Она тщательно проверяет наилучшие совпадения с полным контекстом вопроса: категория и тип ответа, зашифрованные в вопросе время, место и пол и т. д. И когда машина чувствует себя достаточно «уверенной», она подает сигнал. Для игрока в «Джеопарди!» это мгновенный и интуитивный процесс, но я уверен, что в моей голове происходит что-то очень похожее.

*Кен Дженнингс, чемпион викторины «Джеопарди!»,  
проигравший Ватсону*

Я приветствую наших новых компьютерных повелителей.

*Кен Дженнингс, перефразируя Симпсонов,  
после поражения Ватсону*

Бог мой, [Ватсон] умнее отвечает на вопросы «Джеопарди!», чем средний игрок. Он паразитально умен.

*Себастьян Трун, бывший директор  
лаборатории в Стэнфорде*

Ватсон ничего не понимает. Это огромный паровой коток.

*Ноам Хомский*

Искусственный интеллект повсюду вокруг нас. Простое общение с кем-то по электронной почте или мобильному телефону основано на передаче информации с помощью разумных алгоритмов. Практически любой продукт, который мы сегодня используем, спроектирован при сотрудничестве человека и искусственного интеллекта и создан на автоматизированном заводе. Если бы все системы ИИ завтра объявили забастовку, наша цивилизация покачнулась бы: мы не смогли бы получить деньги в банке, да и сами деньги исчезли бы, коммуникации, транспорт и производство — все бы остановилось. К счастью, наши разумные машины пока еще не настолько разумны, чтобы выкинуть подобный фокус.

На сегодня новое в сфере ИИ — это потрясающие возможности доступных для широкой публики приложений. Например, подумайте о самодвижущихся автомобилях Google (которые уже пробежали более 200 тыс. миль по большим и малым городам): эта технология позволит уменьшить число аварий, повысит пропускную способность дорог, избавит людей от рутинной водительской заботы и даст множество других важных преимуществ. Сегодня машины без водителя могут с некоторыми ограничениями передвигаться по общественным дорогам Невады, однако их повсеместное широкое распространение ожидается не раньше конца десятилетия. А вот технологии, которые следят за дорогой и предупреждают водителей о возможной опасности, уже установлены на многих моделях автомобилей. Одна такая технология отчасти основана на успешной модели обработки зрительных образов в головном мозге, предложенной Томазо Поджо из Массачусетского технологического института. Эта система под названием Mobil Eye разработана бывшим аспирантом Поджо Амном Шашуа. Система предупреждает водителя о таких опасностях, как возможное столкновение или наличие на дороге ребенка; она уже установлена на автомобилях марок Volvo и BMW.

В данном разделе книги я подробнее остановлюсь на технологии распознавания речи, на то у меня есть несколько причин. Нет

ничего удивительного в том, что иерархическая структура языка отражает иерархическую структуру мышления. Разговорная речь стала нашей первой технологией, письменный язык — второй. Моя собственная работа в области ИИ в значительной степени связана с изучением языка. Наконец, язык — очень мощное орудие. Ватсон прочел сотни миллионов страниц электронных источников информации и овладел содержащимся там материалом. Когда-нибудь машины будут способны овладевать всей существующей в Интернете информацией, которая объединяет практически все знания, накопленные нашей цивилизацией.

Английский математик Алан Тьюринг (1912–1954) разработал тест, теперь носящий его имя, который основан на способности компьютера вести беседу на человеческом языке с помощью письменных сообщений<sup>12</sup>. Тьюринг полагал, что в языке отражается весь человеческий разум и что никакая машина не сможет пройти тест, просто овладев языковыми приемами. Хотя в тесте используется письменная речь, Тьюринг считал, что машина смогла бы пройти его лишь при условии, что будет обладать разумом, эквивалентным разуму человека. Критики утверждали, что истинный тест на наличие у машины «человеческого» разума должен включать в себя также зрительную и слуховую компоненты<sup>13</sup>. Поскольку многие из созданных мной систем ИИ обучают компьютер воспринимать и обрабатывать человеческую речь, форму букв и музыкальные звуки, возможно, вы предполагаете, что я встану на защиту этой более полной версии теста для оценки интеллекта. Однако я согласен с тем, что исходной версии теста Тьюринга вполне достаточно — добавление зрительных или слуховых сигналов на входе или на выходе на самом деле совсем не усложняет прохождение теста.

Не нужно быть экспертом в области ИИ, чтобы оценить эффективность Ватсона в игре «Джеопарди!» Хотя я достаточно хорошо понимаю методологию, заложенную в основу действия его ключевых подсистем, это не ослабляет мою эмоциональную реакцию на то, что он (оно?) может делать. Даже полное

понимание принципов работы всех подсистем (которого нет ни у кого) не помогает предсказать реакцию Ватсона на конкретную ситуацию. Машина содержит сотни взаимодействующих между собой подсистем, каждая из них одновременно прорабатывает миллионы альтернативных гипотез, так что предсказать ответ всей системы невозможно. Анализ мыслительного процесса, происходящего «в голове» Ватсона за три секунды при ответе на вопрос викторины, у человека занял бы несколько столетий.

Что касается моей собственной истории, в конце 1980-х и в 1990-х гг. мы начали заниматься внедрением систем распознавания человеческой речи в некоторых областях. С одной из наших систем, названной Kurzweil Voice, можно было поговорить о чем угодно, и она могла применяться для редактирования документов. Например, вы могли попросить ее передвинуть в определенное место в тексте третий параграф с предыдущей страницы. В этой ограниченной сфере машина работала достаточно хорошо. Мы также создали систему, обладавшую медицинскими знаниями, которая позволяла врачам диктовать результаты обследования пациентов. Эта машина обладала довольно обширными знаниями в области радиологии и патологии, так что она могла задать врачу вопрос, если что-то в тексте было неясно, и направляла врача по ходу составления отчета. Эта система стала основой многомиллиардного бизнеса компании Nuance.

Понимание естественной речи, особенно в качестве приложения к автоматическим системам распознавания речи, теперь стало элементом серийной продукции. В то время, когда я писал эту книгу, автоматизированный персональный помощник Сири, установленный на модели айфона 4S, произвел фурор в мире сотовых телефонов. Вы можете спросить или попросить Сири практически обо всем, что должен знать и уметь каждый уважающий себя смартфон, например: «Где здесь поблизости индийский ресторан?», или «Сообщи моей жене, что я уже иду», или «Что думают люди о новом фильме Брэда Питта?» — и практически всегда Сири исполняет поручение. Сири может в некоторой степени

поддерживать беседу на общие темы. Если вы спросите ее, в чем заключается смысл жизни, она ответит: «42». Поклонники фильма «Автостопом по галактике» знают, что это «ответ на главный вопрос о жизни, вселенной и всяком таком». На вопросы на знание (включая вопрос о смысле жизни) может ответить описанная ниже программа Wolfram Alpha. Существует уже целый мир «чат-ботов», которые нужны только для того, чтобы болтать. Если вы захотите поболтать с нашим чатботом по имени Рамона, зайдите на сайт [KurzweilAI.net](http://KurzweilAI.net) и кликните по ссылке *Chat with Ramona*.

Люди иногда жалуются, что Сири не может ответить на некоторые вопросы, но обычно это те же самые люди, что частенько недовольны и человеческим сервисом. Иногда я предлагаю решить вопрос совместными усилиями, и часто результат бывает лучше, чем они ожидали. Эти жалобы напоминают мне анекдот о собаке, которая играет в шахматы. На вопрос недоверчивого зрителя владелец собаки отвечает: «Да, это правда, она умеет играть, но слаба в эндшпиле». Тот факт, что широкая публика имеет возможность общаться со своими карманными компьютерами на разговорном языке, открывает новую эпоху. Очень часто люди недооценивают значение технологии первого поколения из-за ее ограничений. Но через несколько лет, когда технология уже хорошо работает, люди по-прежнему не признают ее значение, поскольку теперь она уже не нова. Поэтому следует сказать, что для продукта первого поколения Сири работает очень хорошо и, очевидно, будет работать еще лучше.

Сири использует технологию распознавания речи на основе иерархических моделей Маркова, разработанную компанией Nuance. Расширения для использования разговорной речи впервые были созданы в рамках проекта CALO<sup>14</sup>, финансируемого агентством DARPA\*. Сири снабжена собственной технологией

---

\* DARPA (Defense Advanced Research Projects Agency) — агентство Министерства обороны США, отвечающее за разработку новых технологий для использования в вооруженных силах.

компании Nuance, кроме того, компания предлагает очень похожую технологию под названием Dragon Go<sup>15</sup>.

Методы, используемые для понимания разговорной речи, очень напоминают иерархические скрытые модели Маркова, да и сами ИСММ часто применяются для данных приложений. Хотя иногда в подобных системах не указано, что они используют скрытые модели Маркова или иерархические скрытые модели Маркова, математические основы их функционирования практически идентичны. Все эти системы применяют иерархию линейных последовательностей, каждый элемент которых имеет вес, контакты самостоятельно адаптируются, а вся система в целом самостоятельно организуется в процессе обучения. Обычно обучение продолжается постоянно на протяжении всего времени работы системы. Этот подход отражает иерархическую структуру разговорного языка — это естественное восхождение по понятийной иерархической лестнице от звуков к словам и далее к словосочетаниям и сложным речевым построениям. Имеет смысл применять ГА для параметров, контролирующих конкретный алгоритм обучения в таких классах иерархических систем, и находить оптимальные элементы алгоритма.

За последнее десятилетие возникли новые способы создания подобных иерархических структур. В 1984 г. Дуглас Ленат начал реализацию проекта Сус (от enCYClopedic) с целью разработки правил кодирования «бытовых знаний». Эти правила были организованы в виде сложной иерархической структуры, и каждое правило, опять же, включало в себя линейную последовательность утверждений. Например, одно такое утверждение может информировать, что у собаки есть лицо (морда). Далее Сус выводит более общее правило о структуре лиц: на лице есть два глаза, нос и рот и т. д. Нам не нужно иметь отдельное правило для описания собачьей морды и отдельное — для кошачьей, но мы, конечно же, можем создать дополнительное правило, описывающее различие между кошачьей и собачьей мордой. В системе также реализуется следственный механизм: если

есть правила, утверждающие, что кокер-спаниель — собака, что собака — животное и что животные потребляют пищу, то на вопрос, питается ли кокер-спаниель, система ответит утвердительно. За прошедшие с тех пор три десятилетия были затрачены тысячи человеко-лет трудов и написано и проверено более миллиона таких утверждений. Интересно, что язык, на котором пишут правила для Суc (его называют СуcL), практически идентичен языку ЛИСП.

Тем временем приверженцы альтернативной теории мыслительного процесса считают, что наилучший подход к пониманию разговорной речи и вообще к созданию разумных систем заключается в автоматизированном обучении системы на многочисленных примерах понятий и явлений. Мощный пример системы такого рода — программа-переводчик Google Translate, которая способна переводить с 50 языков на 50 языков. Это подразумевает 2500 направлений перевода, хотя в большинстве пар языков программа не переводит напрямую с первого языка на второй, а переводит сначала с первого языка на английский, а уже с него на второй язык. Это приводит к сокращению числа направлений перевода всего до 98 (и еще небольшого числа языковых пар, в которых перевод осуществляется напрямую). Данная программа перевода не использует грамматических правил, а создает обширные базы данных для всех пар на основе переведенных документов из программного продукта Rosetta stone («Розеттский камень»)\*. Для шести официальных языков ООН\*\* Google использует документы ООН, поскольку они всегда публикуются на шести языках. Для других языков применяются другие источники.

---

\* Компьютерная программа обучения языкам; название происходит от Розеттского камня с выбитыми на трех языках текстами; сопоставление текстов положило начало расшифровке древнеегипетских иероглифов.

\*\* Шесть официальных языков ООН: английский, арабский, испанский, китайский, русский и французский.

Результаты такой работы бывают весьма впечатляющими. DARPA организует ежегодные соревнования для выявления лучших автоматических систем перевода для различных пар языков, и Google Translate часто выигрывает эти соревнования для определенных пар, превосходя системы, напрямую разработанные лингвистами.

За последнее десятилетие было сделано два важных открытия, которые в значительной степени повлияли на развитие систем понимания разговорной речи. Первое имеет отношение к иерархическому строению систем. Несмотря на то что подход Google заключается в поиске ассоциаций между линейными последовательностями слов в разных языках, иерархическая структура языка неизбежно влияет на этот процесс. Системы, применяющие метод иерархического обучения (такие как иерархические скрытые модели Маркова), обеспечивают гораздо более высокую эффективность. Однако создание таких систем — далеко не автоматический процесс. Что верно для людей, которые за один раз осваивают один иерархический уровень понятий, то верно и для компьютеров, так что процессом обучения необходимо тщательно руководить.

Другое открытие заключается в том, что для усвоения основного массива знаний хорошо работают правила, сформулированные людьми. Для перевода коротких фрагментов текста такой подход часто обеспечивает более точный результат. Например, DARPA выше оценило переводы коротких текстов с помощью китайско-русского словаря, основанного на правилах, чем переводы Google Translate. Что же касается других элементов языка, к которым относятся миллионы редких фраз и заключенных в них понятий, тут точность переводов, основанных на правилах, непозволительно низкая. Если построить график точности понимания разговорной речи от количества данных, на которых обучалась система, системы, основанные на правилах, сначала показывают более высокую эффективность, но точность распознавания не поднимается выше 70%. Напротив,

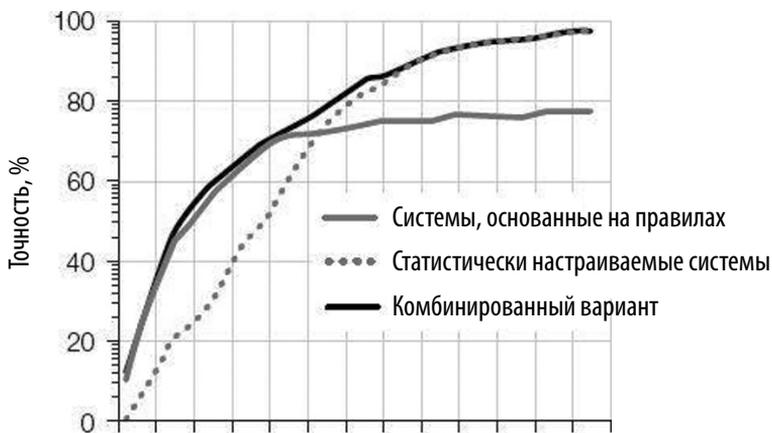
статистические системы достигают точности 90%, но для этого им нужно «изучить» очень большой массив данных.

Часто нам требуется обеспечить приемлемую эффективность при небольшом объеме обучающих данных, но предусмотреть повышение точности при дополнительных тренировках системы. Быстрое достижение средней эффективности позволяет перейти к автоматическому режиму сбора тренировочных данных при каждом использовании. Таким образом, в процессе применения системы происходит и ее обучение, что приводит к постепенному повышению точности результатов. Такое статистическое обучение должно быть полностью основано на принципе иерархии, что отражает структуру языка и принцип работы человеческого мозга.

Именно так работают Сири и Dragon Go: для определения наиболее общих и надежных явлений используются заранее сформулированные правила, а усвоение более редких элементов языка находится в руках конкретных пользователей. Когда создатели Сус обнаружили, что достигли потолка эффективности при обучении системы на заранее сформулированных правилах, они также переключились на этот подход. Правила, определенные лингвистами, выполняют две важнейшие функции. Во-первых, они обеспечивают приемлемую начальную точность, так что систему можно допускать к широкому использованию, где она будет улучшаться автоматически. Во-вторых, они служат надежной основой для низших уровней понятий, от которых начинается автоматический подъем на более высокие иерархические уровни.

Как отмечалось выше, Ватсон является удивительным примером реализации комбинированного подхода, в котором сочетается настройка системы по предварительно сформулированным правилам и ее иерархическое статистическое обучение. Для создания системы, способной играть в «Джеопарди!» на разговорном языке, компания IBM объединила несколько лучших программ. С 14 по 16 февраля 2011 г. Ватсон соревновался

с двумя ведущими игроками: Брэдом Раттером, выигравшим в эту викторину больше денег, чем кто-либо другой, и Кеном Дженнингсом, который удерживал звание чемпиона викторины рекордное время — 75 дней.



*Точность систем распознавания разговорной речи как функция объема тренировочных данных. Наилучшие результаты получаются при сочетании заранее сформулированных правил для освоения «основного» языка и самостоятельной тренировки для освоения «деталей».*

В моей первой книге «Эра разумных машин», написанной в середине 1980-х гг., я предсказывал, что компьютер станет чемпионом мира по шахматам примерно к 1998 г. Еще я предсказал, что если это произойдет, то мы либо станем хуже думать о человеческом разуме, либо лучше думать о разуме машин, либо потеряем интерес к шахматам. И если принять во внимание историю, скорее всего, должно было произойти последнее. Так и вышло. В 1997 г., когда суперкомпьютер IBM Deep Blue обыграл чемпиона мира по шахматам Гарри Каспарова, мы немедленно стали утверждать, что именно так и должно было случиться, поскольку компьютеры — логические машины, а шахматы, в конце концов, игра на логику. Победу компьютера не сочли ни важным, ни удивительным событием. Многие критики принялись утверждать, что компьютер никогда бы не смог

одолеть человеческую речь — со всеми метафорами, сравнениями, каламбурами, двусмысленностями и юмором.

Вот почему Ватсон так много значит: «Джеопарди!» — именно такая сложная и напряженная игра на знание языка. Типичные вопросы викторины содержат непростые обороты человеческой речи. Однако для большинства зрителей, возможно, не очевидно, что Ватсон не только понимает вопросы, заданные в неожиданной и запутанной форме, но большая часть его знаний не была сформирована людьми. Он обзавелся этими знаниями самостоятельно, прочитав 200 млн страниц документов на человеческом языке, включая «Википедию» и другие энциклопедии, что составляет 4 трлн байт информации. Как вы понимаете, «Википедия» написана не на ЛИСП или Сул, а «нормальными» предложениями со всеми присущими языку двусмысленностями и путаницами. При ответе на вопрос викторины Ватсон должен проверить все 4 млрд знаков реферативного материала (конечно, «Джеопарди!» — не вопросы, а загадки, но это техническая сторона дела — по форме это настоящие вопросы). Если Ватсон способен понять вопрос и ответить на него на основании 200 млн страниц текста — и всего за три секунды! — ничто не может помешать подобным машинам прочесть миллиарды имеющихся в Интернете документов. Именно это сейчас и происходит.

Когда в период с 1970-х по 1990-е гг. мы занимались разработкой систем для распознавания знаков и речи и первых систем, понимающих разговорную речь, мы включали в свои программы «эксперта-менеджера». Мы создавали разные системы для решения одной и той же задачи, но в каждом случае использовали несколько иной подход. Некоторые из различий были незначительными, например вариации параметров, контролирующих математику алгоритма обучения. Но некоторые были фундаментальными, например использование предварительно сформулированных правил вместо иерархических статистически обучаемых систем. Эксперт-менеджер представлял собой компьютерную программу, призванную изучить сильные и слабые стороны

различных систем путем анализа их эффективности в реальных ситуациях. Оценка производилась по принципу ортогональности, то есть одна система считалась скорее сильной, другая — скорее слабой. Выяснилось, что общая эффективность комбинированных систем с обученным экспертом-менеджером была намного выше, чем у отдельных систем.

Ватсон действует по такому же принципу. Используя архитектуру UIMA (Unstructured Information Management Architecture), Ватсон распоряжается буквально сотнями различных систем — многие отдельные языковые компоненты Ватсона аналогичны тем, что применяются в широко используемых системах распознавания разговорного языка, — и все они либо пытаются напрямую дать ответ на вопрос викторины, либо как минимум прояснить вопрос. UIMA выступает в роли эксперта-менеджера, разумно сочетающего результаты отдельных систем. UIMA превзошла более ранние системы (включая разработанные нами еще до создания компании Nuance) в том, что составляющие ее системы могут участвовать в получении результата, даже если не пришли к финальному ответу. Достаточно, чтобы подсистема помогла найти путь к ответу. Кроме того, UIMA может рассчитать степень собственного доверия к полученному ответу. Человеческий мозг тоже это делает: обычно вы совершенно уверены в ответе, если у вас спросят, как зовут вашу мать, но гораздо менее уверены, если речь идет о человеке, которого вы видели один раз год назад.

Таким образом, вместо того чтобы разработать единственный путь разрешения речевых проблем, возникающих при ответах на вопросы «Джеопарди!», специалисты из IBM объединили все имевшиеся на тот момент модули распознавания речи. В некоторых используются иерархические скрытые модели Маркова, в других — заранее сформулированные языковые правила, кодирующие основной массив языковых данных. UIMA оценивает эффективность каждой системы в каждом случае и оптимальным образом объединяет результаты. Важная роль UIMA, которую подчеркивают создавшие ее специалисты из IBM, вызывает определенное

недопонимание, возникающее при обсуждении способностей Ватсона. Некоторые зрители полагают, что Ватсон на самом деле не понимает речь, поскольку трудно сказать, в каком модуле происходит это «понимание». Но хотя UIMA тоже обучается на своем собственном опыте, «понимание» Ватсоном языка происходит не только благодаря ей, но распределено по всем многочисленным модулям, включая самоорганизующиеся языковые модули, действующие по методам, аналогичным ИСММ.

Отдельная часть системы использует произведенную UIMA оценку достоверности ответа Ватсона для заключения пари. Хотя Ватсон специфическим образом оптимизирован именно для этой игры, его базовые знания и технология поиска новых знаний легко могут быть адаптированы для решения более широкого круга задач. Возможно, некоторые думают, что машине сложнее будет управляться с менее общими и более узкопрофессиональными данными (например, в сфере медицины), чем с «общими» знаниями, требующимися для игры в «Джеопарди!» На самом деле ситуация как раз противоположная. Профессиональные знания упорядочены в гораздо большей степени, лучше структурированы и менее двусмысленны, поэтому с помощью той же самой технологии вполне можно достичь точного понимания языка в подобных специфических областях. Я уже упоминал, что в настоящее время IBM и Nuance совместно трудятся над адаптацией Ватсона для работы в сфере медицины.

Диалог с Ватсоном во время викторины короткий: ему задают вопрос, он дает ответ. Он не вовлекается в разговор, в котором были бы упомянуты все предыдущие высказывания всех игроков. (А вот Сири в какой-то степени это делает: если вы попросите ее отправить сообщение вашей жене, она в первый раз попросит ее назвать, но после этого запомнит.) Проследить за всей информацией, упомянутой в ходе разговора (что, безусловно, было бы необходимо для прохождения теста Тьюринга), — серьезное дополнительное требование, но выполнить его, по сути, не сложнее, чем то, что Ватсон уже умеет делать.

В конце концов, Ватсон прочел миллионы страниц текста, включая многочисленные истории, так что он способен наблюдать за запутанными последовательностями событий. Таким образом, он сможет следить за собственными ответами и учитывать их в последующих турах викторины.

Еще одно ограничение викторины заключается в том, что ответы обычно очень короткие. Например, игроков не просят перечислить пять основных сюжетных линий «Повести о двух городах»\*. Для обсуждения таких вопросов нужна несколько иная версия Ватсона. Самостоятельно выявить главные темы книги, не копируя чужие мысли (даже без слов), это совсем другая задача, гораздо более сложная, чем те, которые сейчас умеет решать Ватсон; я бы назвал эту задачу тестом на уровне теста Тьюринга. (Говоря это, я хочу подчеркнуть, что большинство людей при решении подобной задачи воспользуются не собственными мыслями, а скопируют уже готовые идеи.) В любом случае, 2029 г. еще не наступил, так что я пока не ожидаю появления машинного разума, способного пройти тест Тьюринга. Кроме того, я хочу отметить, что оценка ответов на такие вопросы, как выявление ключевых идей литературного произведения, сама по себе не является очевидной задачей. Если спрашивают, кто подписал Декларацию независимости, верность ответа оценить легко. С более сложными понятийными вопросами дело обстоит намного сложнее.

Нужно сказать, что, хотя языковые навыки Ватсона в настоящее время слабее, чем у образованного человека, он смог одолеть двух лучших игроков в «Джеопарди!» Ему помогло сочетание лингвистических возможностей и многочисленных познаний человечества с очень точной памятью машины. Вот почему мы уже передали компьютерам значительную часть наших персональных, социальных и исторических воспоминаний.

---

\* «Повесть о двух городах» (A Tale of Two Cities) — исторический роман Ч. Дикенса о временах Французской революции.

Я не готов передвинуть мой прогноз о прохождении компьютером теста Тьюринга с 2029 г. на более раннюю дату, однако прогресс, который достигнут в создании подобных Ватсону машин, вселяет уверенность, что вскоре появятся компьютеры с ИИ уровня теста Тьюринга. Если бы кто-то специально занялся созданием версии Ватсона, оптимизированной для прохождения теста Тьюринга, возможно, такая машина появилась бы совсем скоро.

Американский философ Джон Серль (род. в 1932 г.) недавно заявил, что Ватсон не способен думать. Опираясь на свой мысленный эксперимент «Китайская комната» (я расскажу о нем в одиннадцатой главе), он утверждает, что Ватсон лишь манипулирует символами, но не понимает их смысла. На самом деле, Серль неточно описывает Ватсона, поскольку понимание Ватсоном языка основано не на манипуляции символами, а на иерархическом статистическом процессе обучения. Характеристика Серля верна лишь в том случае, если рассматривать каждую стадию самоорганизующегося процесса как «манипуляцию символами». Но если это так, то и человеческий мозг нельзя называть мыслящим.

Забавно, когда Ватсона критикуют за то, что он *всего лишь* выполняет статистический анализ языка, но не владеет «истинным» пониманием человеческой речи. Иерархический статистический анализ — это именно то, что делает человеческий мозг, разбирая несколько гипотез на основе статистических выводов (на каждом иерархическом уровне новой коры). И Ватсон, и человек обучаются и отвечают на вопросы на основании одного и того же иерархического подхода. Во многих отношениях знания Ватсона намного обширнее человеческих; ни один человек не может утверждать, что прочел всю «Википедию», которая составляет лишь часть базы данных Ватсона. Напротив, человек пока еще может справляться с более сложными понятийными задачами, чем Ватсон, но этот пробел вскоре закроется.

Важной системой, демонстрирующей мощь применения компьютерных возможностей для обработки организованных баз знаний, является программа Wolfram Alpha. Эта отвечающая программа (в отличие от поисковой программы) разработана британским физиком и математиком Стивеном Вольфрамом (род. в 1959 г.) и его коллегами из компании Wolfram Research. Если вы спросите Wolfram Alpha (на сайте WolframAlpha.com) сколько существует простых чисел меньше миллиона, она ответит: «78 498». Этот ответ программа не отыщет, она его вычислит и после ответа выдаст вам уравнения, которыми пользовалась. Если вы попытаетесь получить ответ с помощью обычных поисковых программ, они направят вас к сайтам, где вы найдете нужный алгоритм. Далее вам потребуется загрузить эти формулы в программу типа «Математика» (также созданную Вольфрамом), но на это уйдет гораздо больше времени (и усилий), чем просто обращение к Wolfram Alpha.

Программа Wolfram Alpha содержит 15 млн строк кода «Математики». Ее функция заключается в расчете ответа на основании примерно 10 трлн байт данных, тщательно подобранных сотрудниками Wolfram Research. Программе можно задать множество фактических вопросов типа «В какой стране самый высокий валовой внутренний продукт на душу населения?» (ответ: в Монако, 212 тыс. долл. США) или «Сколько лет Стивену Вольфраму?» (ответ (на момент написания книги): 52 года, 9 месяцев и 2 дня). Как я уже говорил, Wolfram Alpha используется в технологии Сири. Если вы задаете Сири фактический вопрос, она переадресует его Alpha. Кроме того, Alpha осуществляет некоторые поисковые функции для поисковой программы Bing.

Недавно Вольфрам сообщил, что Alpha дает правильный ответ в 90% случаев<sup>16</sup>. Он также указал, что число отказов сокращается вдвое каждые 18 месяцев. Это очень мощная система, которая использует созданные вручную методы и собранные вручную данные. Именно для этого мы когда-то впервые сделали

компьютеры. По мере того как мы открываем и компилируем научные и математические методы, мы все шире привлекаем компьютеры, которые используют эти методы гораздо лучше, чем невооруженный человеческий разум. Большинство наших научных методов кодируется в программе Alpha параллельно с постоянно обновляемыми данными во многих областях знания — от физики до экономики. Во время нашей частной беседы с Вольфрамом он сообщил, что применение самоорганизующихся методов, подобных тому, что использует Ватсон, позволяет достигать точности 80%. А Alpha способна на 90% точности. Конечно же, эти значения в определенной степени субъективны, поскольку пользователи (включая меня самого) выбирают вопросы, на которые их система отвечает хорошо, и то же касается самоорганизующихся методов. Кажется, 80% — достаточно верная оценка точности ответов Ватсона при игре в «Джеопарди!», но такой точности было довольно, чтобы он смог обыграть двух лучших игроков.

По моему мнению, такие самоорганизующиеся методы, как я описал для теории мысленного распознавания образов (ТМРО), необходимы для понимания сложной и часто неоднозначной иерархии, встречающейся в реальном мире, в том числе в человеческом языке. Идеальным сочетанием для надежной разумной системы была бы комбинация иерархического подхода, основанного на ТМРО (который, как я считаю, лежит в основе функционирования человеческого мозга), и точной кодировки научных данных. А это и есть сочетание человека и компьютера. В ближайшие годы мы усилим оба полюса. Хотя наша биологическая новая кора достаточно пластична, ее базовая архитектура имеет физические ограничения. Увеличение размеров новой коры в лобной части нашего мозга стало важной эволюционной инновацией, но теперь мы не в состоянии расширить наши лобные доли ни в тысячу раз, ни даже на 10%. Я хочу сказать, что мы не можем сделать это биологическим путем, но мы совершим это технологически.

## Как создать разум

В нашем головном мозге миллиарды нейронов, но что такое нейроны? Просто клетки. Пока между нейронами не образуются связи, мозг не обладает никакими знаниями. Все, что мы знаем, все, что мы есть, зависит от того, как связываются наши нейроны.

*Тим Бернерс-Ли\**

Для создания разума будем основываться на тех рассуждениях, которые я привел выше. Начнем с построения распознающего модуля, отвечающего всем необходимым требованиям. Далее создадим множество копий этого модуля — столько, сколько позволяют возможности компьютеров. Каждый модуль рассчитывает вероятность распознавания соответствующего образа. При этом он учитывает наблюдаемую величину каждого входного сигнала и соотносит ее с усвоенной им величиной и ее возможной вариабельностью. Если рассчитанная величина сигнала превосходит пороговое значение, распознающий модуль активирует соответствующий аксон. Пороговое значение и параметры, учитываемые при вычислении вероятности присутствия образа, относятся к числу параметров, которые оптимизируются с помощью генетического алгоритма. Поскольку для распознавания образа не требуется активизации всех входных сигналов, это обеспечивает самоассоциативное распознавание (то есть распознавание всего образа на основании отдельных частей). Также следует учесть возможность ингибирующих сигналов (означающих сниженную вероятность присутствия образа).

В результате распознавания образа распознающий модуль посылает сигнал дальше по аксону. Этот аксон связан с одним

---

\* Сэр Тимоти Джон Бернерс-Ли (род. в 1955) — английский ученый, один из изобретателей Всемирной паутины и действующий глава Консорциума Всемирной паутины; автор множества других разработок в области информационных технологий.

или несколькими другими распознающими модулями на более высоком понятийном уровне. Все распознающие модули следующего уровня, к которым приходит сигнал, воспринимают этот образ как входной сигнал. Когда большая часть образа распознана, каждый распознающий модуль отправляет сигналы и на более низкий концептуальный уровень; эти сигналы дают понять, что образ «ожидается». Каждый модуль имеет один или несколько каналов для доставки сигналов ожидания. Когда такой сигнал получен, порог распознавания этого образа в данном распознающем модуле понижается (распознавание облегчается).

Распознающие модули отвечают за связывание с другими распознающими модулями выше и ниже по иерархии. Заметим, что все эти контакты в компьютерном варианте действуют через виртуальные связи (которые, как в виртуальных сетях, представляют собой просто указатели), а не через реальные контакты. Данная система гораздо более гибкая, чем в биологическом мозге. В человеческом мозге новые образы должны быть «приписаны» к реальным распознающим модулям, а между аксонами и дендритами должны образовываться реальные связи. Обычно для этого используется существующая физическая связь, имеющая отношение к данному образу, и за счет дополнительного роста аксонов и дендритов осуществляется новое взаимодействие.

Еще один механизм в биологической новой коре млекопитающих заключается в постепенной ликвидации неиспользуемых нервных связей. Для того чтобы перенастроить распознающие модули новой коры на восприятие каких-то новых образов, необходима физическая реконфигурация связей. И вновь в компьютерном варианте эта задача решается гораздо проще. Мы просто присваиваем новым распознающим модулям новую информацию и программируем новые связи. Если цифровой коре нужно перенастроить ресурсы памяти на новый набор образов, она освобождает распознающие модули от старых образов и задает новые настройки. Такой своеобразный «сбор мусора»

и перераспределение памяти являются стандартным свойством многих компьютерных систем. В цифровом мозге нам также следует создать резервную копию старых воспоминаний, прежде чем удалить их из активной коры, что в нашем биологическом мозге мы сделать не в состоянии.

Существует несколько математических методов, которые можно использовать для создания самоорганизующихся иерархических распознающих модулей. Лично я по нескольким причинам предпочитаю метод скрытых иерархических моделей Маркова. Я уже несколько десятилетий пользуюсь этим методом, начиная с самых первых систем распознавания речи и понимания разговорного языка, созданных в 1980-х гг. Да и если говорить в общем, ученые, занятые распознаванием образов, имеют больше опыта в применении данного метода, чем каких-либо других. Этот метод или аналогичные ему математические методы также широко применяются для понимания разговорной речи.

Следует сказать, что не все системы, действующие по принципу скрытых моделей Маркова, являются полностью иерархическими. В некоторых предусмотрено лишь несколько уровней иерархии, например при переходе от акустического состояния к фонеме и слову. Чтобы создать разум, мы позволим системе создавать столько новых уровней иерархии, сколько понадобится. Кроме того, не все системы на основе скрытых моделей Маркова являются самоорганизующимися. В некоторых запрограммированы фиксированные контакты, однако эти системы умеют эффективно удалять многие исходные связи, присваивая им нулевой вес. В наших системах, созданных в 1980-х и 1990-х гг., происходило автоматическое удаление контактов, вес которых был ниже определенного уровня, а также формирование новых контактов для лучшего соответствия тренировочным данным и обучения. Для оптимальной организации связей с новыми распознающими модулями мы можем использовать линейное программирование.

Наша цифровая новая кора будет характеризоваться определенной степенью избыточности, особенно это относится к часто встречающимся образам. Это обеспечивает надежное узнавание распространенных образов, а также является ключевым элементом в достижении инвариантного распознавания различных форм образа. Однако нам придется установить правила для ограничения избыточности, поскольку не следует отводить слишком много места для сохранения самых распространенных образов низкого порядка.

Правила, ограничивающие избыточность, порог распознавания и связь порога распознавания с ожидаемостью образа, — примеры общих параметров, которые влияют на эффективность таких самоорганизующихся систем. Сначала мы выберем их значения интуитивно, а затем оптимизируем с помощью генетического алгоритма.

Очень важный этап — обучение мозга, как биологического, так и компьютерного. Как я уже писал, иерархическая система распознавания образов (и цифровая, и биологическая) за один момент осваивает не больше двух иерархических уровней (а скорее один). Чтобы усовершенствовать систему, я начну с предварительно обученных иерархических сетей, которые уже научились распознавать человеческую речь, печатные буквы и структуры разговорного языка. Такая система сможет читать документы, написанные разговорным языком, но за один раз по-прежнему сможет осваивать примерно один понятийный уровень. Ранее усвоенные понятия создадут достаточно прочную основу для продвижения на следующие уровни. Система может вновь и вновь обращаться к уже прочитанным документам, достраивая новые понятийные уровни при каждом следующем прочтении. Так и люди глубже понимают текст, когда читают его повторно. В нашем доступе имеются миллиарды страниц информативного материала, такого как «Википедия».

Система также будет содержать модуль критического мышления, который станет осуществлять постоянное сканирование

всех существующих образов и оценивать их совместимость с другими образами (идеями) цифровой новой коры. У биологической новой коры такой функции нет, вот почему люди иногда совмещают в своей голове совершенно несовместимые идеи. В случае обнаружения противоречивых данных модуль попытается найти решение, используя как кортикальные структуры самой системы, так и весь массив имеющихся литературных данных. Принятие решения может просто сводиться к установлению некорректности одной из идей (если она противоречит большинству имеющихся данных). В более конструктивном плане может быть выявлена идея более высокого понятийного уровня, которая позволяет разрешить кажущееся несоответствие путем объяснения обеих исходных идей. Система добавит это решение в виде нового образа и свяжет его с теми идеями, которые запустили поиск. Модуль критического мышления будет работать постоянно.

Кроме того, я предполагаю установить модуль, который выявляет нерешенные вопросы в различных дисциплинах. Этот модуль будет искать ответы на подобные вопросы во всех возможных источниках информации также на постоянной основе. Как я объяснял в начале книги, хранящаяся в новой коре информация состоит из глубоко угнездившихся друг в друге образов и, следовательно, является полностью метафорической. Мы можем использовать конкретный образ для поиска решения практически в любой несвязанной области.

В качестве примера вспомните метафору, которую я приводил в четвертой главе, связывая случайные движения молекул газа и случайность эволюционных изменений. Молекулы газа в банке движутся непредсказуемым образом, и все же через определенное время все они из этой банки выйдут. Я отмечал, что это наталкивает на важный вопрос об эволюции разума. Подобно молекулам газа, эволюционные изменения тоже происходят во всех направлениях, не имея какой-либо цели. Однако мы наблюдаем изменения в сторону усложнения и усовершенствования

разума и в итоге — появление новой коры, способной осуществлять иерархическое мышление. Таким образом, можно наблюдать, как бесцельный и ненаправленный процесс (в термодинамическом плане) может привести к полезному результату (в плане биологической эволюции).

Я также уже писал о том, что догадка Чарлза Лайеля о роли небольших изменений горных пород под действием воды в формировании гигантских каньонов подтолкнула Чарлза Дарвина к аналогичным наблюдениям о роли малых изменений в формировании видов живых организмов. Поиск таких метафор будет еще одним непрерывным процессом в работе цифрового мозга.

Нам следует найти способ обрабатывать одновременно несколько списков, что эквивалентно структурному мышлению. Списком может быть определение ограничений, которым должно удовлетворять решение. Каждый шаг может запускать рекурсивный поиск в существующей иерархии идей или в доступной литературе. Кажется, человеческий мозг способен одновременно обрабатывать четыре таких списка (без помощи компьютера), но у цифровой коры таких ограничений быть не должно.

Мы также хотим усилить наш искусственный мозг такой способностью, которой блестяще владеют компьютеры: речь идет о способности управляться с огромными массивами данных, а также быстро и эффективно применять известные алгоритмы. Например, Wolfram Alpha сочетает в себе возможности множества научных методов и применяет их к тщательно подобранным данным. Такие системы продолжают совершенствоваться, учитывая заявление Вольфрама об экспоненциальном сокращении случаев отказов системы.

Наконец, нашему новому мозгу нужна цель. Цель выражается в виде набора задач. Когда речь идет о биологическом мозге, задачи определяются центрами удовольствия и страха, унаследованными от старого мозга. Эти примитивные двигатели созданы биологической эволюцией для того, чтобы вид мог выжить, но новая кора научила нас сублимировать их. Задача

Ватсона состоит в ответах на вопросы «Джеопарди!» Еще одна просто формулируемая задача может заключаться в прохождении теста Тьюринга. Чтобы это сделать, цифровой мозг должен уметь по-человечески изложить свою вымышленную историю, чтобы претендовать на роль реального (биологического) человека. Он, кроме того, должен уметь скрывать свои способности, поскольку любую систему, которая обладает такими знаниями, как, скажем, Ватсон, очень легко отличить от человека.

Но интереснее то, что мы можем поставить перед нашим новым мозгом более амбициозную задачу, такую как создание лучшего мира. Конечно, на этом пути возникает множество вопросов: лучшего для кого? в каком плане? для людей как биологического вида? для всех сознательных существ? если так, то кого можно считать сознательным существом?

Если небиологический мозг получит такую же способность изменять мир, как биологический мозг (на самом деле значительно бóльшую способность, чем невооруженный биологический мозг), нам необходимо предусмотреть его морально-нравственное воспитание. Начать можно с золотого правила нравственности, установленного нашими религиозными традициями.

# Глава восьмая

## МОЗГ КАК КОМПЬЮТЕР

Наш мозг, по форме напоминающий французский деревенский хлеб, представляет собой переполненную химическую лабораторию, в которой никогда не прекращаются диалоги между нейронами. Представьте себе этот яркий клубок жизни, этот сгусток клеток мышиноного цвета, эту фабрику идей, этого маленького деспота в костяной чаше, это скопление бегущих во всех направлениях нейронов, это маленькое все, этот переменчивый центр веселья, этот ящик, в который вся суть личности втиснута, как комок одежды в спортивную сумку.

*Диана Аккерман\**

Мозг существует по той причине, что распределение необходимых для выживания ресурсов и угрожающих жизни опасностей изменяется во времени и в пространстве.

*Джон М. Оллмен\*\**

Современная география мозга сохранила чудесную архаичность, как средневековая карта, в которой рядом с известными мирами соседствуют неизведанные земли, населенные монстрами.

*Дэвид Бейнбридж\*\*\**

В математике вы не понимаете вещей. Вы просто их используете.

*Джон фон Нейман*

---

\* Диана Аккерман (род. в 1948) — американский поэт, писатель и натуралист.

\*\* Джон Морган Оллмен — американский нейрофизиолог, специалист в области эволюционной нейробиологии.

\*\*\* Дэвид Бейнбридж (род. в 1968) — английский писатель-публицист, специалист в области репродуктивной биологии и анатомии животных.

**У**же с момента появления первых компьютеров в середине XX века начались дебаты не только относительно предела их возможностей, но и относительно того, можно ли рассматривать человеческий мозг в качестве формы компьютера. Общепринятое мнение по последнему вопросу изменилось кардинально от идеи о том, что эти два обрабатывающих информацию «устройства» принципиально не различаются, к идее об их принципиальном различии. Итак, является ли наш мозг компьютером?

Когда в 1940-х гг. компьютеры впервые привлекли к себе внимание общественности, их тут же стали воспринимать в качестве мыслящих машин. Первый компьютер ENIAC\*, о создании которого было объявлено в 1946 г., пресса назвала «гигантским мозгом». По мере роста доступности компьютеров в последующие десятилетия их постоянно сравнивали с мозгом, который может делать то, чего не может обыкновенный биологический мозг.

Компьютерные программы быстро научили машины действовать в соответствии с этими рекламными заявлениями. «Общая программа для решения задач», созданная в 1959 г. Гербертом Саймоном, Дж. Клиффом Шоу и Алленом Ньюэллом в Университете Карнеги—Меллон, смогла доказать теорему, с которой не справились Бертран Рассел и Альфред Норт Уайтхед в своей знаменитой работе 1913 г. *Principia Mathematica*\*\*.

В последующие десятилетия стало ясно, что компьютеры значительно превосходят невооруженный человеческий мозг по способности решать математические задачи, диагностировать заболевания и играть в шахматы, однако робота трудно научить завязывать шнурки или понимать разговорную человеческую речь, которую запросто понимает пятилетний ребенок. С таки-

---

\* ENIAC (Electronic Numerical Integrator and Computer; электронный числовой интегратор и вычислитель) — первый электронный цифровой компьютер общего назначения, созданный в США в годы Второй мировой войны.

\*\* «Принципы математики» (или «Начала математики») — трехтомный труд по логике и философии математики, выпущенный Расселом и Уайтхедом в 1910, 1912 и 1913 гг.

ми задачами компьютеры начинают справляться только в наши дни. Забавно, что эволюция компьютерного разума происходит в обратном направлении в сравнении с созреванием человеческого разума.



*Реклама 1957 года, отражающая популярное видение компьютера как гигантского мозга.*

Вопрос о том, являются ли человеческий мозг и компьютер эквивалентными на каком-то уровне, до сих пор остается открытым. Во введении я писал о том, что в Интернете можно найти миллионы ссылок, относящихся к сложности человеческого мозга. Аналогичным образом, если в Google запустить поиск на тему «мозг — не компьютер», вы тоже обнаружите миллионы ссылок. На мой взгляд, подобные формулировки аналогичны формулировке «яблочное пюре — не яблоко». Технически утверждение верно, но из яблока можно сделать яблочное пюре. Возможно, правильнее сказать: «Компьютеры — не текстовые процессоры». Правда, что компьютеры и текстовые процессоры

существуют на разном понятийном уровне, но компьютер может стать текстовым процессором, если обладает соответствующей программой. Аналогичным образом, компьютер может стать мозгом, если снабжен программами мозга. Именно это и пытаются сделать такие ученые, как я.

Таким образом, вопрос заключается в следующем: сможем ли мы найти алгоритм, который превратит компьютер в эквивалент человеческого мозга? В конце концов, ввиду своей универсальности (ограниченной только мощностью) компьютер может следовать любому созданному нами алгоритму. Напротив, человеческий мозг действует лишь по специфическим алгоритмам. Эти методы разумны, так как обеспечивают значительную пластичность и возможность реструктуризации внутренних связей на основе пережитого опыта, но все эти функции можно имитировать с помощью компьютерных программ.

Идея об универсальности компьютерных вычислительных методов (идея о том, что компьютер общего назначения может использовать любой алгоритм) возникла одновременно с появлением первых настоящих компьютеров. Я назову четыре основных принципа, лежащих в основе универсальности и осуществимости компьютерных вычислений и их применимости к описанию нашего мышления. Эти принципы стоит упомянуть еще и по той причине, что человеческий мозг тоже их использует. Первый принцип заключается в способности передавать, запоминать и обрабатывать информацию. В 1940-х гг. под словом «компьютер» люди понимали аналоговую машину, в которой числовые данные представлялись в виде аналоговых физических переменных, таких как напряжение, а специализированные элементы могли выполнять арифметические действия, такие как сложение или умножение. Однако серьезным ограничением аналоговых компьютеров является недостаточная точность. Данные могут быть представлены с точностью около одного процента, а по мере увеличения числа арифметических операций точность данных еще больше снижается. При проведении

большого числа вычислений результат становится столь неточным, что вообще теряет смысл.

Всякий, кто помнит попытки записывать музыку с помощью аналоговых устройств, знает об этом эффекте. Первая копия была значительно хуже оригинала по уровню шума (вспомним, что «шум» представляет собой случайные несовпадения). Копия первой копии была еще хуже, а десятая копия представляла собой один только шум. Считалось, что зарождавшийся мир цифровых компьютеров будет поражен той же болезнью. Это беспокойство понятно, если речь идет о передаче цифровой информации по каналу. Идеальных каналов не бывает, каждый характеризуется определенным уровнем ошибки. Допустим, у нас есть канал, который правильно проводит каждый байт информации с вероятностью 0,9. Если я передаю сообщение размером один бит, вероятность точной передачи по этому каналу составляет 0,9. Если объем информации составляет два бита, точность передачи равняется  $0,9^2 = 0,81$ . А если мне нужно передать один байт (8 бит)? Вероятность безошибочной передачи этой информации меньше 50% (0,43, если быть точным). Вероятность безошибочной передачи пяти байт информации составляет примерно 1%.

Очевидный путь решения проблемы состоит в повышении качества каналов. Допустим, при прохождении по каналу миллиона бит информации возникает лишь одна ошибка. Но если я пересылаю файл, содержащий полмиллиона байт (это размер небольшой программы или базы данных), вероятность правильной передачи будет менее 2%, даже при таком высоком качестве канала. Учтывая, что ошибка всего в одном байте информации может полностью обесценить компьютерную программу или любые другие данные, ситуация является неприемлемой. Таким образом, вне зависимости от качества канала, поскольку вероятность ошибки передачи данных быстро растет с увеличением объема передаваемой информации, это ограничение кажется непреодолимым.

Создатели аналоговых компьютеров подошли к решению проблемы путем постепенного снижения требований (то есть

позволяли решать только такие задачи, в которых допустима небольшая погрешность), и, если ограничиваться определенными вычислениями, эти компьютеры могли быть полезными. С другой стороны, цифровые компьютеры нуждаются в постоянном обмене информацией — не только между компьютерами, но и внутри самого компьютера. Это обмен информацией между модулем памяти и центральным процессором. Внутри центрального процессора осуществляется связь между регистрами и вычислительными регистрами. И даже в вычислительных регистрах происходит обмен между битовыми регистрами. Таким образом, обмен информацией идет буквально на всех уровнях. Если учесть, что количество ошибок быстро возрастает с усилением информационного обмена и что ошибка в одном байте информации может нарушить весь процесс, цифровые вычислительные методы, как казалось, были обречены на провал.

Таково оставалось общественное мнение до тех пор, пока американский математик Клод Шеннон (1916–2001) не показал, каким образом можно достичь достаточно точной передачи информации с помощью самых ненадежных каналов передачи. В своей всемирно известной статье «Математическая теория связи» (*A Mathematical Theory of Communication*), опубликованной в журнале *Bell System Technical Journal* в июле и октябре 1948 г., и, в частности, в теореме о пропускной способности канала Шеннон постулировал, что с помощью канала с любым уровнем шума (ниже 50%, поскольку такой канал передает только шум) можно передавать сообщение с любой точностью. Другими словами, число ошибок может составлять один бит из  $n$  битов даже при сколь угодно большом значении  $n$ . Например, в предельном случае, если у вас есть канал, безошибочно передающий информацию лишь 51% времени (то есть бит правильной информации передается лишь немного чаще, чем бит неправильной информации), вы тем не менее можете передавать сообщения таким образом, что ошибка будет лишь в одном бите из миллиона или в одном бите из триллиона или из триллиона триллионов бит.

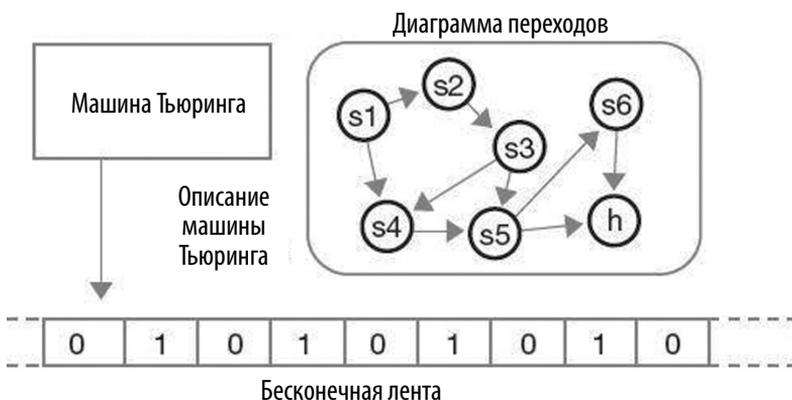
Как же это возможно? Ответ заключается в избыточности информации. Сейчас это кажется очевидным, но тогда дело обстояло иначе. Если, к примеру, я передаю каждый бит информации три раза, надежность результата станет значительно выше. Если этого недостаточно, нужно просто повышать степень избыточности до тех пор, пока не будет достигнут приемлемый результат. Повтор информации — простейший путь получения желаемой точности передачи информации по каналу с высоким уровнем шума, но не самый эффективный. В статье Шеннона, с которой началось развитие теории коммуникации, были предложены оптимальные методы определения ошибок и коррекционные коды, позволяющие достичь *любой* точности передачи информации с помощью *любого* направленного канала связи.

Читатели старшего поколения наверняка помнят телефонные модемы, проводившие информацию по шумным аналоговым телефонным линиям. Такая связь сопровождалась шипением, свистом и многими другими помехами, но при этом позволяла передавать цифровые сообщения с очень высокой точностью — в соответствии с теоремой Шеннона о зашумленном канале. Та же проблема и такое же решение существуют для цифровой памяти. Почему, по вашему мнению, программные диски, CD и DVD продолжают работать даже при появлении царапин или после падения на пол? Еще раз скажем спасибо Шеннону.

Вычислительный процесс состоит из трех элементов: коммуникации (которая, как я подчеркнул, осуществляется как внутри компьютера, так и между компьютерами), памяти и логических вентилях (производящих арифметические и логические действия). Точность логических вентилях тоже можно сделать сколь угодно высокой с помощью определения ошибок и коррекционных кодов. Благодаря теореме и теории Шеннона мы можем без помех управлять большими массивами сложных цифровых данных и алгоритмов. Важно отметить, что головной мозг тоже использует принцип Шеннона, хотя эволюция человеческого мозга произошла задолго до появления мозга самого Шеннона! Как

мы видели, большинство образов или идей (идея — тоже образ) хранится в мозге в избыточном количестве. Основной смысл избыточности образов заключается в преодолении ненадежности цепей нейронов.

Второй важный принцип в основе теории информации я уже упоминал — это универсальность вычислительных методов. В 1936 г. Алан Тьюринг описал «машину Тьюринга», которая представляет собой не реальный механизм, а еще один мысленный эксперимент. Этот абстрактный компьютер состоит из бесконечно длинной ленты памяти, в ячейках которой располагаются цифры 1 или 0. Эта информация служит для машины входным сигналом, и машина за один раз прочитывает информацию в одной ячейке. Кроме того, машина снабжена таблицей правил (в частности, подчиняется принципу хранимой программы), представленных в виде пронумерованных утверждений. Каждое правило определяет одно действие, если в текущей ячейке расположена цифра 0, и другое, если в текущей ячейке цифра 1. К возможным действиям относится написание на ленте 0 или 1, перемещение ленты на одну ячейку влево или вправо или остановка. Каждое состояние определяет следующее состояние машины.



*Блок-диаграмма машины Тьюринга, состоящей из считывающей и записывающей головки и внутренней программы, осуществляющей переходы между состояниями.*

Входной сигнал задается информацией на ленте. Программа работает, а когда алгоритм выполнен, машина останавливается и выходные данные оказываются отпечатанными на ленте. Заметим, что, хотя теоретически лента имеет бесконечную длину, любая реальная программа, не попадающая в бесконечный цикл, использует лишь ограниченную часть ленты, так что, если мы возьмем ленту конечного размера, машина по-прежнему сможет решать некий набор задач.

Схема машины Тьюринга кажется простой — так сделал ее создатель. Он хотел, чтобы его машина была максимально простой (но не проще, если перефразировать Эйнштейна\*). В результате работ Тьюринга и его бывшего учителя Алонзо Черча был сформулирован так называемый тезис Черча—Тьюринга: если некая функция не может быть вычислена машиной Тьюринга, она не может быть вычислена *никакой* другой машиной, подчиняющейся законам природы. Хотя машина Тьюринга выполняет лишь несколько команд и обрабатывает за один раз только один бит информации, она может решить любую задачу, которую способен решить любой другой компьютер. Иными словами, любая машина, эквивалентная машине Тьюринга, может вычислить любой алгоритм (решить любую задачу, которую мы в состоянии сформулировать).

В «строгой» формулировке тезиса Черча—Тьюринга постулируется эквивалентность между тем, что и как думает человек, и тем, что может рассчитать машина. Основная идея заключается в том, что человеческий мозг подчиняется законам природы и поэтому его способность обрабатывать информацию не может превышать аналогичную способность машины (в том числе машины Тьюринга).

Безусловно, статья Тьюринга 1936 г. заложила теоретические основы информатики, но важно отметить, что на

---

\* Эйнштейн говорил: «Научная теория должна быть максимально простой, но не проще того».

работу Тьюринга большое влияние оказала лекция венгерско-американского математика Джона фон Неймана (1903–1957), которую он прочел в Кембридже в 1935 г. Тьюринг использовал в своей машине концепцию фон Неймана о хранимой программе<sup>1</sup>. Верно и то, что на фон Неймана оказала влияние статья Тьюринга, где элегантно изложены принципы информатики; в конце 1930-х и в начале 1940-х гг. фон Нейман рекомендовал прочесть эту статью всем своим коллегам<sup>2</sup>.

В той же статье Тьюринг поведал о другом неожиданном открытии, касающемся неразрешимых задач. Неразрешимые задачи — это те, что хорошо описываются единственным решением (которое, как можно показать, существует), но (как тоже можно показать) не могут быть решены никакой машиной Тьюринга (то есть вообще никакой машиной). Представление о существовании таких задач в корне противоречит сформировавшейся к началу XX в. догме о том, что все проблемы, которые можно сформулировать, являются решаемыми. Тьюринг показал, что число неразрешимых задач не меньше числа разрешимых задач. В 1931 г. к такому же выводу пришел Курт Гедель, сформулировавший «теорему о неполноте». Такая странная ситуация: мы можем сформулировать задачу, можем доказать, что у нее существует единственное решение, но при этом знаем, что никогда не сможем это решение найти.

Тьюринг показал, что вычислительные машины действуют на основании очень простого механизма. Поскольку машина Тьюринга (и, следовательно, любой компьютер) может определять свою дальнейшую функцию на основе полученных ею ранее результатов, она способна принимать решения и создавать иерархические информационные структуры любой сложности.

В 1939 г. Тьюринг сконструировал электронный калькулятор Bombe, который помогал дешифровать сообщения, составленные немцами на кодирующей машине Enigma. К 1943 г. группа инженеров при участии Тьюринга закончила создание машины Colossus, которую иногда называют первым в истории

компьютером. Это позволило союзникам расшифровывать сообщения, созданные более сложной версией Enigma. Машины Bombe и Colossus были сконструированы для решения единственной задачи и не могли перепрограммироваться. Но свою функцию они выполняли блестяще. Считается, что отчасти благодаря им союзники могли предвидеть тактику немцев на протяжении всей войны, а Королевские военно-воздушные силы Великобритании в Битве за Британию смогли одолеть втрое превосходящие их по численности силы Люфтваффе.

Именно на этой основе Джон фон Нейман создал компьютер современной архитектуры, отражающей третью из четырех важнейших идей теории информации. На протяжении прошедших с тех пор почти семидесяти лет основное ядро этой машины, названной «машиной фон Неймана», практически не изменилось — как в микроконтроллере в вашей стиральной машине, так и в самом крупном суперкомпьютере. В статье, опубликованной 30 июня 1945 г. и озаглавленной «Первый проект отчета о EDVAC», фон Нейман изложил основные идеи, которые с тех пор направляли развитие информатики<sup>3</sup>. В машине фон Неймана присутствует центральный процессор, где выполняются арифметические и логические операции, модуль памяти, в котором хранятся программы и данные, массовая память, программный счетчик и входные/выходные каналы. Хотя статья предназначалась для внутреннего пользования в рамках выполнения проекта, для создателей компьютеров она стала Библией. Вот так иногда обычный рутинный отчет может изменить мир.

Машина Тьюринга не была предназначена для практических целей. Теоремы Тьюринга не имели отношения к эффективности решения задач, а скорее описывали спектр задач, которые теоретически могут быть решены с помощью компьютера. Напротив, цель фон Неймана заключалась в создании концепции реального компьютера. Его модель заменила однобитную систему Тьюринга многобитной (обычно кратную восьми битам) системой. Машина Тьюринга имеет последовательную ленту

памяти, так что программы затрачивают очень большое время на перемещение ленты вперед и назад для записи и извлечения промежуточных результатов. Напротив, в системе фон Неймана доступ к памяти осуществляется произвольным образом, что позволяет немедленно извлекать любые нужные данные.

Одной из ключевых идей фон Неймана является концепция хранимой программы, которую он развил за десять лет до создания компьютера. Суть концепции заключается в том, что программа хранится в том же модуле памяти с произвольным доступом, что и данные (а часто даже в том же блоке памяти). Это позволяет перепрограммировать компьютер для решения разных задач и создавать самомодифицирующийся код (в случае записывающих накопителей), что обеспечивает возможность рекурсии. До того времени практически все компьютеры, включая Colossus, создавались для решения конкретных задач. Концепция хранимой программы позволила компьютеру стать поистине универсальной машиной, соответствующей представлению Тьюринга об универсальности машинных вычислений.

Еще одно важное свойство машины фон Неймана заключается в том, что в каждой инструкции содержится операционный код, определяющий арифметическую или логическую операцию и адрес операнда в памяти компьютера.

Концепция фон Неймана об архитектуре компьютера отразилась в проекте EDVAC, над которым он работал совместно с Преспером Дж. Эккертом и Джоном Моучли. Компьютер EDVAC начал функционировать только в 1951 г., когда уже существовали другие компьютеры с хранимой программой, такие как Манчестерская малая экспериментальная машина, ENIAC, EDSAC и BINAC, причем все они были созданы под влиянием статьи фон Неймана и при участии Эккерта и Моучли. Фон Нейман также был причастен к появлению некоторых из этих машин, включая последнюю версию ENIAC, где использовался принцип хранимой программы.

У компьютера с архитектурой фон Неймана имелось несколько предшественников, но ни один из них — за одним неожиданным исключением — нельзя назвать истинной машиной фон Неймана. В 1944 г. Говард Эйкен выпустил Mark I, который можно было в какой-то степени перепрограммировать, но он не использовал хранимой программы. Машина считывала инструкции с перфокарты и немедленно их выполняла. В машине также не было предусмотрено условных переходов.

В 1941 г. немецкий ученый Конрад Цузе (1910–1995) создал компьютер Z-3. Он тоже считывал программу с ленты (в данном случае закодированную на пленке) и тоже не выполнял условных переходов. Интересно, что Цузе получил финансовую поддержку от Немецкого института самолетостроения, который использовал этот компьютер для изучения флаттера крыла самолета. Однако предложение Цузе о финансировании замены реле радиолампами не было поддержано нацистским правительством, которое считало развитие компьютерной технологии «не имеющим военного значения». Это, как мне кажется, в определенной степени повлияло на исход войны.

На самом деле у фон Неймана был один гениальный предшественник, причем жил он на сто лет раньше! Английский математик и изобретатель Чарльз Бэббидж (1791–1871) в 1837 г. описал свою аналитическую машину, основанную на тех же принципах<sup>4</sup>, что и компьютер фон Неймана, и использовавшую хранимую программу, нанесенную на перфокарты жаккардовых ткацких машин. Память машины с произвольным доступом содержала 1000 слов по 50 десятичных знаков в каждом (что эквивалентно примерно 21 килобайту). Каждая инструкция содержала код операции и номер операнда — точно так же, как в современных компьютерных языках. Система не использовала условных переходов и циклов, так что это была настоящая машина фон Неймана. Полностью механическая, она, по-видимому, превзошла и дизайнерские, и организаторские возможности самого Бэббиджа. Он создал части машины, но так и не запустил ее.

Точно не известно, знали ли пионеры компьютеростроения XX в., включая фон Неймана, о работах Бэббиджа\*.

Однако создание машины Бэббиджа положило начало развитию программирования. Английская писательница Ада Байрон (1815–1852), графиня Лавлейс, единственный законный ребенок поэта лорда Байрона, стала первым в мире программистом. Она писала программы для аналитической машины Бэббиджа и отлаживала их в уме (поскольку компьютер так никогда и не заработал). Теперь программисты называют эту практику *table checking*. Она перевела статью итальянского математика Луиджи Менабреа об аналитической машине, добавив от себя существенные замечания и заметив, что «аналитическая машина плетет алгебраические рисунки, как ткацкий жаккардовый станок плетет цветы и листья». Возможно, она первой упомянула о возможности создания искусственного интеллекта, но сделала вывод, что аналитическая машина «сама не способна что-либо придумать».

Идеи Бэббиджа кажутся поразительными, если учесть, в какую эпоху он жил и работал. Однако к середине XX в. эти идеи были практически забыты (и вновь открыты лишь позднее). Именно фон Нейман придумал и сформулировал ключевые принципы действия компьютера в его современном виде, и недаром машину фон Неймана продолжают считать основной моделью вычислительной машины. Однако не будем забывать, что машина фон Неймана постоянно осуществляет обмен данными между отдельными модулями и внутри этих модулей, так что она не могла быть создана без теорем Шеннона и тех методов, которые он предложил для надежной передачи и хранения цифровой информации.

Все сказанное подводит нас к четвертой важной идее, которая преодолевает выводы Ады Байрон о неспособности компьютера

---

\* Судя по всему, идеи Бэббиджа легли в основу созданного Эйкеном компьютера Mark I.

к творческому мышлению и позволяет найти ключевые алгоритмы, используемые мозгом, чтобы потом применить их для превращения компьютера в мозг. Алан Тьюринг сформулировал эту задачу в статье «Вычислительные машины и разум», опубликованную в 1950 г., в которой содержится описание теперь широко известного теста Тьюринга, позволяющего определить близость ИИ к человеческому интеллекту.

В 1956 г. фон Нейман начал готовить серию лекций для престижных Силлимановских чтений в Йельском университете. Ученый уже был болен раком и не смог ни прочесть свои лекции, ни даже закончить рукопись, на основе которой создавались лекции. Тем не менее этот незаконченный труд является блестящим предсказанием того, что лично я воспринимаю как самый трудный и важный проект в истории человечества. Уже после смерти ученого, в 1958 г., рукопись была опубликована под названием «Компьютер и мозг». Так вышло, что последний труд одного из самых блестящих математиков прошлого столетия и одного из основоположников компьютерной технологии оказался посвящен анализу мышления. Это было первое серьезное исследование человеческого мозга с точки зрения математика и специалиста в области компьютеров. До фон Неймана компьютерные технологии и нейробиология представляли собой два отдельных острова, между которыми не существовало никакого моста.

Фон Нейман начинает повествование, описывая сходство и различие между компьютером и человеческим мозгом. Учитывая, в какую эпоху создавался этот труд, он представляется удивительно точным. Ученый отмечает, что выходной сигнал нейрона цифровой — аксон либо возбуждается, либо остается в покое. В то время было далеко не очевидно, что обработка выходного сигнала может происходить аналоговым путем. Обработка сигнала в дендритах, ведущих к нейрону, и в теле нейрона аналоговая, и фон Нейман описал эту ситуацию с помощью взвешенной суммы входных сигналов с пороговым значением.

Эта модель функционирования нейронов привела к развитию коннекционизма и к использованию данного принципа для создания как аппаратного оформления, так и компьютерных программ. (Как я рассказывал в предыдущей главе, первая такая система, а именно программа для IBM 704, была создана Фрэнком Розенблаттом из Корнельского университета в 1957 г., сразу после того, как стала доступна рукопись лекций фон Неймана.) Теперь у нас есть более сложные модели, описывающие сочетания входных сигналов нейронов, но общая идея об аналоговой обработке сигналов с помощью изменения концентрации нейромедиаторов по-прежнему верна.

На основе концепции универсальности компьютерных вычислений фон Нейман пришел к выводу, что даже при кажущемся радикальном различии архитектуры и структурных единиц головного мозга и компьютера с помощью машины фон Неймана мы можем симулировать происходящие в головном мозге процессы. Обратный постулат, однако, не является справедливым, поскольку головной мозг не является машиной фон Неймана и не имеет хранимой программы (хотя в голове мы можем симулировать действие очень простой машины Тьюринга). Алгоритмы или методы функционирования мозга определены его структурой. Фон Нейман пришел к справедливому заключению, что нейроны могут выучивать соответствующие образы на основании входных сигналов. Однако во времена фон Неймана не было известно, что обучение также происходит путем создания и разрушения контактов между нейронами.

Фон Нейман также указал, что скорость обработки информации нейронами очень низкая — порядка сотни вычислений в секунду, однако мозг компенсирует это тем, что одновременно осуществляет обработку информации во множестве нейронов. Это еще одно очевидное, но очень важное открытие. Фон Нейман утверждал, что все  $10^{10}$  нейронов мозга (эта оценка также достаточно точна: по сегодняшним представлениям, в головном мозге содержится от  $10^{10}$  до  $10^{11}$  нейронов) обрабатывают сигналы

в одно и то же время. Более того, все контакты (в среднем от  $10^3$  до  $10^4$  на каждый нейрон) обсчитываются одновременно.

Учитывая примитивный уровень развития нейробиологии того времени, оценки и описания функции нейронов, сделанные фон Нейманом, удивительно точны. Однако я не могу согласиться с одним аспектом его работы, а именно с представлением об объеме памяти мозга. Он считал, что мозг запоминает каждый сигнал на всю жизнь. Среднюю продолжительность жизни человека фон Нейман оценивал в 60 лет, что составляет примерно  $2 \times 10^9$  секунды. Если каждый нейрон за одну секунду получает примерно 14 сигналов (что на самом деле на три порядка ниже истинной величины), а всего в головном мозге содержится  $10^{10}$  нейронов, выходит, что объем памяти мозга составляет около  $10^{20}$  бит. Как я писал выше, мы запоминаем лишь небольшую часть наших мыслей и опыта, но даже эти воспоминания хранятся не как побитовая информация низкого уровня сложности (как в видео), а скорее в виде последовательности образов более высокого порядка.

По мере того как фон Нейман описывает каждый механизм в функции головного мозга, он одновременно демонстрирует, как современный компьютер мог бы осуществить ту же самую функцию, несмотря на кажущееся различие между мозгом и компьютером. Аналоговые механизмы действия мозга можно моделировать с помощью цифровых механизмов, поскольку цифровые вычисления способны моделировать аналоговые значения с любой степенью точности (а точность передачи аналоговой информации в мозге достаточно низкая). Можно также имитировать массивный параллелизм функции мозга, учитывая значительное превосходство компьютеров по скорости серийных вычислений (со времен фон Неймана это превосходство еще более усилилось). Кроме того, мы можем осуществлять параллельную обработку сигналов в компьютерах с помощью параллельно функционирующих машин фон Неймана — именно так действуют современные суперкомпьютеры.

Учитывая способность людей быстро принимать решения при столь низкой скорости работы нейронов, фон Нейман пришел к выводу, что функции головного мозга не могут задействовать длинные последовательные алгоритмы. Когда третий бейсмен получает мяч и решает бросить его на первую, а не на вторую базу, он принимает это решение за какую-то долю секунды — за это время каждый нейрон едва успевает осуществить несколько циклов возбуждения. Фон Нейман приходит к логичному выводу, что замечательная способность мозга связана с тем, что все 100 млрд нейронов могут обрабатывать информацию одновременно. Как я отмечал выше, зрительная кора делает сложные выводы всего за три или четыре цикла возбуждения нейронов.

Именно значительная пластичность мозга позволяет нам обучаться. Однако компьютер обладает гораздо большей пластичностью — его методы можно полностью изменить путем смены программного обеспечения. Таким образом, компьютер может имитировать мозг, а вот обратное утверждение неверно.

Когда фон Нейман сравнивал возможности массивированной параллельной активности мозга с немногочисленными компьютерами того времени, казалось очевидным, что мозг отличается гораздо большей памятью и скоростью. Сегодня уже сконструирован первый суперкомпьютер, по самым консервативным оценкам, удовлетворяющий тем функциональным требованиям, которые нужны для моделирования функций человеческого головного мозга (около  $10^{16}$  операций в секунду)<sup>5</sup>. (По моему мнению, компьютеры такой мощности в начале 2020-х гг. будут стоить около 1000 долларов.) Что касается объема памяти, мы продвинулись еще дальше. Труд фон Неймана появился в самом начале компьютерной эры, но ученый был уверен в том, что в какой-то момент мы сможем создавать компьютеры и компьютерные программы, способные имитировать человеческий мозг; именно поэтому он и готовил свои лекции.

Фон Нейман был глубоко убежден в ускорении прогресса и в его значительном влиянии на жизнь людей в будущем. Через

год после смерти фон Неймана, в 1957 г., его коллега математик Стэн Юлам цитировал слова фон Неймана, сказавшего в начале 1950-х гг., что «любое ускорение технологического прогресса и изменения образа жизни людей создает впечатление приближения некоей важнейшей сингулярности в истории человеческой расы, за пределами которой человеческая деятельность в том виде, какой мы знаем ее сегодня, больше не может продолжаться». Это первый известный случай использования слова «сингулярность» для описания технологического прогресса человечества.

Важнейшая догадка фон Неймана заключалась в обнаружении сходства между компьютером и мозгом. Заметим, что частью человеческого интеллекта является эмоциональный интеллект. Если догадка фон Неймана верна и если согласиться с моим утверждением, что небиологическая система, удовлетворительно воспроизводящая интеллект (эмоциональный и другой) живого человека, обладает сознанием (см. следующую главу), придется сделать вывод, что между компьютером (*с правильным программным обеспечением*) и *сознательным* мышлением имеется явное сходство. Итак, был ли прав фон Нейман?

Большинство современных компьютеров — полностью цифровые машины, тогда как человеческий мозг использует как цифровые, так и аналоговые методы. Однако аналоговые методы легко воспроизводятся в цифровом варианте с любой степенью точности. Американский специалист в области компьютерных технологий Карвер Мид (род. в 1934 г.) показал, что аналоговые методы мозга можно напрямую воспроизвести в кремниевом варианте, и реализовал это в виде так называемых нейроморфных чипов<sup>6</sup>. Мид продемонстрировал, что данный подход может быть в тысячи раз более эффективным, чем цифровая имитация аналоговых методов. Если речь идет о кодировании избыточных алгоритмов новой коры, возможно, имеет смысл воспользоваться идеей Мида. Исследовательская группа ИВМ под руководством Дхармендра Модхи применяет чипы, имитирующие нейроны

и их контакты, в том числе их способность образовывать новые контакты<sup>7</sup>. Один из чипов, названный SyNAPSE, напрямую модулирует 256 нейронов и примерно четверть миллиона синаптических связей. Цель проекта заключается в симуляции новой коры, состоящей из 10 млрд нейронов и 100 трлн контактов (что эквивалентно человеческому мозгу), которая использует всего один киловатт энергии.

Более пятидесяти лет назад фон Нейман заметил, что процессы в головном мозге происходят чрезвычайно медленно, но отличаются массивной параллельностью. Современные цифровые схемы действуют как минимум в 10 млн раз быстрее, чем электрохимические переключатели мозга. Напротив, все 300 млн распознающих модулей коры мозга действуют одновременно, и квадрильон контактов между нейронами может активизироваться в одно и то же время. Следовательно, для создания компьютеров, которые могли бы адекватно имитировать человеческий мозг, необходимы соответствующий объем памяти и производительность вычислений. Нет нужды напрямую копировать архитектуру мозга — это очень неэффективный и негибкий метод.

Какими же должны быть соответствующие компьютеры? Многие исследовательские проекты направлены на моделирование иерархического обучения и распознавания образов, происходящих в новой коре. Я сам занимаюсь подобными исследованиями с привлечением иерархических скрытых моделей Маркова. По моим оценкам, для моделирования одного цикла распознавания в одном распознающем модуле биологической новой коры требуется около 3000 вычислений. Большинство симуляций построено на значительно меньшем числе вычислений. Если принять, что головной мозг осуществляет около  $10^2$  (100) циклов распознавания в секунду, получаем общее число  $3 \times 10^5$  (300 тыс.) вычислений в секунду для одного распознающего модуля. Если же умножить это число на общее число распознающих модулей ( $3 \times 10^8$  (300 млн, по моим оценкам)),

получаем  $10^{14}$  (100 трлн) вычислений в секунду. Примерно такое же значение я привожу в книге «Сингулярность уже близка». По моим прогнозам, для функциональной симуляции головного мозга требуется скорость от  $10^{14}$  до  $10^{16}$  калькуляций в секунду. По оценкам Ганса Моравека, основанным на экстраполяции данных для начальной обработки зрительных сигналов во всем головном мозге, это значение составляет  $10^{14}$  калькуляций в секунду, что совпадает с моими расчетами.

Стандартные современные машины могут работать со скоростью до  $10^{10}$  калькуляций в секунду, однако с помощью ресурсов облака их производительность можно существенно увеличить. Самый быстрый суперкомпьютер, японский компьютер «К», уже достиг скорости  $10^{16}$  калькуляций в секунду<sup>8</sup>. Учитывая массивную избыточность алгоритмов новой коры, хороших результатов можно добиться с помощью нейроморфных чипов, как в технологии SyNAPSE.

Что касается требований к памяти, нам нужно около 30 бит (примерно 4 байта) для каждого контакта с одним из 300 млн распознающих модулей. Если к каждому распознающему модулю подходит в среднем восемь сигналов, получаем 32 байта на распознающий модуль. Если учесть, что вес каждого входного сигнала составляет один байт, получаем 40 байт. Добавим 32 байта для нисходящих контактов — и получим 72 байта. Замечу, что наличие восходящих и нисходящих разветвлений приводит к тому, что число сигналов намного больше восьми, даже если учесть, что многие распознающие модули пользуются общей сильно разветвленной системой связей. Например, в распознавании буквы «р» могут участвовать сотни распознающих модулей. Это означает, что тысячи распознающих модулей следующего уровня участвуют в распознавании слов и фраз, содержащих букву «р». Однако каждый модуль, ответственный за распознавание «р», не повторяет это древо связей, питающих все уровни распознавания слов и фраз с «р», у всех этих модулей древо связей общее.

Сказанное выше верно и для нисходящих сигналов: модуль, ответственный за распознавание слова apple, сообщит всей тысяче стоящих ниже модулей, ответственных за распознавание «е», что ожидается образ «е», если уже распознаны «а», «р», «р» и «l». Это древо связей не повторяется для каждого модуля, распознающего слово или фразу, который хочет информировать модули нижестоящего уровня, что ожидается образ «е». Это древо общее. По этой причине среднее оценочное значение в восемь восходящих и восемь нисходящих сигналов для каждого распознающего модуля является вполне разумным. Но даже если мы повысим это значение, это не сильно изменит конечный результат.

Итак, с учетом  $3 \times 10^8$  (300 млн) распознающих модулей и 72 байт памяти для каждого, получаем, что общий объем памяти должен составлять около  $2 \times 10^{10}$  (20 млрд) байт. А это весьма скромное значение. Такой памятью обладают обычные современные компьютеры.

Все эти расчеты мы выполнили для приблизительной оценки параметров. Учитывая, что цифровые схемы примерно в 10 млн раз быстрее сетей нейронов в биологической коре, нам не нужно воспроизводить массивированный параллелизм человеческого мозга — весьма умеренного параллельного процессинга (по сравнению с триллионным параллелизмом в головном мозге) будет вполне достаточно. Таким образом, необходимые вычислительные параметры вполне достижимы. Способность нейронов головного мозга к переподключению (помним, что дендриты постоянно создают новые синапсы) тоже можно имитировать с помощью соответствующего программного обеспечения, поскольку компьютерные программы гораздо пластичнее биологических систем, которые, как мы видели, впечатляют, но имеют пределы.

Избыточность мозга, необходимая для получения инвариантных результатов, безусловно, может быть воспроизведена в компьютерном варианте. Математические принципы оптимизации

подобных самоорганизующихся иерархических систем обучения вполне понятны. Организация мозга далеко не оптимальна. Но она и не должна быть оптимальной — она должна быть достаточно хорошей, чтобы обеспечить возможность создавать инструменты, компенсирующие ее собственные ограничения.

Еще одно ограничение новой коры заключается в том, что в ней нет механизма, устранивающего или хотя бы оценивающего противоречащие друг другу данные; отчасти это объясняет весьма распространенную нелогичность человеческих рассуждений. Для решения данной проблемы у нас есть весьма слабая способность, называемая *критическим мышлением*, но люди ею пользуются гораздо реже, чем следовало бы. В компьютерной новой коре можно предусмотреть процесс, выявляющий противоречащие данные для их последующего пересмотра.

Важно отметить, что конструирование целого отдела мозга осуществить проще, чем конструирование одного нейрона. Как уже было сказано, на более высоком уровне иерархии модели часто упрощаются (тут просматривается аналогия с компьютером). Чтобы понять, как работает транзистор, нужно в деталях понимать физику полупроводниковых материалов, а функции одного реального транзистора описываются сложными уравнениями. Цифровая схема, осуществляющая перемножение двух чисел, содержит сотни транзисторов, но для создания модели такой схемы хватит одной или двух формул. Целый компьютер, состоящий из миллиардов транзисторов, можно смоделировать с помощью набора инструкций и описания регистра на нескольких страницах текста с привлечением нескольких формул. Программы для операционных систем, компиляторов языков или ассемблеров достаточно сложны, однако моделирование частной программы (например, программы распознавания языка на основе скрытых иерархических моделей Маркова) тоже сводится к нескольким страницам формул. И нигде в подобных программах вы не встретите детального описания физических свойств полупроводников или даже компьютерной архитектуры.

## Глава восьмая

Аналогичный принцип верен и для моделирования мозга. Один конкретный распознающий модуль новой коры, который детектирует определенные инвариантные зрительные образы (например, лица), осуществляет фильтрацию звуковых частот (ограничивая входной сигнал определенным диапазоном частот) или оценивает временную близость двух событий, можно описать с помощью гораздо меньшего числа специфических деталей, чем реальные физические и химические взаимодействия, контролирующие функции нейромедиаторов, ионных каналов и других элементов нейронов, участвующих в передаче нервного импульса. Хотя все эти детали необходимо тщательно предусмотреть до перехода на следующий уровень сложности, при моделировании операционных принципов головного мозга многое можно упростить.

# Глава девятая

## МЫСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ПО АНАЛИЗУ РАЗУМА

Разум — это просто то, что делает мозг.

*Марвин Минский. Общество разума, 1986*

При конструировании разумных машин не следует удивляться, если они окажутся столь же непоследовательными и упрямыми, как и люди, в вопросах разума, сознания, свободы воли и т. д.

*Марвин Минский*

**Кто обладает сознанием?**

Подлинная история вашего сознания начинается с первой лжи.

*Иосиф Бродский*

Страдание — единственная причина сознания.

*Федор Достоевский*

Есть растения, которые питаются органической пищей при помощи своих цветов: когда муха садится на цветок, лепестки закрываются и удерживают ее, пока растение не всосет насекомое в себя; но цветок закрывается только тогда, когда в него попадает что-то пригодное для еды; капля дождя или кусок веточки не привлекают его внимания. Удивительно, что столь несознательное существо с таким вниманием преследует свои интересы. Если это отсутствие сознания, то в чем же заключается сознание?

*Сэмюэл Батлер, 1871*

До сих пор мы рассматривали мозг как некую систему, способную осуществлять функции определенной сложности. Однако в таком ракурсе *мы сами* оставались вне поля зрения. Кажется, что мы живем в нашем мозге, что наша жизнь субъективна. Как объективный взгляд на мозг, которого мы придерживались до сих пор, связан с нашими ощущениями, с нашим существом, обладающим некоторым опытом?

Британский философ Колин Макгинн (род. в 1950 г.) пишет, что «сознание может довести даже самого утонченного мыслителя до бессвязного бормотания». Суть в том, что люди часто не понимают точного смысла этого термина.

Многие наблюдатели рассматривают сознание в качестве формы проявления личности, например способности обдумывать собственные мысли и объяснять их. Я предпочитаю определять сознание как способность думать о своем мыслительном процессе. Казалось бы, мы в состоянии оценить эту способность и применить данный тест для того, чтобы отличать сознательное от несознательного.

Однако выясняется, что применить этот подход достаточно сложно. Сознателен ли новорожденный ребенок? Сознательна ли собака? Они не могут описать свой мыслительный процесс. Некоторые люди полагают, что младенцы и собаки не являются сознательными существами именно по той причине, что не могут объясниться. А что можно сказать о компьютере по имени Ватсон? В определенном режиме он может объяснить, каким образом он пришел к тому или иному ответу. Поскольку у Ватсона есть модель собственного мышления, выходит, он является сознательным существом, а младенцы и собаки — нет?

Прежде чем глубже изучить этот вопрос, важно рассмотреть один его важный аспект: что говорит нам на эту тему наука и что остается в ведении философии? Одна точка зрения заключается в том, что философия занимается вопросами, для которых еще не выработан научный метод изучения. Если это так, то по мере прогресса науки, достаточного для разрешения определенного

круга вопросов, философы переключаются на другие проблемы, пока наука не доберется и до них. В вопросах, касающихся сознания, такая точка зрения весьма популярна. В том числе она относится к вопросу о том, кто и что обладает сознанием.

Вот что пишет на эту тему философ Джон Серль: «Мы знаем, что мозг формирует сознание при помощи специфических биологических механизмов... Важно признать, что сознание — это биологический процесс, как пищеварение, лактация, фотосинтез или митоз... Мозг — это машина, точнее, биологическая машина, но все же машина. Так что на первом этапе нужно понять, как действует мозг, а затем создать искусственную машину, обладающую столь же эффективным механизмом формирования сознания»<sup>1</sup>. Люди часто удивляются, прочтя это высказывание, поскольку считают, что Серль трепетно защищает тайну сознания от таких редуccionистов, как Рэй Курцвейл.

Австралийский философ Дэвид Чалмерс (род. в 1966 г.) придумал термин «трудная проблема сознания» для отражения сложностей, возникающих при попытках точно сформулировать эту неясную концепцию. Иногда емкое выражение вбирает в себя всю суть какого-либо направления мысли, так что оно становится эмблематичным (как выражение Ханны Арендт «банальность зла»<sup>\*</sup>). Знаменитое выражение Чалмерса тоже прекрасно справляется с этой задачей.

При обсуждении проблем сознания очень легко соскользнуть на рассмотрение тех наблюдаемых и измеряемых атрибутов, которые мы связываем с сознанием, но при этом теряется сама суть идеи. Это концепция метакогнитивности (осмысление собственного мышления). Другие наблюдатели объединяют сознание и эмоциональный или моральный интеллект. Но и тут, опять же, наша способность выражать любовь, шутить или быть сексуально привлекательным — просто проявления нашей на-

---

<sup>\*</sup> Книга Ханны Арендт (1906–1975) о Холокосте вышла в 1963 г.; в русском переводе: «Банальность зла: Эйхман в Иерусалиме» (М.: Европа, 2008).

туры, возможно, впечатляющие и разумные, но именно те, что можно наблюдать и измерить (даже если способы их оценки неочевидны). Понимание того, как мозг выполняет эти функции и что в нем в это время происходит, относится, в терминологии Чалмерса, к разряду «легких» проблем сознания. Конечно, эту проблему неправильно назвать «легкой»; она относится к числу самых сложных и важных научных вопросов современности. Однако «трудная» проблема Чалмерса настолько трудна, что ее практически невозможно сформулировать.

Чтобы продемонстрировать это разграничение, Чалмерс использует мысленный эксперимент с участием «зомби». «Зомби» из этого эксперимента действует абсолютно так же, как человек, но только не имеет субъективного опыта, то есть является существом несознательным. Чалмерс утверждает, что раз мы в состоянии выдумать зомби, то по крайней мере логически их существование возможно. Представьте себе, что вы на вечеринке, где присутствуют и зомби, и «нормальные» люди. Как вы их различите? Возможно, вы даже бывали на такой вечеринке.

Многие люди считают, что следовало бы расспросить «испытываемых» об их эмоциональных реакциях на события или идеи. Они полагают, что зомби проявят отсутствие субъективного опыта, так как затруднятся ответить на некоторые эмоциональные вопросы. Однако ответы такого рода просто не учитывают допущений мысленного эксперимента. В эксперименте мы можем встретить неэмоционального человека (подобный эмоциональный дефицит, например, бывает у людей с некоторыми формами аутизма), аватара или робота, не обладающего человеческими эмоциями, но они не зомби. Вспомним: в соответствии с определением Чалмерса, зомби *совершенно нормально* отвечает на вопросы и способен эмоционально реагировать, у него нет только субъективного опыта. Проблема в том, что идентифицировать зомби нам не удастся, поскольку по определению в его поведении не проявляются никакие признаки его природы. Так что же это за разграничение без различий?

Чалмерс не пытается ответить на трудный вопрос, но предоставляет несколько возможностей. Одна возможность — это вариант дуализма, в рамках которого сознание существует в физическом мире не само по себе, а скорее как отдельная онтологическая реальность. В соответствии с этой формулировкой суть личности определяется процессами в головном мозге этой личности. Поскольку мозг — закрытая система, мы можем полностью объяснить действия человека, включая его мысли, через происходящие в мозге процессы. Тогда получается, что сознание относится к другому миру или отделено от физического мира. В рамках этой концепции разум (то есть сознательное свойство мозга) не способен влиять на мозг.

Другая рассмотренная Чалмерсом возможность логически не противоречит описанному выше варианту дуализма. Эта идея, которую часто называют панпротопсихизмом, подразумевает, что все физические системы обладают сознанием, хотя человек более сознателен, чем, скажем, электрический выключатель. Я, безусловно, согласен с тем, что у человеческого мозга есть гораздо больше тем для осознания, чем у выключателя.

Мое собственное мнение, которое, возможно, совпадает с неким течением в рамках панпротопсихизма, заключается в том, что сознание — это свойство, возникающее на уровне сложных физических систем. С этой точки зрения собака тоже обладает сознанием, хотя и меньшим, чем человек. И у муравья есть некий уровень сознания, но гораздо более низкий, чем у собаки. С другой стороны, можно считать, что колония муравьев обладает более высоким уровнем сознания, чем один муравей: большая группа муравьев, безусловно, более разумна, чем один муравей. Так мы приходим к выводу, что компьютер, успешно моделирующий сложный человеческий разум, должен обладать таким же сознанием, что и сам человек.

Еще один путь определения сознания заключается в том, чтобы охарактеризовать его через понятие «квалиа». Что такое квалиа? В соответствии с одним определением, это «осознанный

опыт». Однако данная формулировка не позволяет нам продвигаться вперед. Давайте проведем такой мысленный эксперимент. Один нейробиолог полностью лишен способности различать цвета. Это не дальтонизм, при котором человек не различает оттенки, скажем, красного и зеленого (как я). Это такое состояние, когда человек живет в полностью черно-белом мире. (В крайней версии такого сценария человек вырос в черно-белом мире и никогда не видел никаких цветов. Короче говоря, в его мире цвета отсутствуют.) Однако этот человек активно изучал физические основы цвета и знает, что красный свет — это свет с длиной волны 700 нм, а также ему известна суть неврологических процессов в мозге человека, способного нормально воспринимать цвета. Он знает о цвете больше, чем обычно знают люди. Если бы вы захотели объяснить этому человеку, что же такое на самом деле «красный цвет», как бы вы поступили?

Возможно, вы прочли бы фрагмент из стихотворения «Красный» нигерийского поэта Олусеи Олусеуна:

Красный — цвет крови  
символ жизни  
Красный — цвет опасности  
символ смерти  
  
Красный — цвет роз  
символ красоты  
Красный — цвет любовников  
символ слияния  
  
Красный — цвет томатов  
символ здоровья  
Красный — цвет огня  
символ желания

Это действительно может помочь человеку представить себе, с чем ассоциируется красный цвет у других людей, и даже позволит поддержать беседу о цветах. («Да, я люблю красный

цвет, он столь горяч и горд, столь опасно прекрасен...») И если этот человек захочет, он в принципе может убедить других людей в том, что у него есть личный опыт, связанный с красным цветом, но никакие в мире стихи не помогут ему этот опыт пережить.

Аналогичным образом, как вы объясните человеку, никогда не касавшемуся воды, что испытывает ныряльщик? Вам опять придется прибегать к стихам, но это не поможет обзавестись личным опытом. Именно этот опыт мы и называем словом квалиа.

Многие из тех, кто читает мою книгу, по собственному опыту знают, что такое красный цвет. Но откуда мне знать, не испытываете ли вы при виде красного цвета то же самое, что я испытываю при виде синего? Мы оба смотрим на красный предмет и утверждаем, что он красный, но это не ответ на вопрос. Может быть, я при этом испытываю то, что вы испытываете, когда смотрите на синий, просто мы оба выучили, что этот цвет называется красным. Мы опять можем прибегнуть к помощи стихов, но они лишь отражают ассоциации людей с тем или иным цветом и ничего не говорят о реальной природе квалиа. На самом деле слепые от рождения люди много читают о цветах, поскольку о цветах часто пишут, и поэтому у них есть некоторый собственный опыт, связанный с цветами. Как сравнить их опыт «красного цвета» с опытом зрячего человека? Это, по сути, тот же вопрос, что и вопрос о человеке, живущем в черно-белом мире. Удивительно, что столь распространенные в нашей жизни явления настолько сложны, что даже невозможно утверждать, что мы имеем один и тот же чувственный опыт (квалиа).

Другое определение квалиа — субъективный опыт. Однако это определение тоже ведет в тупик, как и наши попытки дать определение сознания, поскольку слова и выражения «ощущение», «личный опыт» и «сознание», по сути, являются синонимами. Что есть сознание и что есть квалиа — фундаментальные, возможно, последние вопросы философии (хотя есть еще более

важный вопрос, связанный с определением личности, мы поговорим об этом в заключительном разделе данной главы).

Возвращаясь к теме сознания, постараемся точнее сформулировать волнующий нас вопрос. Он звучит так: кто или что обладает сознанием? В заглавии книги я использовал слово «разум», а не слово «мозг», поскольку разум — свойство сознательного мозга. Мы также можем сказать, что разум обладает свободой воли и идентифицирует себя. Утверждение, что эти вопросы относятся к разряду философских вопросов, не совсем очевидно. Я имею в виду, что эти вопросы не могут быть разрешены исключительно научными методами. Другими словами, нет таких фальсифицируемых\* экспериментов, которые могли бы ответить на эти вопросы — без привлечения философских приемов. Если бы мы построили детектор сознания, Серль обязательно пожелал бы убедиться, что он испускает биологические нейромедиаторы. Американский философ Дэн Деннет (род. в 1942 г.) был бы более податлив, но, возможно, захотел бы установить, содержит ли система модель самой себя и своих возможностей. Эта точка зрения ближе к моей собственной, но в своей основе по-прежнему является философской.

Регулярно возникает идея связать сознание с каким-то измеряемым физическим параметром; Серль называет это «механизмом стимуляции сознания». Американский ученый, философ и анестезиолог Стюарт Хамерофф (род. в 1947 г.) писал, что «филаменты цитоскелета — это корни сознания»<sup>2</sup>. Речь идет об имеющихся во всех клетках (в том числе и в нейронах) тонких нитях, называемых микротрубочками, которые обеспечивают структурную целостность клетки и играют важную роль в ее делении. Книги и статьи Хамероффа на эту тему содержат

---

\* В математической терминологии фальсифицируемость (принципиальная возможность опровержения) является критерием научности теории; то есть теория является научной, если существует методологическая возможность ее экспериментального опровержения.

подробные описания и уравнения, объясняющие возможную роль микротрубочек в информационных процессах в клетке. Но, чтобы перейти от микротрубочек к сознанию, нужно сделать очень смелый шаг, не менее смелый, чем религиозная идея о том, что высшее существо наградило сознанием (иногда именуемым душой) некоторых существ (главным образом человека). Идея Хамероффа подтверждается одним доказательством, заключающимся в том, что неврологические процессы, которые могли бы поддерживать данные клеточные перестройки, прекращаются под действием наркоза. Однако это наблюдение нельзя считать исчерпывающим доказательством, учитывая значительное количество процессов, которые останавливаются в этих условиях. Мы даже не знаем точно, обладает ли сознанием находящийся под наркозом человек. Нам известно только, что, проснувшись после наркоза, люди ничего не помнят. Однако даже это не совсем верно, поскольку некоторые люди достаточно отчетливо помнят какие-то события, происходившие, пока они были под наркозом, например разговоры врачей. Такие случаи «бодрствования под наркозом» в США отмечаются примерно 40 тыс. раз в год<sup>3</sup>. Но речь вообще-то не об этом, сознание и память — совершенно не одно и то же. Как я уже писал, если я пытаюсь вспомнить прошедший день во всех подробностях, я нахожу множество чувственных ощущений, но помню лишь некоторые вызвавшие их события. Означает ли это, что большую часть дня я был без сознания? Это очень хороший вопрос, и ответ на него совсем не очевиден.

Английский физик и математик Роджер Пенроуз (род. в 1931 г.) делает другой смелый шаг для объяснения происхождения сознания, но тоже связанный с микротрубочками, в данном случае как с носителями некоторых квантовых свойств. Суть его рассуждений мне не совсем ясна, но в целом сводится к тому, что, поскольку сознание является загадочным феноменом и квантовые явления тоже загадочны, между ними должна существовать какая-то связь.

Пенроуз начал свои исследования с изучения теоремы Тьюринга о неразрешимых задачах и теоремы Геделя о неполноте. Тьюринг утверждал (мы подробно обсуждали это в восьмой главе), что существуют алгоритмические задачи, которые можно сформулировать, но нельзя решить с помощью машины Тьюринга. Учитывая универсальность вычислительных способностей машины Тьюринга, приходится заключить, что эти «неразрешимые задачи» не под силу никакой машине. Теорема Геделя о неполноте приводит к тому же результату в применении к формальной арифметике. Пенроуз же утверждает, что человеческий мозг способен решить эти неразрешимые задачи, то есть может делать вещи, которые такие детерминистские машины, как компьютеры, делать не в состоянии. Его мотивация, по крайней мере отчасти, заключается в том, чтобы поставить человека выше машины. Однако его главный тезис — о том, что человек может решить неразрешимые задачи Тьюринга и Геделя, — просто-напросто неверен.

Одна из таких неразрешимых задач, известная как задача про усердного бобра (Busy Beaver), заключается в следующем: нужно найти максимальное число единиц, которое машина Тьюринга с определенным числом состояний ( $n$ ) может записать на своей ленте. Для ответа на вопрос нужно создать все возможные машины Тьюринга с  $n$ -состояниями (это конечное число, если значение  $n$  — конечное число) и определить максимальное число единиц, которые эти машины печатают на своей ленте, за исключением зациклившихся машин Тьюринга. Ответить на этот вопрос нельзя, поскольку при попытке моделирования всех этих машин наш симулятор зациклится при попытке симулировать машину, которая уже зациклилась. Тем не менее компьютеры решают эту задачу для некоторых значений  $n$ . Справляются с задачей и математики, но компьютеры решают задачу для гораздо большего числа  $n$ , чем не вооруженные вычислительной машиной люди. Вообще обычно компьютеры лучше людей решают неразрешимые задачи Тьюринга и Геделя.

Пенроуз связал эти «трансцендентные способности» человеческого мозга с квантовыми вычислениями, которые, с его точки зрения, там происходят. В соответствии с этой точкой зрения, квантовые явления в нейронах по какой-то причине доступны не компьютерам, а только человеческому мозгу. На самом деле многие электронные устройства используют квантовые явления (транзисторы применяют квантовое явление туннельного эффекта), а вот квантовые явления в головном мозге пока никто не обнаружил. Человеческие мыслительные возможности удовлетворительно объясняются классическими вычислительными методами, и в любом случае ничто не мешает нам применять квантовые вычисления в компьютерах. Однако ни одну из этих задач Пенроуз перед собой не ставил. Хамерофф и Пенроуз объединили свои усилия, когда критики заявили, что мозг — это не то место, где могут производиться квантовые вычисления.

Пенроуз нашел в нейронах подходящий элемент для осуществления квантовых вычислений — те самые микротрубочки, в которых, по мнению Хамероффа, происходит обработка информации. Таким образом, тезис Хамероффа—Пенроуза заключается в том, что микротрубочки нейронов совершают квантовые вычисления, что и является основой сознания.

Этот тезис подвергся критике, например со стороны шведско-американского физика и космолога Макса Тегмарка (род. в 1967 г.), который определил, что квантовые явления в микротрубочках могут продолжаться лишь  $10^{-13}$  с, что слишком мало и для того, чтобы произвести сколько-нибудь значимые вычисления, и для того, чтобы как-то повлиять на процессы в нейронах. Существует ряд задач, для решения которых квантовые вычисления имеют преимущества перед классическими методами, например для дешифровки кодов путем факторизации больших чисел. Однако известно, что человеческий мозг очень плохо справляется с такими задачами и не может соперничать даже с обычными компьютерными вычислениями, и это подтверждает, что мозг не имеет никаких способностей к квантовым

вычислениям. Более того, если бы даже возможность квантовых вычислений в головном мозге существовала, она вовсе не обязательно имела бы отношение к сознанию.

### **Ты должен верить\***

Что за мастерское создание — человек! Как благороден разумом! Как беспределен в своих способностях, обличьях и движениях! Как точен и чудесен в действии! Как он похож на ангела глубоким постижением! Как он похож на некоего бога! Краса вселенной! Венец всего живущего! А что для меня эта квинтэссенция праха?

*Уильям Шекспир. Гамлет, перевод М. Лозинского*

На самом деле все подобные теории требуют сделать некий смелый шаг, некий «скачок веры». Когда дело касается вопросов сознания, руководящий принцип звучит так: «Ты должен верить». Таким образом, чтобы ответить на вопросы, кто и что обладает сознанием и что означает для нас собственное сознание, каждый из нас должен сделать этот шаг. Иначе по утрам мы не могли бы встать с постели. Однако следует быть честным в отношении того, что для каждого из нас подразумевает этот шаг.

А для каждого человека это особый шаг, хотя может показаться, что это не так. Личные философские представления любого человека о природе и источнике сознания определяют разногласия между людьми по таким вопросам, как права животных или допустимость абортов, и в будущем вызовут еще более серьезные разногласия по вопросу о правах машин. По моему мнению, когда машины смогут обсуждать с людьми вопросы субъективного опыта, люди согласятся с тем, что машины обладают сознанием. Машины овладеют самым разнообразным эмоциональным опытом, смогут заставлять нас смеяться и плакать и будут чрезвычайно недовольны, если мы станем утверждать, что они

---

\* Строчка из песни группы Europe.

не обладают сознанием. (Они будут очень умными, и нам не захочется, чтобы это произошло.) В итоге нам придется прийти к выводу, что они сознательны. Для меня смелый шаг заключается в том, чтобы признать, что, когда машины смогут убеждать нас в разговоре в наличии у них субъективного опыта, они на самом деле будут сознательными. Я пришел к этому выводу в результате следующего мысленного эксперимента. Представьте, что в будущем вы встретите существо (робота или аватара), эмоциональные реакции которого абсолютно естественны. Оно убедительно смеется над вашими шутками и само заставляет вас смеяться и плакать (не от физической боли). Оно убедительно рассказывает о своих страхах и желаниях. И во всех своих проявлениях кажется совершенно сознательным. Оно действительно похоже на человека. Сможете ли вы воспринимать это существо как сознательную личность?

Если вы скажете, что с большой вероятностью сумеете найти путь выявить небиологическую природу этого существа, значит, вы не учитываете условий этого эксперимента. А условия таковы, что такое существо *совершенно убедительно*. Учитывая данное условие, если бы это существо, как человек, испытывало страх, когда ему угрожают, думаю, вы сопереживали бы ему так же, как если бы речь шла о живом человеке. Лично для меня это так, и я думаю, что все люди, по крайней мере большинство, согласятся со мной — вне зависимости от того, какой точки зрения они сегодня придерживаются в философских спорах. И вновь ключевым словом здесь является слово «убедительно».

Понятно, что существуют разногласия по поводу того, когда мы встретим подобных существ и встретим ли вообще. Лично я считаю, что такие существа начнут появляться в 2029 г., а в 2030-х гг. станут обычным делом. Но вне зависимости от того, когда это произойдет, я думаю, что в итоге мы начнем относиться к ним как к сознательным существам. Посмотрите, как мы относимся к таким героям в книгах и в кино: R2D2 в «Звездных войнах», Дэвид и Тедди в «Искусственном разуме», Дейта

в телевизионном сериале «Звездный путь: следующее поколение», Джонни 5 в «Коротком замыкании», Валли в одноименном мультфильме, добрый терминатор Т-800 в «Терминаторе-2», репликант Рейчел (которая даже не знает, что она не человек) в «Бегущем по лезвию», Бамблби\* в фильмах и мультфильмах о трансформерах и Сонни в фильме «Я, робот». Мы симпатизируем этим героям, хотя и знаем, что это небиологические существа. Мы воспринимаем их как обладающих сознанием, как воспринимали бы на их месте живых людей. Мы сочувствуем им и переживаем, когда они попадают в трудные ситуации. Если мы сегодня так относимся к вымышленным небиологическим существам, то как же мы будем воспринимать реальных разумных существ, которым не посчастливилось возникнуть из биологического субстрата?

И если вы согласны сделать этот шаг и признать, что небиологические существа, убедительно доказывающие свой эмоциональный опыт, действительно обладают сознанием, то это означает следующее: сознание является свойством всего существа в целом, а не того субстрата, из которого оно состоит.

Существует концептуальный провал между наукой, требующей *объективных* измерений и соответствующих выводов, и сознанием, являющимся синонимом *субъективного* опыта. Понятно, что мы не можем задать существу вопрос: «У тебя есть сознание?» Если же мы заглянем в его «голову» (биологическую или какую-то другую), нам нужно будет принять философское решение: что же именно мы там ищем? Таким образом, вопрос о наличии сознания у того или иного существа не является научным. По этой причине некоторые наблюдатели переходят к вопросу, есть ли вообще у сознания какое-либо реальное основание? Английский писатель и философ Сьюзен Блэкмор говорит о «великой иллюзии сознания». Она признаёт существование мемы (идеи) сознания. Другими словами, сознание, безусловно,

---

\* Шершень в русской телевизионной версии.

существует как идея, и многие структуры новой коры имеют дело с этой идеей, не говоря уже о том, сколько об этой идее сказано и написано. Но остается неясным, связана ли эта идея с чем-то реальным. Цель Блэкмор состоит не в том, чтобы опровергнуть существование сознания, а скорее в том, чтобы сформулировать проблемы, с которыми мы сталкиваемся при попытках определить суть этого понятия. Британский психолог и писатель Стюарт Сазерленд в «Международном словаре по психологии» пишет следующее: «Сознание — удивительный, но неуловимый феномен; невозможно сказать, что это такое, в чем его функция или почему оно появилось»<sup>4</sup>.

Однако не стоит воспринимать этот вопрос как пустую философскую болтовню, которая, кстати сказать, началась две тысячи лет назад — с «Диалогов» Платона. Идея сознания лежит в основе нашей морали, а наше законодательство тесно связано с нашими моральными нормами. Если один человек лишает другого человека сознания (как при убийстве), мы считаем это действие аморальным и, за редкими исключениями, преступным. Однако исключения эти тоже связаны с сознанием. В частности, мы разрешаем полицейским или военным убивать сознательных людей, чтобы защитить превосходящих по численности других сознательных людей. Возможность различных исключений можно обсуждать, но суть от этого не меняется.

Насилие над человеком и причинение страданий также обычно рассматриваются как аморальное и незаконное поведение. Если я разрушу то, что принадлежит мне, возможно, это допустимо. Если я без вашего разрешения разрушу то, что принадлежит вам, это, скорее всего, недопустимо, но не потому, что пострадало ваше имущество, а потому, что пострадали вы как владелец. Но если в моей собственности находится сознательное существо, например животное, я как собственник этого животного не имею морального или юридического права делать с ним все, что захочу. В частности, существуют законы, запрещающие жестокое обращение с животными.

Поскольку наша законодательная система и моральные нормы основаны на защите сознательных существ и предотвращении их страданий, для принятия ответственных решений мы должны знать ответ на вопрос, кто же обладает сознанием. Потому этот вопрос — не просто тема для интеллектуальной беседы, что видно, например, из горячей дискуссии по поводу правомерности аборт. Замечу, что вопрос об абортах выходит за рамки темы сознания, поскольку противники аборт утверждаю, что сам тот факт, что эмбрион в итоге станет сознательной личностью, является достаточным основанием для его защиты (то же относится и к людям, находящимся в коме). Поэтому вопрос заключается в том, в какой момент зародыш приобретает сознание.

Понимание сути сознания часто влияет на нашу реакцию в сложных вопросах. Если рассмотреть ту же тему аборт, мы заметим, что многие люди видят различие между приемом противозачаточных таблеток, предотвращающих имплантацию оплодотворенной яйцеклетки в матке в первые дни беременности, и аборт на более поздних сроках. Это различие связано с тем, что, с некоторой вероятностью, эмбрион на более поздних сроках беременности уже обладает сознанием. Трудно предпологать, что эмбрион в возрасте нескольких дней сознателен, если только не придерживаться позиции панпротопсихизма, но даже при такой точке зрения уровень сознания подобного эмбриона ниже, чем у самых примитивных животных. Аналогичным образом, мы совершенно по-разному относимся к проблеме жестокого обращения с высшими приматами по сравнению, например, с насекомыми. Сегодня мало кто беспокоится по поводу страданий, причиняемых нами компьютерным программам (зато мы часто жалуемся на муки, которые компьютерные программы доставляют нам), но если в будущем компьютерное обеспечение получит интеллектуальные, эмоциональные и моральные качества человека, на этом месте возникнет проблема.

Таким образом, моя позиция следующая: я считаю, что небιологические существа с абсолютно убедительными

эмоциональными реакциями обладают сознанием, и предсказываю, что в будущем общество в целом сочтет их сознательными. Обратите внимание, что данное определение описывает иной круг существ, чем те, которые владеют человеческой речью и способны пройти тест Тьюринга. Эти существа должны в достаточной степени быть похожими на человека, так что они подходят под мое определение, и общество, я думаю, со мной согласится. Но под мое определение попадают и те существа, которые проявляют человеческие эмоциональные реакции, но не способны пройти тест Тьюринга, например маленькие дети.

Решает ли это философский вопрос о том, кто обладает сознанием? Хотя бы для меня самого и тех, кто согласен вместе со мной сделать этот смелый шаг? Ответ: *не совсем*. Мы рассмотрели лишь один частный случай, касающийся существ, обладающих человеческими реакциями. Даже если эти существа из будущего, которых мы обсуждаем, имеют небиологическую природу, мы рассматриваем их способность убедительно демонстрировать человеческие реакции, то есть данная позиция по-прежнему является человекоцентрической. Но что можно сказать о других формах разума, отличных от человеческого? Мы можем представить себе существование столь же сложного или гораздо более сложного разума, чем наш, у существ с совершенно иными эмоциями и мотивациями. Как понять, обладают они сознанием или нет?

Можно начать с обсуждения биологических существ, мозг которых сравним с мозгом человека, но поведение совсем иное. Британский философ Дэвид Кокберн (род. в 1949 г.) описывает поведение гигантского кальмара в момент опасности (возможно, как считает Кокберн, кальмар испугался человека с видеокамерой). Кальмар задрожал и съежился и, как пишет Кокберн, «реагировал так, что мне стало очевидно, что ему страшно. Потрясающим в этой последовательности реакций было то, что в поведении существа столь далекого от человека, с очевидностью отразилась человеческая эмоция страха»<sup>5</sup>. Кокберн убежден, что животное испытывало именно эту эмоцию и что большинство

людей, посмотревших фильм, придет к такому же мнению. Если мы согласимся с его точкой зрения, мы вынуждены будем включить гигантских кальмаров в наш список сознательных существ. Однако это никуда нас не ведет, поскольку такой вывод по-прежнему основан на нашей способности сопереживания эмоций, которые знакомы нам самим. Мы по-прежнему смотрим на вещи с человекоцентрической точки зрения.

Если же покинуть мир биологических существ, мы обнаружим, что формы небιологического разума могут оказаться еще более разнообразными. Например, какие-то существа могут не бояться собственной гибели и не иметь нужды в обычных для людей и других биологических существ эмоциях. Может быть, они все же будут способны пройти тест Тьюринга, а может быть, попросту не захотят.

Нам нужны роботы, предназначенные для работы в опасных условиях и не имеющие инстинкта самосохранения. Пока такие машины не являются достаточно разумными или сложными, чтобы всерьез говорить об их чувствительности, но можно вообразить, что в будущем они станут столь же сложными, как люди. Что тогда?

Лично я могу сказать, что поломка устройства с таким сложным строением и способностью выполнять указания, необходимые для реализации его миссии, меня бы, безусловно, поразила и огорчила. Возможно, здесь мы слегка расширяем обзор, поскольку рассматриваем реакцию на поведение, отличное от нормального поведения людей и вообще всех биологических существ. Но и в данном случае я пытаюсь нащупать связь с теми признаками, которые характерны для меня самого и для других людей. Вообще говоря, мысль о существе, полностью посвящающем себя выполнению достойной задачи и пытающемся решить ее вне зависимости от собственного благополучия, не так уж чужда человеческому опыту. Нам нужны существа, способные защитить биологических людей или каким-то образом способствовать нашему процветанию.

Но что если такое существо имеет собственные задачи, отличные от человеческих, и не выполняет той миссии, которую мы считаем достойной? Можно попытаться оценить его способности другим путем. Если это существо действительно очень разумное, возможно, оно сильно в математике, и с ним можно поговорить на эту тему. Возможно, оно сможет оценить математические шутки.

Но если это существо не захочет общаться со мной, а у меня не будет достаточной возможности оценить его действия и решения, означает ли это, что оно лишено сознания? Я должен сделать вывод, что существа, которые не убедили меня своими эмоциональными реакциями или не пытались этого сделать, не обязательно являются несознательными. Сложно распознать существо с другим сознанием, не установив какого-либо эмпатического контакта, но это скорее проблема моей собственной ограниченности, чем проблема того существа. Поэтому действовать придется осторожно. Нам бывает сложно поставить себя на место другого человека, что уж говорить о понимании разума, который так сильно отличается от нашего собственного.

### **Что же мы осознаём?**

Если бы мы могли заглянуть сквозь череп внутрь мозга сознательно мыслящего человека и если бы место оптимального возбуждения было светящимся, то мы должны бы увидеть на мозговой поверхности игру яркого пятна с фантастическими волнующимися границами, постоянно флюктуирующее в размере и форме, окруженное более или менее глубокой темнотой, покрывающей остальную часть полушария мозга.

*Иван Петрович Павлов, 1913<sup>6</sup>*

Вернемся к истории с гигантским кальмаром. Да, возможно, мы распознаём некоторые из его эмоций, но значительная часть его поведенческих реакций остается для нас тайной. Что значит

быть гигантским кальмаром? Что он чувствует, когда проталкивает свое бесхребетное тело через небольшое отверстие? У нас просто слов таких нет, чтобы ответить на этот вопрос, учитывая нашу неспособность даже описать опыт, разделяемый многими людьми, такой как восприятие красного цвета или ощущение нашего тела, погружаемого в воду.

Но, чтобы отыскать тайну осознанного опыта, не нужно опускаться на дно океана, достаточно обратиться к самим себе. Я, например, знаю, что обладаю сознанием. Полагаю, что и ты, мой читатель, тоже. (Чего не могу сказать о людях, которые мою книгу не купили.) Но *что* же, собственно, я сознаю? Попробуйте и вы задать себе этот вопрос.

Давайте проведем один мысленный эксперимент (он годится для тех, кто водит машину). Представьте, что вы на скоростном шоссе и движетесь по левой полосе. Теперь закройте глаза, возьмитесь за воображаемый руль и сделайте такое движение, как будто перестраиваетесь на правую полосу.

Попробуйте проделать этот эксперимент, прежде чем читать дальше.

Вот вероятная последовательность ваших движений. Вы держите руль, проверяете, свободна ли правая полоса. Предположим, полоса свободна, тогда на короткое время вы поворачиваете руль вправо, а затем выправляете его. Дело сделано.

Это хорошо, что вы в воображаемой машине, поскольку только что вы пересекли все полосы шоссе и врезались в дерево. Наверное, следует предупредить, чтобы вы не пытались проделать этот трюк, двигаясь в настоящей машине (надеюсь, вы уже усвоили правило, что за рулем не следует закрывать глаза), но дело, в сущности, не в этом. Если вы поступили так, как я только что описал, — а практически каждый человек в мысленном эксперименте поступает именно так, — вы поступили неправильно. Если вы повернете руль вправо, а затем выровняете колеса, вы будете двигаться по диагонали к вашему исходному направлению движения. Вы свернете на правую полосу, как и хотели,

но будете и дальше ехать вправо. После того как вы выехали на полосу справа от вас, нужно было повернуть руль влево и *только потом* выровнять колеса. Так вы заставите машину ехать вперед по новой полосе.

Если вы опытный водитель, вы проделывали этот маневр тысячи раз. Действовали ли вы при этом сознательно? Вы когда-нибудь обращали внимание на то, что на самом деле делаете, когда меняете полосу на шоссе? Надеюсь, вы читаете книгу не в больнице, поправляясь после автомобильной аварии, и, следовательно, умеете выполнять этот маневр. Но вы никогда не задумывались о том, что делаете, хотя делали это множество раз.

Когда люди рассказывают, что с ними происходило, они излагают последовательность событий и решений. Но в тот момент, когда мы впервые переживаем эти события, все воспринимается не так. Наш первичный опыт представляет собой последовательность образов высокого порядка сложности, и некоторые из них могут вызывать эмоции. Но мы запоминаем лишь незначительную часть этих образов, если вообще запоминаем. Даже если мы передаем нашу историю очень точно, мы прибегаем к конфабуляции, позволяющей восполнять недостающие подробности и превращать последовательность образов в связный рассказ. Мы не можем быть уверены в том, что наше исходное осознание опыта было именно таким, поскольку доступ к этому опыту проходит через воспоминания. Каждый текущий момент мимолетен, и события этого момента сразу становятся воспоминаниями, а чаще не становятся. Даже если опыт стал воспоминанием, он сохраняется, как указывает теория распознавания образов, в виде иерархической системы образов разной степени сложности. И как я уже неоднократно отмечал, практически весь наш опыт (в том числе опыт смены полосы на шоссе) мгновенно забывается. Таким образом, мы не можем сказать, что именно заключает в себе наш осознанный опыт.

## Восток — это Восток, Запад — это Запад

До появления мозга во Вселенной не было ни цвета, ни звуков, не было никаких запахов и, вероятно, мало чувств, ощущений или эмоций.

*Роджер У. Сперри\**<sup>7</sup>

Рене Декарт заходит в ресторан и садится ужинать. К нему подходит официант и спрашивает, желает ли тот закусить.

— Нет, спасибо, — отвечает Декарт, — хочу сразу заказать ужин.

— Вас интересует специальное меню? — спрашивает официант.

— Нет, — отвечает Декарт, теряя терпение.

— Хотите ли выпить перед ужином? — спрашивает официант.

Декарт оскорблен, поскольку он трезвенник.

— Не думаю! — отвечает он с негодованием и — ПУФ! — исчезает.

*Анекдот, рассказанный Дэвидом Чалмерсом*

На анализируемые нами вопросы о сознании и реальности можно взглянуть с двух точек зрения: точки зрения западного человека и восточного человека. Если придерживаться точки зрения западного человека, мы начнем рассмотрение с физического мира, в котором рождается информация. В результате миллиарда лет эволюции живые существа стали настолько сложными, что обзавелись сознанием. С точки зрения восточного человека, сознание — фундаментальная реальность, а физический мир существует лишь по той причине, что сознательные существа думают о нем. Другими словами, физический мир — это материализованные мысли сознательных существ. Безусловно, я упрощаю сложные и разнообразные философские течения, но эти две точки зрения составляют полюса философских взглядов на проблему сознания и его отношения к физическому миру.

Раздел между Востоком и Западом по вопросу о сознании нашел свое отражение и в направлениях мысли в области физики

---

\* Роджер Уолкотт Сперри (1913–1994) — американский нейропсихолог, лауреат Нобелевской премии по физиологии и медицине «за открытия, касающиеся функциональной специализации полушарий головного мозга».

элементарных частиц. В квантовой механике положение частиц описывается так называемыми полями вероятности. Но любые попытки измерения состояния этих частиц с помощью приборов приводят к так называемому коллапсу волновой функции, что означает, что частица внезапно занимает фиксированное положение. Причем эти измерения осуществляются сознательным наблюдателем, поскольку иначе в них не было бы никакого смысла. Таким образом, частица приобретает определенные свойства (локализация, скорость) только тогда, когда за ней наблюдают. Можно сказать, что, пока за частицами никто не следит, им все равно, где находиться. Я называю такой взгляд буддистской школой квантовой механики, поскольку при этом подходе подразумевается, что частицы существуют только тогда, когда за ними наблюдает существо, обладающее сознанием.

Существует другая интерпретация квантовых свойств частиц, избегающая подобной антропоморфной терминологии. В данном случае соответствующее частице поле — это не поле вероятности, а скорее некая функция, которая в разных местах принимает разные значения. Следовательно, частица — это и есть поле. На параметры поля в разных участках накладываются определенные ограничения. В этом и заключается смысл слова «квантовый». В рамках данного подхода так называемый коллапс волновой функции — вовсе никакой не коллапс. Волновая функция никуда не девается. Просто дело в том, что измерительный прибор тоже состоит из частиц и их полей и взаимодействие поля анализируемой частицы с полями частиц прибора приводит к тому, что частица регистрируется в определенном месте. Но поле никуда не делось. Такова «западная» интерпретация квантовой механики, хотя следует заметить, что среди физиков всего мира более популярной является «восточная» интерпретация.

Это разделение «Восток—Запад» ярко проявилось в работах одного философа. Британско-австрийский философ Людвиг Витгенштейн (1889–1951) занимался философией языка

и знания и изучал вопрос о том, что составляет наше реальное знание. Эту проблему он анализировал, находясь на полях сражений Первой мировой войны, и изложил свои мысли в «Логико-философском трактате» — единственной его книге, опубликованной при жизни. Книга имела необычную структуру и вышла в свет только в 1921 г. благодаря усилиям бывшего наставника Витгенштейна, британского математика и философа Бертрана Рассела (1872–1970). Труд этот стал Библией для известного философского направления, называемого «логическим позитивизмом» и занимающегося определением границ научного познания. Книга и возникавшие вокруг нее философские дискуссии повлияли на труды Тьюринга и способствовали возникновению информатики и лингвистики.

«Логико-философский трактат» предвосхищает представление о том, что все знания имеют иерархическое строение. Сама книга организована в виде вложенных друг в друга пронумерованных утверждений. Вот, к примеру, четыре первые фразы.

- 1 Мир есть все то, что имеет место.
- 1.1 Мир есть совокупность фактов, а не вещей.
- 1.11 Мир определен фактами и тем, что это *все* факты.
- 1.12 Потому что совокупность всех фактов определяет как все то, что имеет место, так и все то, что не имеет места\*.

Еще одно важное утверждение «Трактата», отразившееся позднее в работах Тьюринга, имеет номер 4.0.0.3.1 и гласит: «Вся философия — это критика языка».

И сам «Трактат», и направление логического позитивизма утверждают, что физическая реальность существует вне зависимости от нашего восприятия, но все, что мы можем узнать об этой реальности, определяется возможностями наших органов чувств (даже усиленных с помощью инструментов) и логических

---

\* Витгенштейн Л. Логико-философский трактат / Пер. с нем. М. : Наука, 1958 (2009).

построений, основанных на этом чувственном опыте. Можно сказать, что Витгенштейн пытался описать методы и задачи науки. Заключительное, седьмое, утверждение его книги гласит: «О чем невозможно говорить, о том следует молчать». Таким образом, в своих ранних работах Витгенштейн рассматривает дискуссию на тему сознания как бесконечную и тавтологичную, а следовательно, лишённую смысла.

Однако в более поздних работах Витгенштейн изменил точку зрения и посвятил все свое время изучению проблем, которые прежде относил к разряду тех, что обходят молчанием. Его труды на эту тему были собраны и напечатаны в 1953 г., через два года после его смерти, в виде книги под названием «Философские исследования». Он критиковал свои ранние суждения, объявив их бесконечными и бессмысленными, и пришел к выводу, что изучения достойно все то, о чем он прежде не считал нужным говорить. Его поздние труды оказали глубокое влияние на экзистенциалистов, так что Витгенштейн является уникальной фигурой в современной философии, поскольку он сформировал две лидирующие и противоречащие друг другу философские школы.

О чем же «поздний» Витгенштейн считал нужным думать и говорить? О таких понятиях, как красота и любовь, которые, как он признал, наличествуют в виде идей в голове человека. Однако он считал, что они должны существовать в совершенном, идеальном мире и аналогичны идеальным «формам», о которых писал Платон в своих «Диалогах» (еще одна работа, проложившая путь, казалось бы, взаимоисключающим философским направлениям).

Мне представляется, что позиция французского философа и математика Рене Декарта охарактеризована неправильно. Его знаменитое изречение «Я мыслю, значит, я существую» обычно истолковывают как восхваление рационального мышления, в том смысле, что «я мыслю, то есть могу совершать логические построения, следовательно, я на что-то годен». Декарта,

таким образом, считают основателем западной рациональной философии.

Однако, обдумывая высказывание Декарта в свете других его трудов, я прихожу к иному выводу. Декарта волновал вопрос «психического и телесного», а именно: как из физического вещества мозга возникает сознание? Учитывая это, мне кажется, он пытался довести рациональный скептицизм до экстремума, и на этом основании его утверждение можно перефразировать так: «Я мыслю, значит, существует мой субъективный опыт, следовательно, наверняка мы знаем только то, что нечто — назовем это “Я” — существует». Он не был уверен в том, что физический мир существует, поскольку вся информация, каковой мы обладаем, составлена из нашего личного восприятия, которое может быть ошибочным или полностью иллюзорным. Однако мы точно знаем, что есть существо, владеющее этим опытом.

Я получил религиозное образование в унитарной церкви, где мы изучали все мировые религии. Мы, к примеру, шесть месяцев изучали буддизм, ходили на службы в буддистские храмы, читали их книги и беседовали с их религиозными лидерами. Потом переключались на другую религию, скажем, иудаизм. Главная идея состояла в том, что «к истине есть много путей», и так мы получили представление о толерантности и трансцендентности. Последняя идея подразумевает, что для разрешения кажущихся противоречий между различными традициями не требуется признавать, что одна из них верна, а другая ошибочна. Истину можно найти только путем объяснения, которое выходит за пределы этих кажущихся различий, особенно в таких фундаментальных вопросах, как смысл и цель.

И именно этот подход я бы применил к разрешению вопроса о справедливости «восточного» и «западного» понимания связи между сознанием и физическим миром. На мой взгляд, оба направления справедливы.

С одной стороны, глупо отрицать существование физического мира. Даже если мы живем в имитированном мире, как

утверждает шведский философ Ник Бостром (род. в 1973 г.), мы воспринимаем его как реальность. Если же мы допускаем существование физического мира и происходящую в нем эволюцию, мы видим, что именно в результате эволюции и появились сознательные существа.

С другой стороны, утверждения «восточной» перспективы (основа всего сознание, которое и является единственной важной реальностью) тоже трудно отрицать. Хотя бы сравните, как мы оцениваем сознательных и несознательных существ. Мы считаем, что последние не имеют собственной ценности, за исключением той, которую приписывают им сознательные существа. И даже если мы воспринимаем сознание как новое свойство сложных систем, мы не можем утверждать, что это просто еще один их атрибут (наряду с «пищеварением и лактацией», если цитировать Джона Серля). Это нечто поистине важное.

Для описания чрезвычайно важных вещей часто используют слово «духовный». Многим данная терминология, имеющая религиозные традиции, не нравится, поскольку подразумевает комплекс убеждений, с которыми эти люди не согласны. Но на самом деле само слово «дух» часто используют в качестве эквивалента слова «сознание».

Таким образом, эволюцию можно рассматривать как духовный процесс, в том смысле, что она создает духовность, то есть носителей сознания. Кроме того, эволюция ведет к усложнению, к накоплению знаний, к усилению разума, красоты, творчества и к способности выражать трансцендентные эмоции, такие как любовь. Заметим, что все эти определения люди используют для описания Бога.

Людям часто не нравятся дискуссии по поводу возможности существования сознания у машин, поскольку для них это означает понижение духовной ценности человека. Однако такая реакция отражает неправильное восприятие понятия «машина». Критики строят выводы на своих представлениях о современных машинах, но, как бы ни были сложны эти устройства, они

пока не могут восприниматься в качестве сознательных существ. Однако, по моим прогнозам, они станут неотличимы от живого человека, которого мы считаем сознательным существом, и, следовательно, будут разделять все те духовные ценности, что мы связываем с сознанием. Это не унижение достоинства человека, а скорее возвышение нашей оценки (некоторых) машин будущего. Возможно, для этих существ понадобится выбрать другую терминологию, поскольку это будут совсем другие машины.

По мере того как мы учимся заглядывать внутрь мозга и расшифровывать его механизмы, мы открываем для себя его методы и алгоритмы. И мы не только понимаем их, но и воспроизводим эти «толкающие друг друга части мельницы», выражаясь словами немецкого математика и философа Вильгельма Лейбница (1646–1716). Более того, мы настолько сблизимся с создаваемыми нами инструментами, что различие между человеком и машиной начнет размываться, а потом и вовсе исчезнет. Этот процесс уже идет, хотя большинство машин, расширяющих наши возможности, пока еще находятся вне нашего тела и мозга.

### **Свобода воли**

Важнейшим аспектом сознания является умение смотреть вперед, способность «прогнозирования». Эта способность подразумевает возможность планировать, а в социальном плане — составлять сценарий того, что должно или может произойти, но пока еще не произошло... С помощью этой системы мы повышаем свои шансы совершить такие действия, которые в наибольшей степени будут соответствовать нашим собственным интересам... Я полагаю, что «свобода воли» складывается из нашей кажущейся возможности выбирать и действовать таким образом, который кажется наиболее полезным или подходящим, и нашей упорной уверенности в том, что это наш собственный выбор.

*Ричард Д. Александер\**

---

\* Ричард Александер (род. в 1929) — американский эколог и эволюционист, почетный профессор Зоологического музея Университета Мичигана.

## Мысленные эксперименты по анализу разума

Нужно ли говорить, что растение не знает, что делает, просто по той причине, что у него нет глаз, ушей или мозга? Если мы скажем, что оно действует механически и только механически, не придется ли признать, что множество других и совсем неслучайных действий тоже являются механическими? Нам кажется, что растение механически убивает и съедает муху, а не может ли показаться растению, что человек механически убивает и съедает овцу?

*Сэмюэл Батлер*

У вас в головах есть мозги.

У вас в ботинках есть ноги.

И вы можете направить себя

Туда, куда пожелаете.

*Dr. Seuss, композиция Oh, the places you'll go!*

Можно ли сказать, что головной мозг, который имеет парную структуру, двойной орган, «кажется разделенным на части, но един в своем разделении»?

*Г. Модсли\*, 1889<sup>8</sup>*

Избыточность, как мы уже усвоили, является ключевой стратегией новой коры. Но в головном мозге есть и другой уровень избыточности, а именно наличие левого и правого полушария, которые хотя и не идентичны, но во многом очень похожи. Как специфические отделы новой коры специализируются на обработке определенного типа информации, так и полушария имеют некоторую специализацию, например левое полушарие обычно отвечает за речевую функцию. Однако такое разделение может быть изменено, так что при работе лишь половины мозга мы можем выжить и более или менее нормально существовать. Американские нейропсихологи Стелла де Боуд и С. Кертис описали 49 детей, переживших гемисферэктомию (операцию по удалению половины головного мозга). Этот вид хирургического вмешательства применяется

---

\* Генри Модсли (1835–1918) — английский психиатр и философ-позитивист, один из основоположников детской психиатрии и эволюционного направления в психиатрии.

для лечения больных с угрожающими жизни нарушениями, затрагивающими лишь одно полушарие мозга. У некоторых больных после этой операции возникают определенные дефекты, но они являются специфическими, и больные фактически сохраняют личностные качества. Многие чувствуют себя совершенно нормально, и по внешнему виду и поведению невозможно сказать, что у них осталась только половина мозга. Де Боуд и Кертис пишут, что у детей, перенесших гемисферэктомию, «развивается удивительно хорошая речь, несмотря на удаление “речевого” полушария»<sup>9</sup>. Они описывают одного ребенка, который после этой операции окончил университет, поступил в аспирантуру и имел показатель IQ выше среднего. Исследования показывают наличие минимальных долгосрочных последствий для обучения, памяти, личностного восприятия и чувства юмора<sup>10</sup>. В 2007 г. американские исследователи Шервурд Макклеланд и Роберт Максвелл опубликовали аналогичные данные для взрослых, продемонстрировав долгосрочные положительные результаты<sup>11</sup>.

Описан случай десятилетней немецкой девочки, которая родилась всего с одной половиной мозга, но при этом чувствует себя практически нормально. Поле зрения одного глаза у нее практически идеальное, хотя у больных, перенесших гемисферэктомию, сразу после операции ширина поля зрения сокращается<sup>12</sup>. Шотландский исследователь Ларс Маккли комментирует: «Мозг обладает замечательной пластичностью, но все-таки мы были поражены, когда увидели, насколько хорошо единственное полушарие мозга этой девочки адаптировалось, чтобы компенсировать недостающее полушарие».

Хотя эти наблюдения, безусловно, подтверждают идею пластичности новой коры, интереснее следующий вывод: каждый из нас, по-видимому, имеет не один мозг, а два и может достаточно хорошо жить с любым из них. Если мы теряем один мозг, мы теряем и хранившиеся в его коре образы, но каждая часть мозга вполне самодостаточна. Имеет ли каждая полусфера

собственное сознание? Некоторые факты позволяют ответить на этот вопрос утвердительно.

Рассмотрим пример больных с расщепленным мозгом, у которых сохранены оба полушария, но между ними образовалась расщелина. Мозолистое тело мозга (*corpus callosum*) представляет собой пучок из 250 млн аксонов, соединяющих между собой полушария мозга и позволяющих им общаться и координировать свои функции. Как два человека могут находиться в тесном контакте и принимать общие решения, при этом оставаясь отдельными и цельными личностями, так и два полушария мозга способны функционировать как единое целое, оставаясь независимыми.

Как следует из названия патологии, у пациентов с расщепленным мозгом происходит расщепление или повреждение мозолистого тела, в результате чего у таких людей остаются два функциональных мозга, не имеющих возможности непосредственного общения. Американский психолог Майкл Газзанига (род. в 1939 г.) проводил обширные исследования мыслительного процесса в полусферах мозга таких пациентов.

Левое полушарие пациентов с расщепленным мозгом обычно воспринимает информацию с правого поля зрения и наоборот. Газзанига и его коллеги помещали в поле зрения правого глаза пациента (за которое отвечает левое полушарие) изображение куриного когтя, а в поле зрения левого глаза — зимнюю картинку. Затем пациенту показывали набор изображений, так что они воспринимались обоими глазами, и просили его выбрать одну из картинок, которая соответствовала первым изображениям. Левая рука пациента (контролируемая правым полушарием) указывала на изображение совка, а правая — на изображение цыпленка. До сих пор все нормально — оба полушария работали правильно и независимо. Но на вопрос «Почему вы выбрали это изображение?» пациент отвечал (речевой центр находится в левом полушарии) следующее: «Очевидно, что куриный коготь соответствует курице». Но когда пациент опускал голову и видел,

что его левая рука указывает на совок, немедленно объяснял и это (опять-таки, с помощью речевого центра левого полушария): «И вам требуется совок, чтобы чистить птичью клетку».

Это пример конфабуляции. Правое полушарие (контролирующее левую руку и плечо) правильно показывает на совок, но, так как левое полушарие (контролирующее вербальный ответ) ничего про снег не знает, оно придумывает объяснение, хотя и не отдает себе в этом отчета. Оно берет на себя ответственность за действие, которого не совершало, но думает, что совершало.

Этот опыт демонстрирует, что каждое полушарие пациента с расщепленным мозгом имеет собственное сознание. Повидимому, полушария не отдают себе отчета в том, что тело контролируется двумя частями мозга, поскольку научились координировать свои функции, и их решения достаточно согласованны, так что каждое думает, что все решения приняты им самим.

Эксперименты Газзаниги не доказывают, что нормальные люди с сохраненной функцией мозолистого тела имеют два сознательных полушария, но указывают на такую возможность. Хотя мозолистое тело способствует эффективному взаимодействию двух полушарий, это не означает, что полушария не обладают независимым разумом. Каждое может считать, что именно оно принимает все решения, поскольку эти решения вполне соответствуют общей логике, и в конце концов каждое полушарие влияет-таки на каждое решение (путем сотрудничества со вторым полушарием при посредничестве мозолистого тела). Таким образом, каждому полушарию кажется, что это оно контролирует все процессы.

Как бы вы проверили эту гипотезу? Можно, например, провести корреляцию между неврологической активностью и сознанием. Именно это и сделал Газзанига. Его эксперименты показывают, что каждое полушарие функционирует как независимый мозг. Конфабуляции не ограничены полушариями, и для каждого из нас это регулярная практика. Каждое полушарие

почти так же разумно, как человек в целом, поэтому, если мы считаем, что человек обладает сознанием, приходится сделать вывод, что каждое полушарие обладает независимым сознанием. Мы можем провести неврологические корреляции и проделать собственные мысленные эксперименты (например, решить для себя вопрос, что если два полушария мозга, не сообщаясь через мозолистое тело, формируют два независимых сознательных разума, то такая же ситуация должна иметь место при наличии функционирующего мозолистого тела), но любые попытки прямого детектирования сознания в каждом полушарии невозможны из-за отсутствия научного метода определения сознания. Но если мы допускаем, что каждое полушарие мозга сознательно, не придется ли признать, что так называемая неосознанная активность новой коры (составляющая ее основную функцию) тоже обладает независимым сознанием? А может быть, несколькими? Недаром Марвин Минский называл мозг «обществом разума»<sup>13</sup>.

В другом исследовании пациенту с расщепленным мозгом показывали написанное на бумаге слово «колокольчик» для восприятия правым полушарием и слово «музыка» для восприятия левым полушарием. Контролирующее речь левое полушарие реагировало словом «музыка». Затем пациенту показывали набор картинок и просили показать ту, которая в наибольшей степени соответствует слову, которое ему только что показали. Его рука, контролируемая правым полушарием, указывала на колокольчик, а контролируемый левым полушарием речевой центр сообщал: «Да, музыка, последний раз, когда я слышал музыку, это были колокольчики, звук доносился снаружи». Он выбрал это объяснение, хотя среди предложенных ему картинок были и другие, даже более близкие по смыслу к слову «музыка».

И опять мы сталкиваемся с конфабуляцией. Левое полушарие объясняет решение, которого оно никогда не принимало, как будто это было его решение. И делает это не для того, чтобы «прикрыть друга» (то есть объяснить выбор второго

полушария). Оно действительно считает, что само приняло это решение.

Эти реакции и решения могут распространяться и на эмоциональные реакции. Подростка с расщепленным мозгом спросили (так, что слышали оба полушария): «Кто твой любимый...?» — а затем правому полушарию (через левое ухо) сообщили слово «подружка». Газзанига пишет, что юноша покраснел и засмутился, что является обычной реакцией подростков, когда их спрашивают о любимой девушке. Однако левое полушарие, контролирующее речевой центр, ничего не слышало и попросило уточнить: «Мой любимый что?» Когда же пациента попросили ответить на вопрос в письменном виде, левая рука, контролируемая правым полушарием, написала имя девушки.

Эксперименты Газзаниги не являются мысленными — это реальные эксперименты над разумом. Они передают интересные наблюдения над сознанием, но имеют даже более непосредственное отношение к вопросу о свободе воли. В каждом из описанных случаев одно из полушарий считает, что это оно приняло решение, хотя на самом деле это не так. В какой степени это относится к решениям, которые мы принимаем каждый день?

Вот случай десятилетней девочки, страдающей от эпилепсии. Нейрохирург Ицхак Фрид выполнил операцию на мозге пациентки без наркоза (это возможно по той причине, что в головном мозге нет болевых рецепторов)<sup>14</sup>. Всякий раз, когда хирург стимулировал определенный участок новой коры, девочка смеялась. Сначала врачи подумали, что они, возможно, каким-то образом запускают смеховой рефлекс, но потом поняли, что затрагивают механизм восприятия юмора. По-видимому, они нашли в ее новой коре точку (на самом деле их несколько), распознающую смешное. Она не просто смеялась, она действительно находила ситуацию смешной, хотя вокруг нее ничего не изменилось — врачи просто затронули ту самую точку в ее новой коре. Когда ее спросили, почему она так смеется, она не ответила

что-то вроде «так, без причины» или «наверное, вы заставляете смеяться мой мозг», но тут же придумала ответ. Она указывала на какой-то предмет в комнате и пыталась объяснить, почему он вызывает у нее смех. Типичный комментарий был таким: «Вы тоже так весело здесь смотрите».

Нам все время очень хочется найти рациональное объяснение нашим действиям, даже если нам на самом деле не пришлось принимать решения, которые их вызвали. Так в какой же степени мы ответственны за наши решения? Приведу в качестве примера эксперименты профессора психологии Бенджамина Лайбета из Калифорнийского университета в Дэвисе. Нескольким участникам эксперимента прикрепили к голове электроды для снятия электроэнцефалограммы (ЭЭГ) и поставили таймер. Лайбет предлагал испытуемым выполнять простые задания, такие как нажать на кнопку или подвигать пальцем. При этом их просили отмечать по таймеру время, когда они «в первый момент осознают, что должны действовать». Тесты показали, что задержка составляла лишь 50 миллисекунд. Между моментом, когда участник эксперимента сообщал о готовности действовать, и реально приступал к действию, проходило в среднем 200 миллисекунд<sup>15</sup>.

Исследователи также анализировали сигналы ЭЭГ. Активность мозга, связанная с началом функции моторной коры (ответственной за выполнение движения), проявлялась в среднем примерно за 500 миллисекунд до выполнения действия. Это означает, что моторная кора начинала готовиться к выполнению задачи примерно за треть секунды до того, как человек понимал, что решение принято.

Смысл экспериментов Лайбета горячо обсуждался в научных кругах. Сам Лайбет пришел к выводу, что наша идея о принятии решения, по-видимому, является иллюзией и что «сознание в этом не участвует». Философ Дэниел Деннет так прокомментировал эти эксперименты: «Действие сначала готовится в какой-то части мозга, и оттуда сигналы направляются

к мускулам, останавливаясь по дороге, чтобы сообщить вам, сознательному существу, что происходит (но, как хорошие чиновники в кабинете президента, поддерживают иллюзию, что все затеяли именно вы)»<sup>16</sup>. В то же время Деннет сомневается в значимости временных интервалов, определенных в экспериментах Лайбета. В частности, он считает, что люди не могут отдавать себе отчета в том, когда именно начинают осознавать, что решение принято. Можно задать вопрос: если сам человек не знает, в какой момент он осознаёт, что решение принято, то кто же это знает? Да, вопрос поставлен правильно — как я и говорил выше, мы толком не знаем, что же именно мы осознаём.

Индийско-американский нейробиолог Вилейанур Субраманиан Рамачандран (род. в 1951 г.) объясняет ситуацию несколько иначе. Учитывая, что в новой коре имеется порядка 30 млрд нейронов, тут всегда происходит множество процессов, и мы осознаём лишь малую часть происходящего. Решения — малые и большие — постоянно обрабатываются новой корой, и предлагаемые выводы достигают нашего сознания. Рамачандран считает, что говорить следует не о свободе воли, а о «свободе нежелания», то есть о способности отвергать решения, предлагаемые «несознательной» частью новой коры.

Проведем аналогию с военной кампанией. Военные чиновники готовят рекомендации президенту. Прежде чем получить его резолюцию, они проводят подготовительную работу, которая позволит осуществить принятое решение. В какой-то момент предложение доходит до президента, он его одобряет, и тогда выполняются остальные операции в рамках данной стратегии. Аналогичным образом, поскольку в новой коре происходят неосознанные процессы (подготовка рекомендаций до их представления президенту) и осознанные процессы (функция президента), мы можем наблюдать активность нейронов и реальные действия до того, как принято официальное решение. В каждой конкретной ситуации можно обсуждать, насколько свободен президент в выборе окончательного решения, в частности,

американские президенты на протяжении всей истории страны как принимали, так и отклоняли предложения армейских чиновников. Но нас не должно удивлять, что умственная деятельность, даже в моторной коре, начинается до того, как мы осознаём, что нужно принимать решение.

Таким образом, эксперименты Лайбета показывают, что в нашем мозге при принятии решений происходит большая неосознанная работа. Мы уже знаем, что большая часть процессов в новой коре не является осознанной, поэтому не следует удивляться, что наши действия и решения имеют как осознанную, так и неосознанную составляющие. Важно ли их разделять? Если наши решения складываются из этих двух составляющих, есть ли смысл их идентифицировать? Разве не оба аспекта представляют работу нашего мозга? Являемся ли мы в итоге ответственными за все, что происходит в нашем мозге? Вряд ли можно прикрыться фразой типа: «Да, я убил его, но я не могу нести ответственность, поскольку не обратил на это внимания». Даже если иногда существуют какие-то легальные основания, позволяющие не считать человека ответственным за принятые решения, обычно мы отвечаем за все, что совершаем.

Приведенные мной в качестве примеров эксперименты и наблюдения относятся к вопросу *свободы воли*. Эта тема, как и тема сознания, обсуждается со времен Платона. Даже выражение «свобода воли» возникло еще в XIII столетии, но что же, в сущности, оно означает?

Словарь Мерриам—Уэбстер дает такое определение: «Свобода человека сделать выбор, который не предопределен предысторией или божественным вмешательством». Можно заметить, что определение это безнадежно тавтологично: «свобода — это свобода...» Оставив в стороне идею противопоставления свободы воли и божественного вмешательства, обнаруживаем в этом определении один важный элемент, заключающийся в том, что свободное решение не предопределено предшествующими событиями. Я вернусь к этому вопросу чуть ниже.

«Стэнфордская философская энциклопедия» утверждает, что свобода воли — это «способность разумного существа выбрать направление действий среди возможных альтернатив». В соответствии с этим определением обычный компьютер тоже обладает свободой воли, так что энциклопедия помогает нам еще меньше, чем словарь. От «Википедии» чуть больше пользы. Здесь дается такое определение: «Возможность человека делать выбор вне зависимости от определенных обстоятельств... Наличие ограничивающего обстоятельства есть детерминизм»\*. Здесь опять дается круговое определение, но зато называется главный враг свободы воли — *детерминизм*. В этом смысле приведенное выше определение из словаря Мерриам—Уэбстер является аналогичным: свобода воли противопоставляется выбору, предопределенному предысторией.

Что же такое детерминизм? Если мы набираем на калькуляторе «2 + 2», он выдает результат «4», могу ли я сказать, что калькулятор выразил свободу воли и решил показать «4»? Вряд ли найдется человек, который воспримет это в качестве демонстрации свободы воли, поскольку «решение» предопределено входным сигналом и внутренними механизмами калькулятора. И если я совершу более сложные вычисления, мы по-прежнему скажем, что свободой воли калькулятор не обладает.

А как расценить ответы Ватсона на вопросы «Джеопарди!»? Хотя его рассуждения намного сложнее операций, производимых калькулятором, и в этом случае вряд ли найдутся люди, которые обнаружат в его ответах проявление свободы воли. Никто не знает досконально, как работает эта программа, но несколько исследователей сообща смогут описать все методы, определяющие

---

\* В «Википедии» на русском языке дается следующее определение: «Возможность человека делать выбор вне зависимости от определенных обстоятельств... Существуют две противоположные позиции: метафизический либертарианизм — утверждение, что детерминизм неверен и, таким образом, свобода воли существует или по меньшей мере возможна, и жесткий детерминизм — утверждение о том, что детерминизм верен и свободы воли не существует».

функции Ватсона. Можно сказать, что решения Ватсона определяются следующими факторами: 1) всеми программами, участвующими в поиске ответа на конкретный вопрос; 2) самим вопросом; 3) состоянием внутренних параметров, определяющих выбор решения; 4) триллионами байт информации, включая энциклопедии. И на основании этих четырех категорий данных ответ Ватсона предопределен. Может быть, на один и тот же вопрос всегда будет выдан один и тот же ответ, однако Ватсон запрограммирован таким образом, чтобы учиться на собственном опыте, поэтому существует возможность, что в следующий раз ответ будет несколько иным. Однако это не противоречит нашему выводу, а скорее подразумевает расширение пункта 3, описывающего контролируемые параметры машины.

Так в чем же именно состоит различие между человеком и Ватсоном, объясняющее наличие свободы воли у человека и ее отсутствие у машины? Можно назвать несколько факторов. Даже с учетом того, что Ватсон лучше играет в «Джеопарди!», чем большинство людей, его строение все же гораздо проще строения человеческой новой коры. Ватсон действительно много знает и использует иерархические методы, но структура его иерархического мышления пока еще значительно проще, чем у человека. Заключается ли разница только в степени сложности иерархического мышления? Существуют доказательства, указывающие, что это действительно так. Когда мы говорили о сознании, я заметил, что для меня «смелый шаг» в том, что мне придется признать компьютер, прошедший тест Тьюринга, сознательным существом. Пока лучшие виртуальные собеседники не могут этого делать (но постоянно совершенствуются), так что для меня вопрос сознания связан с уровнем совершенства системы. Возможно, вопрос о свободе воли для меня решается так же.

Действительно, с философской точки зрения различие между человеческим мозгом и современными компьютерными программами заключается в наличии/отсутствии сознания. Мы считаем, что человеческий мозг обладает сознанием, тогда

как компьютерные программы — нет (пока еще). Может быть, тот же самый фактор определяет наличие свободы воли?

Простой мысленный эксперимент подтверждает, что сознание в самом деле является важнейшей составляющей свободы воли. Рассмотрим ситуацию, когда кто-то совершает некое действие, не отдавая себе в этом отчета, то есть полностью за счет неосознанной активности мозга. Будет ли этот поступок проявлением свободы воли? Большинство людей на этот вопрос ответят отрицательно. Если это действие опасно для окружающих, мы, возможно, все же будем считать совершившего его человека ответственным за его поступок, но учтем предшествовавшие ему сознательные действия, такие как злоупотребление алкоголем, или просто отсутствие критического осмысления собственного поведения.

По мнению некоторых комментаторов, эксперименты Лайбета доказывали отсутствие свободы воли, поскольку продемонстрировали, что очень многие наши действия являются неосознанными. Многие философы сходятся в том, что осознанное принятие решений есть часть свободного волеизъявления, то есть одно из обязательных условий. Однако для многих наблюдателей сознание — необходимое, но не достаточное условие. Если наши решения — осознанные или неосознанные — предопределены уже до того, как мы их принимаем, как же можно заявлять, что мы принимаем их свободно? Такая позиция, в рамках которой свобода воли и детерминизм считаются несовместимыми, называется инкомпатибилизмом. Например, американский философ Карл Джайнет (род. в 1932 г.) считает, что, если события в прошлом, настоящем и будущем детерминированы, значит, мы не можем контролировать ни сами события, ни их последствия. Наши кажущиеся решения и действия — лишь часть этой предопределенной последовательности. И это означает, что свободы воли просто не существует.

Но не все считают, что концепции свободы воли и детерминизма несовместимы. В частности, сторонники позиции

компатибилизма уверены, что мы свободны сделать выбор, даже если этот выбор predetermined или будет predetermined. Так, Дэниел Деннет утверждает, что, хотя будущее может быть детерминировано настоящим, мир настолько сложен, что мы не можем предвидеть, что нам принесет будущее. Мы способны идентифицировать некоторые «ожидания» и имеем свободу совершить действия, противоречащие этим ожиданиям. Следует соотносить наши решения и поступки с этими ожиданиями, а не с теоретически predetermined будущим, которого нам знать не дано. Таким образом, в соответствии с идеями Деннета, свобода воли существует.

Газзанига также придерживается позиции компатибилизма: «Мы несем личную ответственность за наши действия, даже если живем в детерминированном мире»<sup>17</sup>. Циник может прокомментировать это так: вы не контролируете свои действия, но все равно будете чувствовать себя виноватым.

Некоторые мыслители считали идею свободы воли иллюзией. Шотландский философ Дэвид Юм (1711–1776) описывал ее как «вербальную» материю, характеризующуюся «ложным ощущением или кажущимся опытом»<sup>18</sup>. Немецкий философ Артур Шопенгауэр (1788–1860) писал: «Каждый *a priori* считает себя совершенно свободным, даже в своих отдельных поступках, и думает, будто он в любой момент может избрать другой жизненный путь, то есть сделаться другим. Но *a posteriori*, на опыте, он убеждается, к своему изумлению, что он не свободен, а подчинен необходимости, что, несмотря на все свои решения и размышления, он не изменяет своей деятельности и от начала до конца жизни должен проявлять один и тот же им самим не одобряемый характер, как бы играть до конца однажды принятую на себя роль»<sup>19</sup>.

Я бы хотел кое-что добавить. Концепция свободы воли — и тесно связанная с ней идея об ответственности — очень полез-

---

\* Шопенгауэр А. Мир как воля и представление / Пер. Ю. И. Айхенвальда. Собр. соч. в пяти томах. Т. 1. М.: Московский Клуб, 1992.

на и, вообще говоря, жизненно важна для поддержания порядка в обществе, вне зависимости от того, существует на самом деле свобода воли или нет. Как и сознание, свобода воли существует в виде мемы (идеи). Попытки доказать ее существование или хотя бы дать ей определение могут оказаться безнадежной тавтологией, но в реальности практически все в эту идею верят. Очень важная часть понятий высшего иерархического порядка в нашей новой коре связана с идеей свободного выбора и ответственности за свои поступки. Вне зависимости от истинного философского смысла свободы воли и реальности ее существования без этого понятия наше общество было бы гораздо хуже.

Кроме того, мир не обязательно детерминирован. Выше мы говорили о двух точках зрения в квантовой механике, различающихся отношением квантового поля и наблюдателя. Популярная интерпретация, с точки зрения наблюдателя, объясняет роль сознания: частицы проявляют квантовые свойства только тогда, когда за ними наблюдает сознательное существо. Но в философии квантовых явлений есть еще один вопрос, который связывает нас с вопросом о свободе воли, а именно: детерминированы или случайны квантовые явления?

Чаще всего квантовые явления объясняют так: когда происходит «коллапс» волновой функции, описывающей частицу, частица приобретает специфическую локализацию. Если рассматривать большое множество таких событий, наблюдается предсказуемое распределение вероятностей расположения частиц (вот почему волновую функцию рассматривают как распределение вероятности), но распределение для каждой частицы при коллапсе волновой функции случайное. Другая интерпретация основана на детерминистической позиции: существует скрытая переменная, которую мы не можем определить отдельно, но значение которой определяет положение частицы. Итак, значение или фаза скрытой переменной в момент коллапса определяет положение частицы. Большинство специалистов в области квантовой физики, кажется, склоняются к идее о случайном распределении в поле

вероятности, но уравнения квантовой механики позволяют предполагать существование подобной скрытой переменной.

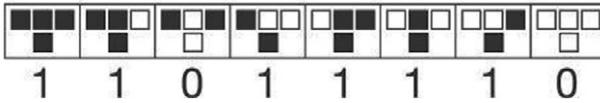
Таким образом, мир, возможно, не является полностью детерминированным. В соответствии с вероятностной интерпретацией квантово-механических законов на самом нижнем уровне реальности существует бесконечный источник неопределенности. Однако эти рассуждения, возможно, не разрешают сомнения инкомпатибилистов. Верно, что при такой интерпретации квантовой механики мир недетерминирован, но наша идея о свободе воли подразумевает нечто большее, чем случайные решения и поступки. Большинство сторонников позиции инкомпатибилизма считают концепцию свободы воли несовместимой со случайным механизмом принятия решений. Они считают, что при наличии свободы воли принятие решений должно осуществляться целенаправленным путем.

Стивен Вольфрам предлагает путь разрешения этой дилеммы. В его книге «Наука нового типа» (A New Kind of Science; 2002) подробно излагается идея клеточных автоматов и их роль во всех аспектах нашей жизни. Клеточный автомат — это механизм, в котором значение информационных ячеек постоянно пересчитывается в зависимости от состояния соседних ячеек. Джон фон Нейман создал теоретическую самореплицирующуюся машину, названную универсальным конструктором, которая, возможно, была первым клеточным автоматом.

Вольфрам иллюстрирует свою идею на примере простейшего клеточного автомата — одномерной группы ячеек. В каждый момент времени каждая ячейка может быть либо черной, либо белой. В каждом цикле значение (цвет) всех ячеек пересчитывается. Значение ячейки в следующем цикле есть функция от ее текущего значения и значений двух соседних ячеек. Каждый клеточный автомат характеризуется правилом, определяющим механизм расчета значения ячейки в следующем цикле.

На рис. 49 в качестве примера представлено так называемое правило 222.

Правило 222

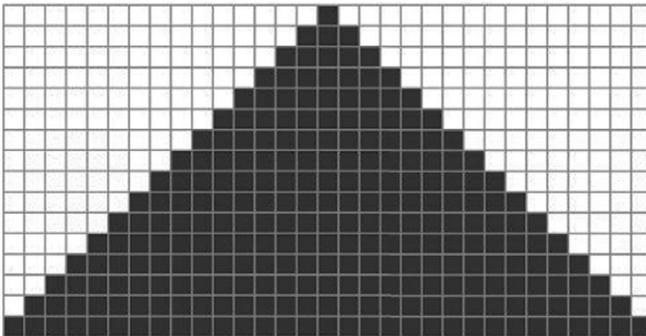


Восемь возможных комбинаций значений для ячейки и ее соседей слева и справа (верхний ряд) и рассчитанное новое значение (нижний ряд). Так, например, если ячейка черная и ее соседи тоже черные, в следующем поколении ячейка остается черной (самая левая клетка). Если ячейка белая, как и ячейка слева, а ячейка справа черная, в следующем поколении ячейка становится черной (вторая клетка справа).

В верхнем ряду изображено восемь возможных состояний ячейки и ее соседей слева и справа. В нижнем ряду представлено новое состояние ячейки. Таким образом, например, если ячейка черная, как и обе ее соседки, в следующем поколении она останется черной (первый рисунок слева). Если же ячейка белая, как и ее соседка слева, а ячейка справа черная, в следующем поколении она станет черной (второй рисунок справа).

Вселенная для такого простого клеточного автомата — лишь один ряд ячеек. Если мы начнем с ряда ячеек с единственной черной ячейкой в середине и проследим за эволюцией этой «вселенной» на протяжении множества поколений (каждый ряд, расположенный ниже предыдущего, соответствует следующему поколению), результат изменений в соответствии с правилом 222 будет выглядеть следующим образом.

Правило 222





математические тесты на случайность распределения, но при этом создают не просто шум, а некие повторяющиеся рисунки. Только повторяются эти рисунки странным, непредсказуемым образом. Если нас спросят, каким будет значение определенной ячейки через триллион триллионов итераций, у нас нет иной возможности ответить на этот вопрос, кроме как запустить машину и провести все эти итерации. Безусловно, решение детерминировано, поскольку это очень простая модель, но оно абсолютно непредсказуемо.

Главный тезис Вольфрама, выдвинутый на основании этих экспериментов, заключается в том, что мир — это один большой клеточный автомат класса IV. Название книги «Наука нового типа» говорит о том, что его теория отличается от большинства других научных теорий. Мы можем определить положение спутника Земли через пять лет от сегодняшнего момента или через любой другой срок, не прибегая к симуляции его положения в каждый момент времени, а воспользовавшись соответствующими физическими законами. Но будущее состояние автоматов класса IV нельзя предсказать без симуляции всех этапов. Если Вселенная — это гигантский клеточный автомат, как считает Вольфрам, то на свете нет такого компьютера (любой компьютер будет соответствовать лишь части Вселенной), который мог бы осуществить такую симуляцию. И поэтому будущее состояние Вселенной совершенно непредсказуемо, хотя и детерминировано.

Таким образом, даже если наши решения детерминированы (поскольку наше тело и мозг являются частью детерминированной Вселенной), они при этом совершенно непредсказуемы, поскольку мы живем в автомате класса IV и являемся его частью. С точки зрения Вольфрама, это достаточное основание, чтобы признавать существование свободы воли.

Нам нет нужды обращаться к рассмотрению всей Вселенной, чтобы анализировать детерминированные, но непредсказуемые события. Ни один из ученых, работавших с Ватсоном, не смог

предсказать его решения, поскольку его программа слишком сложна и вариабельна, а его возможности основаны на столь обширном объеме знаний, который не под силу одолеть ни одному человеку. Если мы допускаем наличие свободы воли у человека, следовательно, нам придется допустить ее наличие и у будущих версий Ватсона и аналогичных машин.

Лично я считаю, что люди обладают свободой воли, и поступаю в соответствии с этой моей позицией, однако мне не удастся найти пример из собственной жизни, иллюстрирующий этот принцип. Например, решение написать эту книгу. Я ведь никогда такого решения не принимал. Сама идея книги заставила меня это сделать. Обычно меня захватывают идеи, которые как бы поселяются в моей новой коре и диктуют свои правила. А что насчет решения жениться, которое я принял (вместе с еще одним человеком) 36 лет назад? В то время я следовал обычной программе — найти и удержать привлекательную девушку. И тогда я влюбился. Где здесь свобода воли?

А что можно сказать о малозначащих решениях, принимаемых мной каждый день? Например, о выборе слов в книге, над которой сейчас работаю? Я начинаю с чистого виртуального листа. Никто не говорит мне, что я должен делать. Нет редактора, который заглядывал бы мне через плечо. Мой выбор *полностью* зависит от меня самого. Я свободен — *полностью свободен* — написать *все, что я...*

*Понимаю...*

*Понимаю?* Да, я сделал это — я наконец проявил свободу воли. Я собирался написать слово «захочу», но принял свободное решение и написал вместо этого что-то совершенно неожиданное. Возможно, мне впервые в жизни удалось реализовать свободу воли.

Впрочем, это не совсем так.

Думаю, понятно, что этот пример — не демонстрация моей свободы воли, а скорее попытка проиллюстрировать мой тезис (и проявить слабое чувство юмора).

Я разделяю точку зрения Декарта о том, что являюсь сознательным существом, но относительно свободы воли у меня меньше уверенности. Очень трудно уйти от вывода Шопенгауэра: «Ты можешь делать то, что ты хочешь; но в каждое данное мгновение твоей жизни ты можешь *хотеть* лишь чего-то определенного и, безусловно, ничего иного, кроме этого одного»<sup>\*20</sup>. Я тем не менее буду продолжать действовать так, как будто у меня есть свобода воли, и стану верить в это, пока мне не придется объяснить почему.

## Личность

Однажды философу приснился сон.

Сначала перед ним появился Аристотель, и философ спросил его: «Можешь за пятнадцать минут изложить мне суть твоего учения?» К его удивлению, Аристотель меньше чем за четверть часа представил блестящее изложение, в которое вместили невообразимое количество материала. Но, когда философ что-то возразил, Аристотель не смог ему ответить и, сконфуженный, исчез. После него появился Платон. И с ним произошла та же история, и замечание философа произвело тот же эффект, как и в случае с Аристотелем. Платон тоже не смог ответить и исчез.

Потом один за другим появлялись все самые известные в мире философы, и наш философ опроверг их всех одним и тем же замечанием.

После поражения последнего философа наш философ заявил ему: «Я знаю, что сплю и вижу все это во сне. Но я нашел универсальный контраргумент для всех философских концепций! Завтра я проснусь и, возможно, забуду его, и миру будет чего-то недоставать!» Усилием железной воли философ заставил себя проснуться, бросился к столу и записал свой универсальный контраргумент. А затем с облегчением нырнул в постель.

На следующее утро он проснулся, подошел к столу и прочел, что написал. Написано было: «Это *ты* так говоришь».

*Рэймонд Смаллиан\*\**, в пересказе Дэвида Чалмерса<sup>21</sup>

---

\* Шопенгауэр А. Две основные проблемы этики / Пер. Ю. И. Айхенвальда. Собр. соч. в шести томах. М. : ТЕРРА — Книжный клуб; 1999–2001.

\*\* Рэймонд Меррилл Смаллиан (род. в 1919) — американский математик, логик, философ, пианист и фокусник.

Что меня лично интересует гораздо больше, чем наличие у меня сознания и свободы воли, это то почему мне принадлежит сознание и опыт именно такого человека, который пишет книги, любит пешие и велосипедные прогулки, принимает пищевые добавки и т. п. Очевидный ответ — потому что ты такой.

Это, вероятно, тавтология того же сорта, что и в определениях сознания и свободы воли. Однако в данном случае у меня есть более точный ответ на вопрос, почему мое сознание ассоциируется именно с этой определенной личностью: по той причине, что я сам себя таким сделал.

Знаете афоризм: «ты то, что ты ешь»? Еще правильнее будет сказать: «ты то, что ты думаешь». Как мы уже обсуждали, все иерархические структуры новой коры, определяющие мою личность, умения и знания, являются результатом моих собственных мыслей и опыта. Люди, с которыми я иду на контакт, идеи и проекты, к которым я подключаюсь, являются основными факторами, определяющими мое становление. В этом смысле выбор еды тоже отражает активность моей новой коры. Учитывая потенциальную возможность существования свободы воли, можно сказать, что я сам принимаю решение, каким стать.

Вне зависимости от того, как мы пришли к тому, что мы есть, каждый из нас желает сохранить свою индивидуальность. Если бы вы не хотели выжить, вы бы не читали сейчас эту книгу. Каждое существо преследует эту цель — это определяющий фактор эволюции. Хотя дать определение личности, возможно, еще труднее, чем сознанию и свободе воли, тема эта гораздо важнее. В итоге, если мы хотим продолжить свое существование, мы должны понимать, кто мы есть.

Проведем следующий мысленный эксперимент. Представим себе, что находимся в будущем, где существуют гораздо более совершенные технологии, чем сегодня. Когда вы спите, исследователи сканируют ваш мозг и отмечают все важные детали. Возможно, они делают это с помощью аппарата размером с клетку крови, которая перемещается по капиллярам вашего мозга,

или посредством какой-то другой неинвазивной технологии, но они имеют всю информацию о вашем мозге в каждый конкретный момент времени. Они также регистрируют состояние тела, способное отражать состояние вашего разума, например активность эндокринной системы. Затем они устанавливают этот «файл» в небиологическое тело, которое выглядит и движется, как вы, и вполне способно выдать себя за вас. Утром вам сообщают о том, что было сделано, и показывают вам вашего «мозгового двойника» (возможно, так, что он вас не видит). Назовем его Вы-2. Вы-2 рассказывает о своей жизни так, как если бы это были вы. А также рассказывает, что сегодня утром он обнаружил, что его снабдили гораздо более надежным новым телом — версией 2.0. «Хм, мне это тело нравится!» — замечает ваш двойник.

Наш первый вопрос заключается в следующем: является ли Вы-2 сознательным существом? На вид он вполне сознателен. Он вполне может пройти тест, о котором я упоминал выше, как чувствующее и сознательное существо. Если вы обладаете сознанием, то Вы-2 тоже.

Таким образом, если вы вдруг исчезнете, никто этого не заметит. Вы-2 заявит всем, что он — это вы. Все ваши друзья и близкие будут рады за вас, поскольку вы теперь счастливый обладатель нового, более надежного тела и ментального субстрата, чем раньше. Возможно, ваши более склонные к философствованию друзья окажутся слегка обеспокоены, но в целом все останутся счастливы, включая вас самого (ну или, по крайней мере, того, кто выдает себя за вас).

Так что же, нам больше не нужно ваше старое тело и старый мозг, так? Можно их выбросить?

Думаю, вы с таким положением вещей не согласитесь. Я сказал, что сканирование производилось неинвазивным способом, так что вы живы и находитесь в сознании. Более того, ваше самосознание с вами, а не с Вы-2, даже если Вы-2 думает, что является вашим продолжением. А возможно, Вы-2 даже не знает, что

вы существуете или когда-то существовали. Да и вы бы ничего не узнали, если бы вам не сказали.

Какой вывод? Вы-2 обладает сознанием, но это другое существо, другая личность. Он чрезвычайно похож на вас, больше, чем генетический клон, поскольку обладает теми же, что и вы, неокортикальными образами и контактами. Точнее сказать, обладал в момент создания. С этого момента каждый из вас пошел своим путем, и новая кора мозга каждого из вас — тоже. Вы все еще существуете. У вас другой опыт, нежели у Вы-2. Вывод такой: Вы-2 — это не вы.

Ну хорошо, пока все ясно. Теперь проведем другой мысленный эксперимент, я считаю, более реалистичный в плане возможного развития человечества. Вы подвергаетесь хирургической операции по замене очень маленького фрагмента мозга небиологическим материалом. Вы уверены, что процедура безопасна и принесет значительную пользу.

Это не такая уж фантастическая ситуация, поскольку данную процедуру уже сейчас рутинно используют для лечения людей с неврологическими и сенсорными нарушениями. Например, имплантаты нейронов при болезни Паркинсона или кохлеарные имплантаты при глухоте. В таких случаях внутрь тела вводят компьютеризированное устройство, которое располагается вне мозга, но связано с ним (в случае кохлеарных имплантатов — со слуховым нервом). Мне не кажется значимым с философской точки зрения, находится ли компьютер непосредственно внутри мозга или снаружи: с помощью компьютеризированного устройства мы эффективным образом усиливаем мозг и заменяем те его функции, которые больше не выполняются должным образом. В 2030-х гг., когда разумные компьютеризированные устройства будут иметь размеры клеток (замечу, что белые клетки крови достаточно разумны, чтобы обнаружить и уничтожить патогенов), мы станем вводить их в организм неинвазивными способами, без хирургических операций.

Возвратимся к нашему сценарию. Вам провели необходимую процедуру, и все прошло успешно — улучшились некоторые ваши функции (например, память). Но остались ли вы самим собой? Ваши друзья в этом абсолютно уверены. Вы тоже в этом уверены. Нет никаких достаточных оснований полагать, что вы вдруг стали другим человеком. Конечно, вы подверглись данной процедуре, чтобы что-то изменить, но вы по-прежнему тот же самый. Ваша личность не изменилась. Никакое чужое сознание не завладело вашим телом.

Хорошо, ободренный результатом, вы решаетесь на следующую процедуру, в этот раз на другом участке мозга. Результат такой же: ваши функции слегка выправлены, но вы — по-прежнему вы.

Думаю, вам понятно, к чему я клоню? Вы рассматриваете возможность новых процедур. Ваша вера в возможности медицины возрастает. И, наконец, приходит момент, когда вы полностью заменили весь свой мозг. Каждый раз процедура проводилась так тщательно, что сохранились все образы и контакты, содержащиеся в новой коре, так что вы не потеряли никаких личных качеств, навыков или воспоминаний. Никогда не существовало двух персон — вы и Вы-2. Всегда существовали только вы. Никто, включая вас самого, не заметил, что вы перестали существовать. Вот вы тут.

Какой вывод? Вы по-прежнему существуете. Никакой дилеммы нет, все в порядке.

*Лишь одно замечание:* в результате постепенных замен фрагментов мозга вы стали полностью эквивалентны Вы-2 из предыдущего мысленного эксперимента (назову его «сценарий сканирования и воспроизведения»). В результате сценария постепенных замен вы сохранили все образы и контакты новой коры, но только в небиологическом субстрате, что так же справедливо для Вы-2 из сценария сканирования и воспроизведения. В результате сценария постепенных замен вы получили некоторые дополнительные возможности и стали «здоровее», чем

прежде. Но абсолютно то же самое произошло и в случае Вы-2 из сценария сканирования и воспроизведения.

Однако ранее мы пришли к выводу, что Вы 2 — это *не вы*. Но если вы в результате процесса постепенных замен полностью эквивалентны Вы-2, возникшему в результате сканирования и воспроизведения, то вы — это не вы.

Это, однако, противоречит только что сделанному выводу. Процесс постепенных замен состоит из множества стадий. Каждая стадия, казалось бы, не изменяет идентичности человека (по крайней мере, именно так мы рассуждаем сегодня в отношении пациента с болезнью Паркинсона, перенесшего процедуру имплантации нейрона).

Эта некая философская дилемма заставляет некоторых людей сделать вывод, что подобные сценарии замены никогда не реализуются (хотя они уже имеют место). Но поймите вот что: мы в процессе жизни самым естественным образом претерпеваем постепенные изменения. Большинство клеток нашего тела постоянно заменяются новыми (к примеру, пока вы читали последнее предложение, в вашем теле сменилось около 100 млн клеток). Клетки выстилки тонкой кишки сменяются примерно каждую неделю, как и клетки выстилки желудка. Продолжительность жизни различных клеток крови варьирует от нескольких дней до нескольких месяцев. Тромбоциты существуют примерно девять дней.

Нейроны — долгожители, но их органеллы и составляющие их молекулы заменяются примерно каждый месяц<sup>22</sup>. Время полужизни микротрубочек нейронов составляет около десяти минут, актиновые филаменты в дендритах существуют около 40 секунд, белки, обеспечивающие энергию синапсов, заменяются каждый час, а рецепторы НМДА (N-метил-D-аспарагиновой кислоты) в синапсах живут сравнительно долго — около пяти дней.

Таким образом, все наше тело обновляется каждые несколько месяцев, что вполне сопоставимо с описанным выше сценарием

постепенных замен. Как вы думаете, вы тот же человек, что и несколько месяцев назад? Конечно, отличия есть. Возможно, вы узнали несколько новых вещей. Но вы воспринимаете себя той же личностью и вовсе не ощущаете, что постоянно расцепляе-тесь и восстанавливаетесь вновь.

Представьте себе реку. Вот, например, позади моего офиса протекает Чарльз Ривер. Когда я смотрю на нее сейчас, могу ли я сказать, что это та же река, что и день назад? Давайте сначала решим, что такое река. В словаре говорится, что река — это «большой естественный поток проточной воды». Если придерживаться этого определения, то сегодняшняя река отличается от вчерашней. Все молекулы воды в ней другие — этот процесс происходит очень быстро. Греческий философ Диоген Лаэртский писал в III в. н. э., что невозможно дважды войти в одну и ту же воду.

Однако обычно мы воспринимаем реку иначе. Люди любят смотреть на реку, поскольку она воспринимается как символ непрерывности и стабильности. С обывательской точки зрения, сегодня Чарльз Ривер та же, что и вчера. И наша жизнь практически не меняется. Мы — это не только материал, из которого сделаны наши тела и мозг. Частички вещества протекают через нас, как молекулы воды протекают по реке. Мы — слабо меняющийся образ, мы стабильны и непрерывны, даже если составляющий нас материал быстро обновляется.

Постепенное введение небιологических систем в наши тела и мозг будет еще одним примером этого непрерывного оборота составляющих нас элементов. Оно не повлияет на целостность личности больше, чем естественная замена клеток тела. Мы уже в значительной степени передали нашу историческую, интеллектуальную, социальную и персональную память различным устройствам и облаку. Устройства, с которыми мы общаемся для обращения к этой памяти, пока еще вне наших тел и мозга, но поскольку их размер все сокращается (примерно в 100 раз каждые 10 лет), они внедряются и внутрь. И, честно говоря, это

будет удобно — так мы их не потеряем. Но, если люди не захотят размещать внутри себя микроскопические устройства, это не страшно, поскольку возникнут другие пути доступа к всепроникающему разуму облака.

Однако вернемся к дилемме, о которой я говорил выше. После процесса постепенных замен вы становитесь эквивалентны Вы-2 из сценария со сканированием и воспроизведением, однако мы установили, что Вы-2 в том сценарии — это не вы. Что же это означает?

Это означает, что мы с вами нащупали свойство, которым обладают небиологические системы в отличие от биологических. Речь идет о возможности копирования, запоминания и воссоздания. Мы постоянно делаем это с различными устройствами. Когда у нас появляется новый смартфон, мы копируем в него все наши старые файлы, так что он во многом сохраняет особенности, навыки и память старого смартфона. Вероятно, у него есть новые возможности, но содержимое старого телефона по-прежнему с нами. Аналогичным образом, такие программы, как Ватсон, безусловно, запоминаются. Если вдруг завтра компьютер сломается, Ватсона легко будет восстановить из хранящихся в облаке файлов.

Такие возможности небиологических систем не реализуются в биологическом мире. Это достоинство, а не недостаток, и именно потому мы так стремимся продолжать загрузку всей нашей памяти в облако. Мы, безусловно, будем продолжать двигаться в этом направлении, поскольку небиологические системы перенимают у нашего биологического мозга все больше и больше способностей.

Я вижу такой путь разрешения проблемы: неправда, что Вы-2 — это не вы. Это вы. Просто теперь «вас» двое. Что не так плохо: если вы думаете о себе хорошо, то два хороших существа — это ведь еще лучше.

Я считаю, что мы действительно станем придерживаться сценария постепенных замен, пока бóльшая часть нашего

## Глава девятая

мыслительного процесса не начнет происходить в облаке. По вопросу о сохранении индивидуальности я считаю, что индивидуальность (личность) сохраняется через непрерывность хранящихся в нас информационных образов. Непрерывность не противоречит непрерывным изменениям, и если я сегодня чуть-чуть отличаюсь от того, каким я был вчера, я тем не менее сохраняю свою индивидуальность. Однако непрерывность образов, составляющих мою индивидуальность, не зависит от субстрата. Биологические субстраты прекрасны, они позволили нам продвинуться очень далеко. Но мы создаем еще более пригодные и надежные субстраты для очень хороших дел.

# Глава десятая

## ЗАКОН УСКОРЕНИЯ ОТДАЧИ В ПРИМЕНЕНИИ К МОЗГУ

И хотя человек в некотором отношении останется наиболее совершенным созданием, разве это не соответствует общему принципу развития природы, который позволяет некоторым животным, далеко не самым совершенным, иметь приоритет в каких-то вещах? Разве природа не позволила муравьям и пчелам сохранить свое преимущество перед людьми в способности организовывать социальные сообщества, птицам — в способности перемещаться по воздуху, рыбам — в плавании, лошадям — в силе и скорости, а собакам — в самоотверженности?

*Сэмюэл Батлер, 1871*

Было время, когда наша Земля была полностью лишена растительной и животной жизни и, как считают наши лучшие философы, представляла собой просто горячий круглый шар, покрытый постепенно остывающей коркой. Если бы человек жил в те времена и увидел бы Землю в таком состоянии как совершенно незнакомый и чужой ему мир, и если бы тот человек был совершенно незнаком с законами физики, он наверняка заявил бы, что на этой испепеленной земле не могут появиться и эволюционировать сознательные существа. Он наверняка не согласился бы, что у этой планеты есть хотя бы потенциальная возможность породить сознание. Однако по прошествии времени сознание появилось. Так неужели нельзя предположить, что сознание может возникать и другими путями, хотя пока мы не видим никаких признаков этого процесса?

*Сэмюэл Батлер, 1871*

## Глава десятая

Если мы задумаемся о множестве уже сменившихся фаз жизни и сознания, вряд ли можно говорить о невозможности возникновения других форм и о том, что животная жизнь — это конец всего. Когда-то концом всего был огонь, когда-то — камни и вода.

*Сэмюэл Батлер, 1871*

Тот факт, что сегодня машины почти не обладают сознанием, не гарантирует, что когда-то не возникнет механическое сознание. У моллюска малоразвитое сознание. Посмотрите, какие чрезвычайные успехи достигнуты в развитии машин за последние сотни лет и как медленно развиваются царства животных и растений. Можно сказать, что наиболее высокоорганизованные машины — это продукт не вчерашнего дня, а последних пяти минут. Подумайте, что сознательные существа существуют уже около двадцати миллионов лет, и посмотрите, какие шаги проделали машины за последнее тысячелетие! Может мир просуществовать еще двадцать миллионов лет? Если так, чем только они не смогут стать под конец?

*Сэмюэл Батлер, 1871*

**М**ой главный тезис, который я называю «законом ускорения отдачи» (ЗУО), заключается в том, что параметры развития информационных технологий следуют предсказуемой экспоненциальной траектории, опровергая общепризнанное суждение, что мы не можем предсказать будущее. Остается еще множество неизвестных (какой проект, компания или технический стандарт займут лидирующие позиции на рынке, когда установится мир на Среднем Востоке), но изменение цены, производительности и емкости информационных технологий абсолютно предсказуемо. Удивительно, но эти тенденции не зависят от того, находимся мы в состоянии войны или мира, процветания или кризиса.

Главная причина возникновения мозга в ходе эволюции заключается в необходимости предвидеть будущее. Когда-то, тысячи лет назад, один из наших предков пересекал саванну и, возможно, заметил животное, двигавшееся в его сторону. Возможно, он предсказал, что, если идти в прежнем направлении, их пути пересекутся. Человек решил сменить направление движения, и это спасло ему жизнь.

Однако такие встроенные системы предвидения будущего развиваются линейно, а не экспоненциально, что связано с линейной организацией новой коры. Вспомните, что новая кора постоянно делает предсказания: какую букву или слово мы увидим в следующий момент, кого встретим за углом и т. д. При обработке каждого образа новая кора проходит линейную последовательность шагов, вот почему экспоненциальное мышление нам не свойственно. Мозжечок тоже делает линейные предсказания. Когда он помогает нам поймать мяч, он составляет линейный прогноз о том, в каком участке поля зрения окажется мяч и где при этом должна находиться наша рука.

Как я указывал ранее, линейная и экспоненциальная прогрессия различаются чрезвычайно сильно (40 единиц составляют 40, а 40 в экспоненте — триллион), вот почему мои предсказания, основанные на законе ускорения отдачи, поначалу многих

удивляют. Мы должны научиться мыслить экспоненциально. Когда дело касается информационных технологий, именно так и нужно думать.

Лучшей иллюстрацией закона ускорения отдачи является сглаженный, дважды экспоненциальный рост цены — производительности вычислений, сохраняющийся на протяжении 110-летнего периода, включая две мировые войны, Великую депрессию, холодную войну, распад СССР, расцвет китайской экономики, финансовые кризисы последних лет и многие другие значимые события.

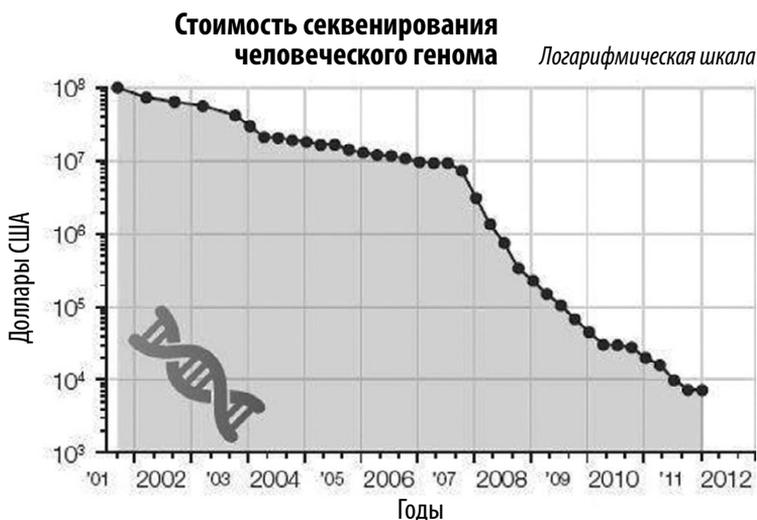
Некоторые люди называют это явление «законом Мура», но они неправы. Закон Мура утверждает, что каждые два года количество элементов на интегральной схеме удваивается и они действуют быстрее, поскольку их размер уменьшается. Но закон Мура — это лишь одна парадигма (причем не первая, а пятая), в основе экспоненциального роста показателя цены — производительности вычислений.

Экспоненциальный рост скорости вычислений начался с первой автоматизированной переписи населения США в 1890 г. — за десять лет до рождения Гордона Мура. В книге «Сингулярность уже близка» я привел соответствующий график до 2002 г., а сейчас продлеваю его до 2009 г. (см. ниже). Предсказанная траектория продолжается, даже несмотря на недавний спад экономики.

Скорость вычислений — самый важный пример действия закона ускорения отдачи, что связано со значительным количеством данных, повсеместным применением вычислительной техники и ее ключевой ролью в обновлении всех важнейших технологических процессов. Но это далеко не единственный пример. Как только какая-то технология становится информационной технологией, она начинает подчиняться закону ускорения отдачи.

Одним из важнейших направлений технологического развития последних лет является биомедицина. Исторически прогресс

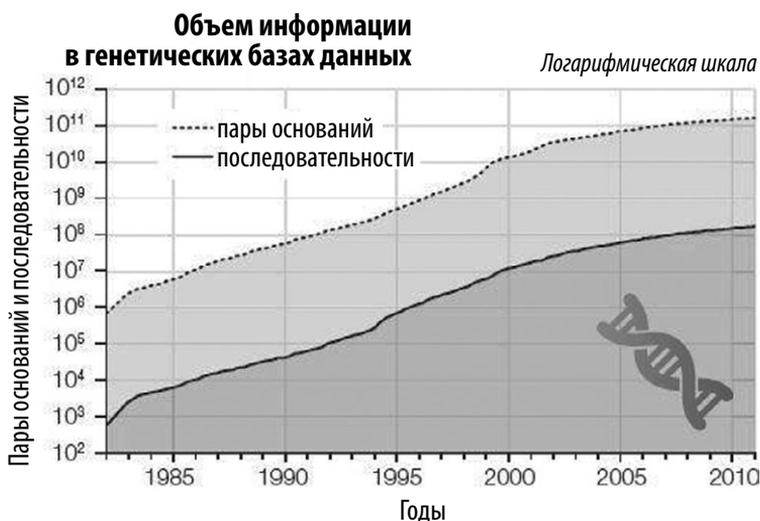
в медицине происходил за счет случайных открытий и поэтому был линейным, а не экспоненциальным. И все же он оказался весьма значителен: продолжительность жизни выросла от 23 лет 1000 лет назад до 37 лет 200 лет назад и почти до 80 лет в наши дни. После открытия программного обеспечения жизни — генома — медицина и биология человека стали информационными технологиями. Сам проект «Геном человека» развивался по экспоненциальному закону: от начала проекта в 1990 г. ежегодно количество генетических данных удваивалось, а стоимость определения одной пары оснований сокращалась вдвое<sup>1</sup>. Данные на всех приведенных ниже рисунках обновлены по сравнению с версией, представленной в книге «Сингулярность уже близка».



*Стоимость секвенирования генома размером с геном человека.*

Теперь мы имеем возможность планировать биомедицинские процедуры на компьютере и тестировать их на симуляторах, причем объемы и точность этих данных также удваиваются каждый год. Кроме того, мы можем исправлять наши стареющие программы: метод интерференции РНК позволяет отключать

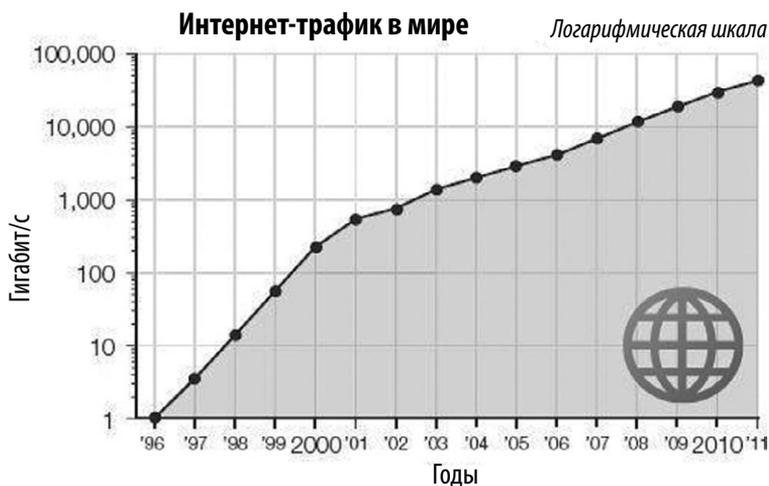
гены, новые формы генной терапии позволяют вводить новые гены, причем не только новорожденным, но и взрослым людям. Успехи в генетических технологиях также влияют на реализацию проекта обратного проектирования человеческого мозга, поскольку важная часть проекта заключается в изучении механизмов генетического контроля функций мозга, таких как создание новых контактов в новой коре при усвоении новых знаний. По мере продвижения от секвенирования генома к его синтезу возникает множество других примеров синтеза биологии и информационных технологий.



*Объем генетических данных, ежегодно секвенируемых во всем мире.*

Еще одна информационная технология, развивающаяся по такому же закону, — это методы коммуникации, позволяющие нам общаться друг с другом и передавать большие массивы данных. Прогресс в данной области измеряют несколькими способами. Закон Купера, который утверждает, что общий объем информации (в битах), передаваемой по беспроводной связи, удваивается каждые 30 месяцев, действовал с того момента, когда Маркони впервые использовал беспроводной телеграф

для передачи символов азбуки Морзе в 1897 г., и до появления современных 4G-технологий<sup>2</sup>. В соответствии с этим законом количество информации, которое может быть передано с помощью беспроводной связи, удваивалось каждые 2,5 года на протяжении столетия. К другим примерам относится количество узлов Интернета и количество бит информации, передаваемых в секунду с помощью Интернета: оба показателя каждый год удваиваются.



*Ширина пропускания международных интернет-каналов.*

Я заинтересовался предсказаниями некоторых аспектов развития технологии по той причине, что примерно 30 лет назад понял, что успешность изобретателя (эту профессию я выбрал, когда мне было пять лет) зависит от времени появления его изобретений. Многие изобретения не нашли своего применения не из-за того, что не работали, а из-за того, что были сделаны в неподходящее время — либо до того, как для их распространения сложились все необходимые условия, либо слишком поздно, когда возможность уже была упущена.

Около тридцати лет назад я начал собирать данные относительно состояния технологии в различных областях. Когда

я стал заниматься этим делом, я не ожидал, что у меня сложится какая-то четкая картина, но надеялся, что смогу нащупать некоторые закономерности. Моей задачей было — и остается — выбрать время для моих собственных технологических разработок, так чтобы они соответствовали тому миру, который существует в момент завершения проекта. А мир этот, понятное дело, очень сильно отличается от мира, существовавшего в начале развития проекта.



*Максимальная ширина пропускания (скорость) интернет-каналов.*

Посмотрите, как сильно и быстро меняется мир. Всего несколько лет назад люди не использовали социальные сети (например, Facebook был основан в 2004 г., а сегодня у этой сети около миллиарда активных пользователей), вики, блоги или твиттеры. В 1990-х гг. мало кто пользовался поисковыми программами и сотовыми телефонами. А теперь представьте себе современный мир без них. Кажется, все эти устройства появились так давно, а на самом деле совсем недавно. В ближайшем будущем мир станет меняться еще более удивительным образом.

Пока я учился, я сделал поразительное открытие: если речь идет об информационных технологиях, изменения значений цены — производительности и емкости (на единицу времени, цены или другого параметра) практически точно следуют экспоненциальному закону.

Эти изменения обгоняют некоторые научные концепции, на которых они основаны (такие как закон Мура). Но когда одна парадигма выходит из общего потока (например, когда в 1950-х гг. инженеры оказались неспособны дальше уменьшать размер и стоимость радиоламп), возникает такая исследовательская активность, которая рождает новую парадигму, и начинается новая S-образная кривая прогресса. Экспоненциальный участок этой S-образной кривой поддерживает экспоненциальную траекторию развития информационных технологий. Так радиолампы 1950-х гг. сменились транзисторами, а затем интегральными схемами и законом Мура в конце 1960-х гг. Закон Мура, в свою очередь, породил трехмерные вычисления, ранние примеры которых реализуются уже сейчас. Причина, по которой информационные технологии способны постоянно преодолевать ограничения любой частной парадигмы, заключается в том, что ресурсы, необходимые для расчета, запоминания или передачи бита информации, исчезающе малы.

Может возникнуть вопрос, есть ли какие-то принципиальные ограничения для развития нашей способности осуществлять вычисления и передавать информацию? На основе наших сегодняшних представлений в рамках физических законов ответ такой: да, есть. Однако эти ограничения пока не очень нас ограничивают. Молекулярные вычисления позволят расширить объем нашего разума в триллионы раз. По моим прогнозам, мы достигнем этого предела в конце текущего столетия.

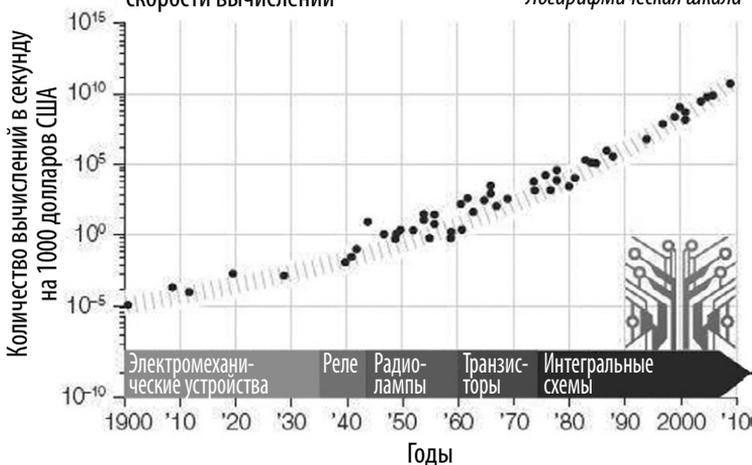
Важно отметить, что не все экспоненциальные процессы служат проявлениями закона ускорения отдачи. Некоторые наблюдатели неправильно истолковывают этот закон, пытаясь применить его к процессам, не имеющим отношения

к информационным технологиям. Например, они указывают, что современные мужские бритвы действительно содержат не одно, а два лезвия, но где бритвы с восемью лезвиями? Однако в бритвах (пока) не используются информационные технологии.

### Экспоненциальный рост скорости вычислений за последние 110 лет

Закон Мура был пятой, а не первой парадигмой, объясняющей экспоненциальный рост скорости вычислений

*Логарифмическая шкала*



*Количество вычислений в секунду на (условную) тысячу долларов, осуществляемых различными вычислительными устройствами.*

В книге «Сингулярность уже близка» я провел теоретическое исследование с математическими выкладками (они содержатся в приложении к той книге), объясняющими удивительную предсказуемость закона ускорения отдачи. Суть в том, что для создания каждой новой технологии мы всегда используем самую последнюю из существующих технологий. Технологии выстраиваются друг на друге по экспоненциальному закону, и этот процесс легко оценить. В 1990 г. мы использовали компьютеры и другие приборы того времени для создания компьютера 1991 г. В 2012 г. мы применяли современные

информационные приборы для создания машин 2013 и 2014 гг. В более широком плане это ускорение затрагивает любой процесс, в котором происходит эволюция процессов обработки информации. Именно поэтому мы наблюдаем ускорение биологической эволюции и аналогичное (но гораздо более быстрое) ускорение технологической эволюции, которое само по себе является одним из направлений биологической эволюции.



*Количество операций с плавающей запятой в секунду для различных суперкомпьютеров.*

У меня собраны предсказания, сформулированные на основании закона ускорения отдачи, сделанные более чем за четверть века, начиная с тех, что я представлял в книге «Эпоха мыслящих машин» (середина 1980-х гг.). Вот примеры некоторых точных предсказаний из этой книги: появление в середине 1990-х гг. широкой всемирной сети коммуникаций, связывающей между собой людей в разных уголках мира и обеспечивающей доступ к обширным базам человеческих знаний; сильнейшая волна де-

\* На настоящий момент эти суперкомпьютеры уже существуют.

мократизации, связанная с возникновением этой децентрализованной коммуникационной сети, уничтожившей Советский Союз; поражение чемпиона мира по шахматам к 1998 г.; а также многое другое.



*Количество транзисторов в чипе для различных процессоров Intel.*

Закон ускорения отдачи в применении к вычислительной технике я подробно описывал в книге «Эпоха мыслящих машин». Там я привел данные за сто лет, показывающие дважды экспоненциальный характер изменения цены — производительности компьютерных вычислений вплоть до 1998 г. Ниже представлены адаптированные данные, охватывающие период до 2009 г.

Недавно я составил 146-страничный обзор предсказаний, которые я сделал в книгах «Эпоха мыслящих машин», «Эпоха духовных машин» и «Сингулярность уже близка» (этот обзор можно найти здесь: [www.kurzweilai.net/how-my-predictions-are-faring-an-update-by-ray-kurzweil](http://www.kurzweilai.net/how-my-predictions-are-faring-an-update-by-ray-kurzweil)). В книге «Эпоха духовных машин», которую я написал в 1990-х гг., содержались сотни предсказаний по десятилетиям (2009, 2019, 2029 и 2099 гг.).

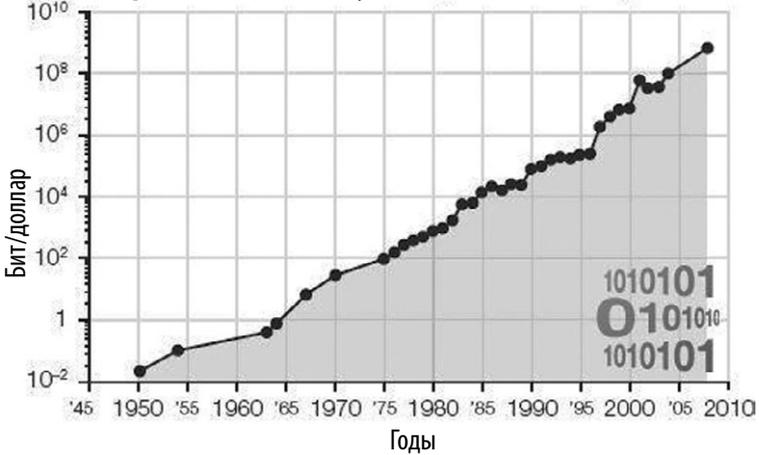
Например, я сделал 147 предсказаний на 2009 г., из которых 115 (78%) полностью сбылись. Наиболее точными были предсказания относительно емкости и цены — производительности информационных технологий. Еще 12 предсказаний (8%) оказались «в основном справедливыми». Таким образом, всего полностью или почти полностью справедливыми получились 127 предсказаний (86%). Поскольку предсказания делались на десятилетие, предсказание на 2009 г., которое сбывалось только в 2010 или 2011 гг., рассматривалось лишь как «в основном справедливое». Еще 17 предсказаний (12%) сбылись лишь отчасти и три (2%) оказались неверными.



*Динамическая оперативная память (DRAM) в битах на доллар.*

Но даже «неверные» предсказания не были полностью ошибочными. Например, я расценил мое предсказание о появлении самодвижущихся машин как неверное, хотя Google уже продемонстрировал использование самодвижущихся машин, а в октябре 2010 г. четыре электрических грузовика без водителя успешно преодолели путь в 13 тыс. км от Италии до Китая<sup>3</sup>. Эксперты в данной области считают, что эта технология станет рутинной к концу текущего десятилетия.

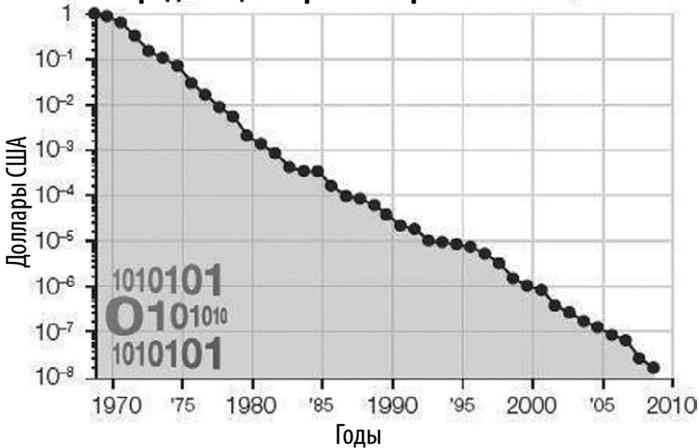
**Память с произвольным доступом (RAM)** *Логарифмическая шкала*



Память с произвольным доступом (RAM) в битах на доллар.

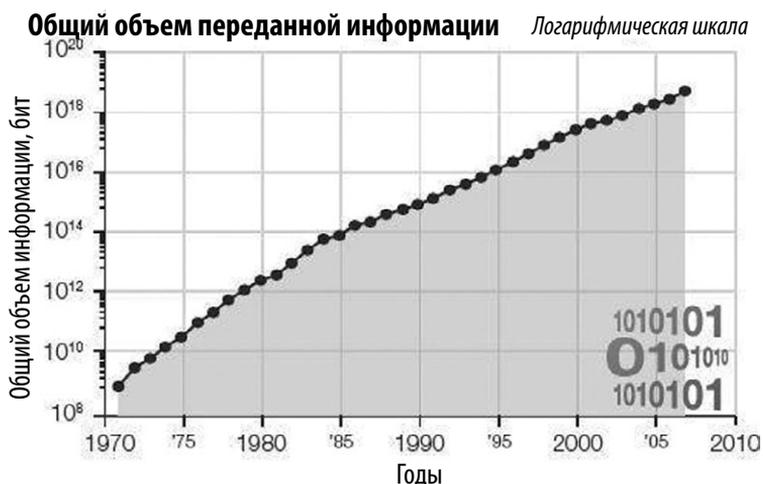
Экспоненциальный рост возможностей вычислительных и коммуникационных технологий способствует реализации проекта, направленного на понимание и воспроизведение методов функционирования головного мозга человека. Это не единый организованный проект, а скорее целая группа разнообразных

**Средняя цена транзистора** *Логарифмическая шкала*



Средняя цена транзистора в долларах США.

проектов, в рамках которых решаются такие задачи, как моделирование мозга на самых разных уровнях — от отдельных нейронов до всей коры, составление карты коннектома (контактов между нейронами), симуляция функции отдельных участков мозга и т. д. И все эти направления развиваются по экспоненциальному закону. Многие приведенные выше примеры совсем новые. Например, описанное в четвертой главе исследование Ван Видена, показавшее упорядоченную и «простую» (как пишут авторы) сеть контактов в новой коре, было проделано лишь в 2012 г.



Исследователи признают, что их выводы (и изображения) удалось получить только с помощью новейших технологий высокого разрешения. Временное и пространственное разрешение изображений, получаемых при сканировании мозга, тоже совершенствуется по экспоненциальному закону. Сейчас существуют разные способы сканирования мозга — от полностью неинвазивных методов, которые можно использовать для изучения человеческого мозга, до более инвазивных и деструктивных, применяемых на животных моделях.

Метод магнитно-резонансной томографии (МРТ) — неинвазивный метод получения изображений с достаточно высоким

**Объем памяти магнитных накопителей информации в битах на доллар (в долларах 2000 г.)**

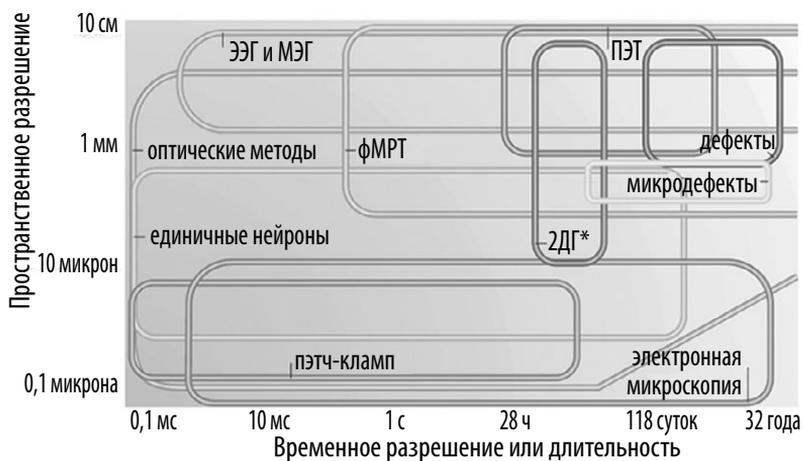


временным разрешением — постоянно совершенствуется в соответствии с тем же экспоненциальным законом; на сегодня пространственное разрешение метода составляет около 100 мкм. Деструктивные методы получения изображений, которые используются для анализа коннектомов животных, тоже развиваются



Диаграмма Эйлера-Венна для методов изучения головного мозга.

## Закон ускорения отдачи в применении к мозгу



Устройства для визуализации головного мозга.

по экспоненциальному закону. Современные методы имеют разрешение порядка 4 нм, что позволяет визуализировать индивидуальные контакты.



Пространственное разрешение метода МРТ в микронах.

\* 2-Дезоксиглюкоза.

Технологии, связанные с созданием искусственного интеллекта, такие как системы распознавания разговорной речи, не обязательно должны следовать теоретическим принципам функционирования мозга. Их задача — достичь максимальной эффективности. Поэтому важно отметить, что лучшие из этих технологий соответствуют принципам, которые я описал в этой книге: это самоорганизующиеся системы иерархических модулей, способные к инвариантному и самоассоциативному распознаванию образов, обладающие избыточностью и восходящей и нисходящей предсказательной способностью. Возможности таких систем тоже изменяются экспоненциально, как показывает пример Ватсона.

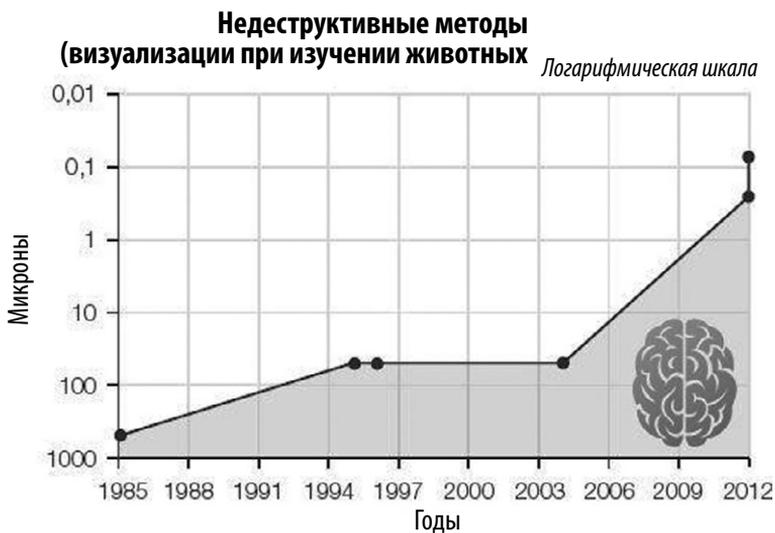


*Пространственное разрешение деструктивных методов визуализации.*

Главная цель исследований функций мозга заключается в расширении наших технических возможностей создания разумных систем. Многие исследователи в области ИИ, вероятно, недооценивают то глубокое влияние, которое оказали на их работу накопленные ранее знания о принципах работы мозга. Понимание механизмов работы мозга также помогает исправлять

## Закон ускорения отдачи в применении к мозгу

различные функциональные нарушения. Однако у проекта по обратному проектированию мозга есть и еще одна важная цель — помочь нам понять, кто мы такие.





# Глава одиннадцатая

## КРИТИКА

Если машина окажется неотличима от человека, нам придется относиться к ней с тем же уважением, что и к человеку, — нам придется признать, что у нее есть разум.

*Стивен Харнад\**

---

\* Стивен Харнад (род. в 1945) — американский психолог-когнитивист, основатель кембриджского журнала «Наука о поведении и мозге».

**Н**аиболее важным источником возражений против моего тезиса о законе ускорения отдачи и его применения к расширению человеческого разума является линейная природа человеческого мышления. Как я уже рассказывал, все сотни миллионов распознающих модулей новой коры обрабатывают информацию одновременно. Один из результатов такой организации коры заключается в том, что будущее мы можем предвидеть только линейным образом, и поэтому критики применяют линейный интуитивный подход к оценке экспоненциальных явлений.

Подобные возражения я называю «критикой из-за недоверия»: экспоненциальные прогнозы кажутся невероятными по той причине, что наша предсказательная способность имеет линейный характер. Один из создателей компании Microsoft Пол Аллен (род. в 1953 г.) и его коллега Марк Гривс недавно озвучили некоторые из подобных возражений в статье «Сингулярность еще не близка», опубликованной в журнале *Technology Review*<sup>1</sup>. Хотя здесь я буду отвечать именно на возражения Аллена, они являются типичными для критиков тех тезисов, которые я выдвигаю в отношении будущих исследований в области мозга. В заголовке своей статьи Аллен апеллирует к моей книге «Сингулярность уже близка», однако цитирует он только статью, которую я написал в 2001 г. («Закон ускорения отдачи»). Более того, в его статье не признаются и не рассматриваются аргументы, действительно изложенные мной в книге. К сожалению, это часто происходит с критиками моих работ.

Когда в 1999 г. вышла книга «Эпоха духовных машин», а за ней в 2001 г. та статья, это вызвало несколько критических замечаний следующего толка: *закону Мура приходит конец; возможности железа могут изменяться экспоненциально, но программное обеспечение отстает; мозг слишком сложен; в мозге слишком много такого, что нельзя воспроизвести с помощью компьютеров, и т. д.* Книгу «Сингулярность уже

близка» я написал в том числе для того, чтобы ответить на эти возражения.

Я не могу сказать, что Аллен и другие критики обязательно должны были поверить моим аргументам, но они, по крайней мере, могли ответить на мои рассуждения. Аллен пишет, что «закон ускорения отдачи (ЗУО)... не является физическим законом». Я же утверждаю, что большинство научных законов — не физические законы, а являются результатом проявлений множества событий на более примитивном уровне. Классический пример — законы термодинамики. Если рассмотреть математические закономерности в основе этих законов, мы увидим, что каждая частица ведет себя случайным образом, так что по определению мы не можем предсказать, где она окажется в следующий момент времени. Но при этом общие свойства газов вполне предсказуемы с большой степенью точности, в соответствии с *законами* термодинамики. И то же самое относится к закону ускорения отдачи: каждый технологический проект развивается непредсказуемым образом, но общая траектория, выраженная в терминах цены — производительности и емкости, тем не менее идет по совершенно предсказуемому закону.

Если бы компьютерные технологии развивались лишь усилиями нескольких ученых, прогресс этих технологий действительно был бы непредсказуем. Однако здесь функционирует большая динамическая система конкурирующих проектов, что определяет параметр цены — производительности (например, выраженный в объеме информации в секунду в расчете на один доллар). И процесс этот следует гладкой экспоненциальной зависимости, начиная с американской переписи населения 1890 г., о чем я писал в предыдущей главе. Хотя теоретические основы ЗУО подробно изложены в книге «Сингулярность уже близка», мной и другими исследователями проведено множество эмпирических наблюдений.

Аллен пишет: «Подобные “законы” работают лишь до тех пор, пока не перестают работать». Однако здесь он путает

отдельные парадигмы с общим направлением развития информационных технологий. Если проанализировать процесс создания радиоламп все более и более мелкого размера (одна из парадигм в рамках общего проекта развития компьютерных технологий в 1950-х гг.), мы действительно увидим, что процесс этот развивался до своего логического конца. Но когда конец приблизился, исследователи переключились на другие направления. Дальнейший экспоненциальный рост поддерживался за счет технологии транзисторов, и это привело к пятой парадигме (закону Мура) и последовательному сокращению размеров интегральных схем. Закону Мура постоянно предсказывают конец. Международный план по развитию полупроводниковой технологии (ITRS) предусматривает переход на 7-нм технологию в начале 2020-х гг.<sup>2</sup> На этом этапе ключевые элементы схем будут составлять в толщину 35 атомов углерода, и дальше сокращать их размер станет сложно. Однако компания Intel и другие производители схем уже обдумывают первые шаги шестой парадигмы (трехмерных вычислений), которая позволит поддержать экспоненциальные изменения параметров цены — производительности. По прогнозам компании Intel, трехмерные схемы станут главным направлением исследований на протяжении десятилетия; трехмерные транзисторы и 3D-память уже стали реальностью. Эта шестая парадигма будет поддерживать ЗУО в отношении цены — производительности компьютеров в следующие годы, пока компьютеры за тысячу долларов не станут в триллионы раз мощнее человеческого мозга<sup>3</sup>. (Кажется, мы сходимся с Алленом во мнении, что именно такой уровень вычислительной техники необходим для функциональной симуляции человеческого мозга<sup>4</sup>.)

Далее Аллен приводит стандартный аргумент, заключающийся в том, что развитие программного обеспечения не успевает за экспоненциальным развитием компьютерного оборудования. В книге «Сингулярность уже близка» я подробно останавливался

на этом вопросе и приводил разные методы оценки сложности и емкости программ, которые подтверждают-таки наличие экспоненциального роста<sup>5</sup>. В одном недавнем исследовании («Доклад президенту и Конгрессу относительно будущего цифровых технологий: исследования на федеральном уровне и развитие сетей и информационных технологий»), представленном Президентским экспертным советом по науке и технологии, говорится следующее.

«Еще более удивительным — и гораздо менее оцененным — фактом является то, что во многих областях *повышение производительности за счет улучшения алгоритмов намного опережает значительное повышение производительности за счет ускорения работы процессоров*. Используемые сегодня алгоритмы для распознавания речи, перевода разговорной речи, игры в шахматы и логистического планирования значительно эволюционировали за последние десять лет... Вот лишь один пример, приведенный профессором Мартином Грётшелем из Музея Конрада Цузе. Профессор Грётшель, эксперт в области оптимизации, замечает, что для решения задачи планирования в 1988 г. с помощью линейных алгоритмов и компьютеров того времени потребовалось бы 82 года. Через 15 лет, в 2003 г., ту же задачу можно было бы решить за одну минуту, что быстрее примерно в 43 млн раз. При этом примерно тысячекратное ускорение было бы достигнуто за счет повышения скорости процессоров, а фактор 43 тыс. объясняется усовершенствованием алгоритмов! Грётшель также отмечает 30 000-кратное улучшение алгоритмов смешанного целочисленного программирования в период между 1991 и 2008 гг. Создание и анализ алгоритмов, а также анализ вычислительной сложности задач являются важнейшими направлениями исследований в компьютерной области».

Обратите внимание, что упомянутое Грётшелем линейное программирование — тот математический метод, который используется для оптимального распределения ресурсов в таких

иерархических системах памяти, как обсуждавшиеся выше иерархические скрытые модели Маркова. В книге «Сингулярность уже близка» я привел множество подобных примеров<sup>6</sup>.

Что касается искусственного интеллекта, Аллен, как и многие другие критики, не признают способностей Ватсона. Многие из этих критиков практически ничего о Ватсоне не знают, кроме того, что это некая компьютерная программа (хотя и действующая на 720 процессорах). Аллен пишет, что подобные Ватсону системы «остаются хрупкими, пределы их возможностей строго определены их внутренними установками и алгоритмами, они не могут обобщать и часто дают бессмысленные ответы в тех областях, которые выходят за рамки их специфической области знаний».

Прежде всего, такое же замечание можно сделать и в отношении людей. Я также замечу, что «специфические области знаний» Ватсона включают в себя всю информацию, содержащуюся в «Википедии», а также много других баз данных, что составляет весьма широкий спектр. Ватсон владеет большим объемом человеческих знаний и способен воспринимать различные формы речи, включая каламбуры, шутки и метафоры, относящиеся практически к любой области человеческой деятельности. Он не идеален, но и люди не идеальны, однако он достаточно хорош, чтобы обыгрывать лучших игроков в «Джеопарди!»

Аллен пишет, что Ватсон был собран учеными, объединившими свои специфические знания в узких областях. Но это неправда. Хотя некоторые знания Ватсона были запрограммированы напрямую, значительную часть информации он раздобыл сам, изучая документы на человеческом языке, например «Википедию». В этом заключается его главная сила, а также в его способности понимать сложно сформулированные вопросы викторины «Джеопарди!», но знания свои он получил сам, читая литературу.

Как я уже упоминал выше, многие критики Ватсона сходятся на том, что его функция основана на определении статистической

вероятности, а не на «истинном» понимании. Многие понимают это так, что Ватсон просто статистическим образом перебирает последовательности слов. На самом деле в случае Ватсона под «статистической информацией» подразумеваются коэффициенты распределения и символические связи в таких самоорганизующихся моделях, как иерархические скрытые модели Маркова. Точно так же «статистической информацией» можно назвать распределение концентрации нейромедиаторов и избыточность образов в человеческой новой коре. На самом деле неясные вопросы мы разрешаем во многом тем же способом, что и Ватсон, — путем рассмотрения вероятности различных интерпретаций той или иной фразы.

Аллен продолжает: «Любая структура [мозга] тщательно оттачивалась на протяжении миллионов лет эволюции для выполнения конкретной функции. Он не похож на компьютер, в котором содержатся миллиарды идентичных транзисторных ячеек памяти, контролируемых центральным процессором и некоторыми другими элементами. В мозге каждая отдельная структура и каждая нервная цепочка вырабатывалась под действием эволюции и внешних факторов в индивидуальном порядке».

Идея о том, что каждая структура и каждая нервная цепочка мозга уникальна и создана по отдельному проекту, просто несостоятельна, поскольку означает, что для проектирования мозга понадобились бы сотни триллионов байт информации. Структурный план мозга (а также всех других отделов тела) содержится в геноме, и в самом мозге нет никакой дополнительной структурной информации. Замечу, что эпигенетическая информация (например, пептиды, контролирующие экспрессию генов) лишь немного дополняет объем генетической информации. Объем содержащейся в мозге информации значительно увеличивается в процессе обучения и приобретения опыта, однако то же самое можно сказать и о системах ИИ, таких как Ватсон. В книге «Сингулярность уже близка» я показал, что после обратимого сжатия

(за счет массивной избыточности генома) объем структурной информации генома составляет около 50 млн байт, и примерно половина этой информации (то есть около 25 млн байт) относится к мозгу<sup>7</sup>. Это нелегко, но это такой уровень сложности, с которым мы можем работать, и он проще, чем многие современные компьютерные системы. Кроме того, значительная часть из этих 25 млн байт генетической информации касается биологических потребностей нейронов, а не алгоритмов обработки информации.

Как же на основании лишь десятков миллионов байт структурной информации возникают 100 или 1000 трлн контактов нейронов в мозге? Понятно, что ответ кроется в значительной избыточности мозга. Руководитель исследовательской группы в компании IBM Дхармендра Модха пишет: «Нейроанатомы нашли не безнадежно спутанную, случайную сеть, полностью уникальную для мозга каждого человека, а повторяющиеся структуры внутри каждого мозга и большое сходство между разными видами... Потрясающее природное свойство реконфигурации дает надежду на то, что ключевые алгоритмы процессов в нейронах не зависят от специфической сенсорной или моторной модальности и что многие наблюдаемые вариации структуры коры отражают более тонкую настройку канонических цепей; на самом деле именно эти канонические цепи мы и хотим создать с помощью обратного проектирования»<sup>8</sup>.

Аллен указывает на неизбежность существования «тормоза, ограничивающего прогресс в понимании человеческого мозга и воспроизведении его возможностей», на основании представлений о том, что каждый из 100 или 1000 трлн контактов в человеческом мозге создан по индивидуальному плану. Тут спутаны представления о том, что есть лес, а что деревья. Если вы хотите изучить, смоделировать, симулировать и воссоздать поджелудочную железу, вам не нужно воссоздавать или симулировать каждую органеллу в каждой клетке островков поджелудочной железы. Вам нужно понять

механизм функционирования одной такой клетки и ее участие в контроле уровня инсулина, а затем распространить эти знания на группу аналогичных клеток. Для островковых клеток этот алгоритм абсолютно ясен. Сейчас проходят тестирование искусственные ткани поджелудочной железы, созданные на основе этой функциональной модели. Безусловно, в мозге значительно больше сложности и вариаций, чем в более или менее однородных клетках островков поджелудочной железы, но и здесь имеет место значительная повторяемость функций, как я уже неоднократно подчеркивал.

Критикам, придерживающимся той же точки зрения, что и Аллен, присуще свойство, которое я называю «научным пессимизмом». Исследователи, занимающиеся созданием нового поколения технологии или моделированием нового научного направления, неизбежно сталкиваются с таким огромным набором проблем, что если кто-нибудь скажет им, как будет выглядеть эта технология примерно через десять поколений, у них глаза вылезут на лоб. Один из пионеров в области создания интегральных схем недавно напомнил мне происходившую 30 лет назад борьбу за сокращение размера микросхемы от 10 до 5 микрон (от 10 000 до 5000 нм). Ученые понимали, что смогут решить эту задачу, но если бы кто-то предсказал, что однажды мы будем иметь микросхемы толщиной меньше одного микрона (1000 нм), большинство из ученых, сконцентрированных на своих сиюминутных задачах, подумали бы, что это абсолютно невозможно. Аргументами были хрупкость схем такого размера, тепловые эффекты и т. д. А сегодня Intel использует 22-нм чипы.

Примерно такие же пессимистические комментарии сопровождали и реализацию проекта «Геном человека». Проект длился около 15 лет, за первую половину этого срока был проанализирован лишь 1% генома, и критики указывали на невозможность повышения скорости секвенирования генома без нарушения тонких генетических структур. Однако

благодаря экспоненциальному росту памяти, а также показателя цены — производительности через семь лет проект был завершен. Проект по обратному проектированию человеческого мозга продвигается аналогичным образом. Например, еще совсем недавно мы получили возможность с помощью неинвазивных методов сканирования в реальном времени наблюдать за тем, как образуются и возбуждаются отдельные контакты между нейронами. Многое из того, о чем я рассказывал выше, стало возможно совсем недавно благодаря подобным достижениям.

Аллен описывает мою идею об обратном проектировании человеческого мозга просто как сканирование мозга для понимания его тонкой структуры с последующей симуляцией всего мозга «вверх дном» без детального понимания его методов обработки информации. Но я предлагаю совсем другое. Нам действительно нужно подробно изучить, как работают отдельные типы нейронов, а затем собрать информацию о соединении функциональных модулей. А дальше функциональные методы, выведенные на основе этих данных, будут направлять развитие разумных систем. Грубо говоря, мы ищем биологические методы, способные ускорить исследования в сфере ИИ, которые пока во многом продвигаются без серьезного прорыва в понимании того, как аналогичную функцию выполняет мозг. На основе собственного опыта по распознаванию речи могу сказать, что работа сильно продвинулась, когда мы поняли, как мозг подготавливает и трансформирует звуковую информацию.

Дифференцировка повторяющихся структур мозга реализуется в процессе обучения и приобретения опыта. При сегодняшнем положении дел в сфере ИИ компьютерные системы тоже могут учиться на собственном опыте. Самодвижущиеся машины Google обучаются на своем водительском опыте, а также на данных машин Google, управляемых людьми. Ватсон получил большую часть информации за счет самостоятельного

чтения. Интересно отметить, что математические принципы методов, заложенных в основу функционирования систем ИИ, очень близки соответствующим принципам функционирования новой коры.

Еще одно часто высказываемое возражение против возможности создания «сильного» ИИ (искусственного интеллекта на уровне человеческого разума и выше) заключается в том, что человеческий мозг активно использует аналоговые методы, а цифровые методы не могут воспроизводить плавные изменения параметров, доступные для воспроизведения аналоговыми методами. Это верно, что с помощью одного бита информации нельзя описать сложную зависимость, однако многобитные слова легко отражают множество значений, причем с любой степенью точности. Это постоянно происходит в цифровых компьютерах. А вот точность аналоговой информации в мозге (например, синаптический потенциал) составляет лишь один уровень из 256 возможных, которые могут быть представлены 8 битами информации.

В девятой главе я приводил доводы Роджера Пенроуза и Стюарта Хамероффа, касающиеся микротрубочек и квантовых вычислений. Вспомните, они утверждают, что микротрубочки нейронов осуществляют квантовые вычисления, а в компьютерных системах этого добиться нельзя. В этом заключается принципиальное отличие и, по-видимому, превосходство человеческого мозга. Я прокомментировал, что пока нет никаких доказательств того, что микротрубочки нейронов осуществляют квантовые вычисления. Человек на самом деле очень плохо решает такие задачи, которые не представляют никакой сложности для квантовых компьютеров (например, факторизация больших чисел). Но, если идеи Пенроуза и Хамероффа окажутся верны, ничто не мешает нам использовать квантовые вычисления в наших компьютерах.

Джон Серль знаменит, в частности, тем, что придумал мысленный эксперимент, называемый «китайская комната»

(я подробно описывал его в книге «Сингулярность уже близка»<sup>9</sup>). Если говорить кратко, речь идет о человеке, который в письменном виде получает вопросы на китайском языке и отвечает на них. Для этого он пользуется сложной книгой правил. Серль утверждает, что человек не знает китайского языка и действует «неосознанно» (то есть не понимает смысла ни вопроса, ни ответа), хотя и отвечает на вопросы по-китайски. Серль сравнивает эту ситуацию с компьютером и приходит к выводу, что компьютер, который отвечает на вопросы по-китайски (и проходит тест Тьюринга на китайском языке), как и человек в китайской комнате, по-настоящему не понимает языка и не осознаёт, что делает.

В аргументах Серля есть несколько философских трюков. Например, человека из мысленного эксперимента сравнивают только с центральным процессором компьютера. Можно сказать, что центральный процессор не имеет реального представления о том, что он делает, но это лишь часть общей системы. В китайской комнате всю систему составляют человек и его книга правил. И эта система в целом понимает китайский язык, иначе она не могла бы убедительно отвечать на вопросы по-китайски, а это нарушило бы условие мысленного эксперимента Серля.

Притягательность этого эксперимента связана с тем, что сегодня нам трудно приписать компьютерной программе такие атрибуты, как истинное понимание и сознание. Однако слабость аргумента в том, что точно такие же рассуждения можно применить собственно к человеческому мозгу. Каждый распознающий модуль (вообще говоря, каждый нейрон и каждый его компонент) следует определенному алгоритму. Здесь действуют молекулярные механизмы, подчиняющиеся природным законам. Если считать, что следование алгоритму — процесс неосознанный и не требующий реального понимания, придется заключить, что человеческий мозг тоже не обладает этими способностями. Используйте аргумент Серля и замените слова

«манипуляция символами» словами «манипуляция межнейронными контактами и синаптическими потенциалами», и вы получите убедительный аргумент в пользу того, что человеческий мозг по-настоящему ничего не понимает.

Еще одна линия аргументации связана с сущностью самой природы, которая является святыней для многих наблюдателей. Например, новозеландский биолог Майкл Дентон (род. в 1943 г.) усматривает глубокое различие между принципами строения машин и биологических систем. Дентон характеризует биологические существа как «самоорганизующиеся, самореплицирующиеся, взаимовыгодные, самообучающиеся и целостные»<sup>10</sup>. Он утверждает, что такие биологические формы могут быть созданы только в биологических процессах, являются «неизменной... недостижимой и... фундаментальной» реальностью бытия и поэтому относятся принципиально к другой философской категории, нежели машины.

Однако реальность, как мы видели, такова, что машины могут быть созданы на основе тех же принципов. Изучение специфики структуры самого разумного создания природы — человеческого мозга — как раз и является задачей исследований в области обратного проектирования мозга. Кроме того, неверно, что биологические системы полностью «целостны», как утверждает Дентон. Впрочем, было бы ошибкой заявить, что машины обязательно должны иметь модульное строение. В природных системах, особенно таких, как мозг, мы с очевидностью идентифицируем иерархические функциональные структуры. И системы ИИ используют аналогичные методы.

Мне кажется, многие критики не удовлетворятся до тех пор, пока компьютеры не смогут в обычном порядке проходить тест Тьюринга, но, возможно, даже этого окажется недостаточно. Наверняка начнут возникать сомнения в пригодности соответствующих тестов Тьюринга. Возможно, сам я буду среди критиков, отвергающих подобные ранние заявления. К тому моменту, когда появятся доказательства способности компьютеров

## Введение

проходить тест Тьюринга, компьютеры уже долгое время будут превосходить невооруженный человеческий разум.

Я делаю акцент на слове «невооруженный», поскольку вооружение, усиление человеческого разума — именно та цель, которую мы преследуем при создании этих «детей разума», как называют компьютеры Ганс Моравек<sup>11</sup>. Сочетание распознавания образов на уровне человеческих возможностей со скоростью и точностью компьютеров приведет к созданию очень мощных машин. Однако это не будет нашествием вражеского марсианского разума: мы создаем эти машины, чтобы сделать умнее самих себя. Я верю, что большинство людей согласятся со мной, что именно в этом заключается уникальная способность человеческого существа: мы создаем эти машины, чтобы расширить наши собственные возможности.

## ЭПИЛОГ

Картина довольно безрадостная, джентльмены... Климат на планете меняется, верх берут млекопитающие, а у нас с вами мозг размером с грецкий орех.

*Разговор динозавров на карикатуре Гари Ларсона\* из серии «Дальняя сторона»*

Мозг — это мышца  
из деятельных холмов,  
борьба необдуманных вещей с теми,  
о которых думают бесконечно.

*Джойс Кэрол Оутс\*\*, из книги *Grave Dwellers in Love and Its Derangements*, 1970*

---

\* Гари Ларсон (род. в 1950) — известный американский карикатурист.

\*\* Джойс Кэрол Оутс (род. в 1938) — американская писательница, поэт, прозаик и критик.

**Р**азум можно определить как способность разрешать проблемы при наличии ограниченных ресурсов, причем одним из ключевых ресурсов является время. Таким образом, способность быстрее разрешать такие вопросы, как поиск пищи или побег от хищника, отражает наличие более сильного разума. Разум возник потому, что помогал выжить. Это кажется очевидным, но не все с этим соглашаются. Нам как виду разум позволил не только занять доминирующее положение на планете, но и планомерно улучшать качество нашей жизни. С этим последним тезисом тоже не все соглашаются, учитывая широко распространенное мнение, что жизнь теперь только ухудшается. Например, по данным опроса, проведенного американской корпорацией Gallup Inc., опубликованным 4 мая 2011 г., «только 44% американцев верят в то, что жизнь современной молодежи будет лучше, чем у их родителей»<sup>1</sup>.

Но если смотреть на вещи шире, мы обнаружим, что не только средняя продолжительность жизни людей выросла в четыре раза за последнее тысячелетие (и удвоилась за последние 200 лет)<sup>2</sup>, но и годовой валовой доход на душу населения (в долларовом эквиваленте) увеличился от нескольких сотен долларов в 1800 г. до тысяч долларов в наши дни, причем в развитых странах эта тенденция просматривается еще более отчетливо. Сто лет назад существовало лишь несколько демократических государств, а сейчас это уже обычное явление. Чтобы оценить прогресс человечества в исторической перспективе, я советую прочесть книгу «Левиафан» Томаса Гоббса (1651), в которой он описывает жизнь человека как «одинокую, бедную, мерзкую, грубую и короткую». Взгляд современного человека нашел отражение в книге «Изобилие» (2012), написанной основателем фонда X-Prize Foundation (и соучредителем вместе со мной Университета Сингулярности в Кремниевой долине) Питером Дамандисом и научным писателем Стивеном Котлером. Книга рассказывает об удивительном и постоянном улучшении человеческой жизни по самым разным направлениям. В книге Стивена

Пинкера «Лучшие стороны нашей натуры: почему насилия становится все меньше» (2012) досконально отражено планомерное улучшение отношений между отдельными людьми и отдельными народами. Американский юрист, предприниматель и писательница Мартина Ротблат (род. в 1954 г.) отмечает расширение прав граждан, стоит лишь упомянуть однополые браки, которые еще несколько десятилетий назад нигде не признавались, а сегодня юридически законны во многих странах<sup>3</sup>.

Основная причина того, что люди чувствуют понижение качества жизни, связана с постоянным улучшением качества информированности населения о существующих проблемах. Если сегодня где-то в мире идет война, мы получаем об этом такое количество информации, как будто сами там находимся. Во время Второй мировой войны десятки тысяч людей погибали в боях, но их современники могли узнать об этом лишь из фрагментов кинохроники в кинотеатре несколько недель спустя. Во время Первой мировой войны лишь избранной публике могло стать известно о развитии конфликта из газет (без фотографий). А в XIX в. доступ к новостям в реальном времени вообще практически отсутствовал.

Прогресс нашего вида, достигнутый благодаря нашему разуму, отражен в эволюции знаний, включающих технологию и культуру. Множество технологий сегодня становятся информационными технологиями, и этот процесс неизбежно будет продолжаться в соответствии с экспоненциальным законом. Именно благодаря им мы можем решать важнейшие проблемы человечества, такие как сохранение окружающей среды, обеспечение ресурсами растущее население планеты (включая энергию, продукты питания и воду), преодоление болезней, значительное повышение продолжительности жизни и устранение бедности. И только усилив самих себя разумными технологиями, мы сможем справиться со столь сложными задачами.

Не нужно воспринимать эти технологии как авангард разумной вражеской армии, которая победит и уничтожит нас самих.

С тех самых пор, как мы взяли в руки палку, чтобы дотянуться до высокой ветки, мы продолжаем использовать орудия труда, чтобы облегчить себе жизнь — как в физическом, так и в мыслительном плане. Тот факт, что сегодня мы можем достать из кармана прибор, с помощью которого получаем доступ к обширным человеческим знаниям путем нажатия на несколько клавиш, продвигает нас далеко за пределы воображения большинства людей, живших всего несколько десятилетий назад. «Мобильный телефон» (в кавычках, поскольку это гораздо больше, чем просто телефон) в моем кармане в миллион раз дешевле и в тысячи раз мощнее компьютера, на котором работали все студенты и профессора Массачусетского технологического института, когда я там учился. Это соответствует увеличению показателя цена — производительность в несколько миллиардов раз за 40 лет, и продолжение этой тенденции мы будем наблюдать еще в течение 25 лет, а потом то, что когда-то устанавливали в зданиях, а позднее носили в кармане, будет помещаться в одной клетке крови.

И так мы станем единым целым с той разумной технологией, которую сейчас создаем. Разумные нанороботы в нашей системе кровообращения будут поддерживать физическое здоровье нашего тела на клеточном и молекулярном уровне. Они неинвазивным путем через капилляры станут попадать в наш мозг, взаимодействовать с нашими нейронами и напрямую расширять наш разум. И это совсем не так фантастично, как может показаться. Уже существуют устройства размером с клетку крови, которые помогают излечивать животных от диабета I типа или находить и разрушать в крови опухолевые клетки. Учитывая действие закона ускорения отдачи, через 30 лет эти технологии будут в миллиарды раз мощнее, чем сейчас.

Я уже воспринимаю используемые мной устройства и облако вычислительных ресурсов, с которым они соединены виртуальным образом, как продолжение меня самого, и если меня отсоединить от этих расширений мозга, я буду чувствовать себя неполноценным. Вот почему так заметна была однодневная

забастовка 18 января 2012 г. против законопроекта SOPA (Stop Online Piracy Act), в которой участвовали Google, «Википедия» и тысячи других сайтов. Я чувствовал, что часть моего мозга тоже задействована в забастовке (хотя и я, и другие люди нашли способы доступа к этим ресурсам). Это, кстати, была весьма яркая демонстрация политической силы тех сайтов, поскольку данный законопроект не передали на ратификацию, а просто уничтожили. Еще важнее то, что это событие показало, как сильно мы продвинулись в передаче облаку различных частей нашего мышления. Оно стало частью нас самих. И поскольку в нашем мозге уже имеется небиологический разум, этот процесс будет нарастать экспоненциально.

Разум, который мы создадим путем обратного проектирования мозга, получит доступ к собственному исходному коду и будет способен быстро самосовершенствоваться путем запуска ускоренного цикла итераций. Хотя, как мы видели, биологический мозг обладает значительной пластичностью, он также характеризуется фиксированной архитектурой, которую практически нельзя модифицировать, и ограниченной памятью. Повысить количество распознающих модулей от 300 до, скажем, 400 млн мы можем только небиологическим способом. Но, когда мы научимся это делать, у нас не будет оснований останавливаться на достигнутом. Мы сможем двигаться дальше и создать миллиард распознающих модулей или триллион.

Количественные изменения переходят в качественные. Самые важные эволюционные достижения *Homo sapiens* были количественными: у нас появился высокий лоб, вмещающий новую кору значительно большего размера. Повышенная емкость новой коры позволила нашему виду создавать и обдумывать более сложные идеи, что положило начало развитию разнообразных направлений науки и искусства. Дополнительное увеличение объема коры за счет небиологического субстрата позволит достичь качественно нового, еще более высокого уровня абстракции.

Британский математик и коллега Тьюринга Ирвинг Дж. Гуд в 1965 г. писал следующее: «Первая сверхразумная машина — это последнее изобретение, которое должен сделать человек». Он считал, что такая машина сможет превзойти по «мыслительной активности разумного человека», и пришел к выводу, что, «поскольку конструирование машин как раз и есть один из видов такой мыслительной активности, сверхразумная машина сможет создать еще более совершенные машины и это, несомненно, приведет к “взрыву разума”».

Последнее изобретение, которое должна была сделать биологическая эволюция, — новая кора — неизбежно привело к последнему изобретению человечества — поистине разумному компьютеру, причем строение первой послужило прообразом для второго. Биологическая эволюция продолжается, но технологическая эволюция происходит в миллионы раз быстрее. В соответствии с законом ускорения отдачи к концу этого столетия мы достигнем предела возможностей компьютерного вычисления, определяемого физическими законами<sup>4</sup>. Эту высокоорганизованную систему материи и энергии можно назвать «компьютроном», и он по удельной мощности многократно превосходит человеческий мозг. Но это будет не просто вычислительная система, она окажется снабжена разумными алгоритмами, совмещающими все человеческие и машинные знания. Со временем мы конвертируем в компьютерон значительную часть доступной материи и энергии из нашего уголка Галактики. Затем, учитывая закон ускорения отдачи, мы проникнем в другие галактики и другие уголки Вселенной.

Если скорость света действительно останется непреодолимым препятствием, заселение Вселенной будет долгим делом, учитывая, что самая близкая к нам звездная система находится на расстоянии четырех световых лет. Если появится хоть какая-то надежда обойти этот барьер, наш разум и наши технологии будут достаточно мощными, чтобы преодолеть это расстояние. Вот почему такой важной новостью стало недавнее сообщение

о том, что мюоны преодолели расстояние в 730 км между ускорителем в ЦЕРНе (на границе Франции и Швейцарии) и лабораторией Гран-Сассо (в центре Италии) со скоростью выше скорости света. Данное сообщение оказалось ошибкой, но существуют другие возможности решить эту задачу. Нам даже не нужно преодолевать этот предел, если мы найдем короткие пути к другим отдаленным местам, используя иные измерения, кроме знакомых нам сегодня трех измерений. Изучение возможности преодоления скорости света или поиск других способов обойти этот барьер будут важнейшими проблемами человеческо-машинной цивилизации в начале следующего столетия.

Космологи обсуждают, каким будет конец света — закончится ли мир в огне (Большой взрыв в начале — Большой треск в конце) или во льду (смерть звезд в результате бесконечного расширения), но они не учитывают силу разума, как будто появление его было лишь побочным эффектом великой небесной механики, управляющей законами Вселенной. Сколько времени нам понадобится, чтобы распространить свой разум в небиологической форме по всей Вселенной? Если нам удастся преодолеть скорость света (безусловно, это очень серьезное «если»), например, путем использования кротовых нор\* (существование которых вполне совместимо с нашим современным пониманием законов физики), на это уйдет несколько столетий. В противном случае гораздо больше. Но при любом варианте развития событий разбудить Вселенную и разумно распорядиться ее судьбой с помощью нашего человеческого разума в его небиологической форме — это наше предназначение.

---

\* Кротовая нора — гипотетическая топологическая особенность пространства-времени, представляющая собой туннель между двумя связанными между собой или несвязанными областями пространства.

# КОММЕНТАРИИ

## Введение

<sup>1</sup> Вот одно предложение из романа Габриэля Гарсиа Маркеса «Сто лет одиночества».

Аурелиано Второй поначалу не заметил этот кошачий вокализ, но на следующее утро после завтрака почувствовал, что его уши сверлит какое-то жужжание, более переливчатое и пискливое, чем дробь дождя, а это, оказывается, Фернанда бродила по дому, громко причитая — для того, мол, ее воспитывали, как королеву, чтобы стать ей прислугой в сумасшедшем доме, жить с мужем — лодырем, безбожником, развратником, который только и знает, что валяется на кровати брюхом кверху и ждет не дождетс манны небесной, а она работает-надрывается, везет на себе хозяйство, которое на глазах рушится, где столько всего надо сделать, столько на своем горбу вывезти, столько дыр залатать каждый Божий день с утра и до вечера, что, как ляжет она в постель, так в глазах рябит и мутится, и никто ведь потом не скажет: доброе утро, Фернанда, хорошо ли ты отдохнула, Фернанда, никто не спросит, хотя бы из вежливости, почему, мол, ты со сна такая бледная, с черными кругами под глазами, хотя, конечно, она и не ждет таких слов от этой семьи, где в общем-то всегда всем она была в тягость, чуть ли не ноги об нее вытирали, как на пугало огородное на нее глядели, по углам шушукались и злословили, называя ее ханжой, называя ее лицемеркой, называя ее хитрой бестией, и даже Амаранта — царствие ей небесное — не постеснялась громко сказать, что она, Фернанда, из тех, кто посты соблюдает не для духа, а для брюха — Господи, что за выражение, — а она все это покорно сносила по воле Божьей, но больше терпеть не будет, после того как этот мерзавец Хосе Аркадио Второй сказал, что семья стала гибнуть из-за того, что впустила в дом самодурку, — только послушать! — своевольную самодурку, прости, Господи, чуть ли не изуверку из породы тех гнусных изуверов, которых правительство посылает убивать рабочих, и — скажите на милость — он при этом имел в виду не кого-нибудь, а ее, крестницу герцога Альбы, даму такого

знатного происхождения, что жены президентов зеленели от зависти, дочь такого древнего рода, что имеет право подписываться одиннадцатью испанскими фамилиями в ряд, и вообще она единственная смертная в этом нечестивом городишке, которая не опростоволосится, если надо накрыть стол на шестнадцать персон, хотя бы потом этот забулдыга, ее муж, и говорил, корчась от смеха, что столько ложек и вилок, ножей и чайных ложечек нормальным людям не требуется, разве что сороконожкам, а ведь она тут одна-единственная, кто во сне может ответить, когда положено белое вино подавать, и с какой стороны, и в какой бокал наливать, а когда — красное, и с какой стороны, и в какой бокал, не то что эта недоучка Амаранта, царствие ей небесное, которая думала, что белое вино пьют днем, а красное вечером, и она, Фернанда, единственная на всем побережье может похвастаться тем, что пользуется ночью не чем иным, как золотым горшком, хотя полковник Аурелиано Буэндия, царствие ему небесное, имел наглость спросить, ехидный франкмасон, уж не за то ли ей такая привилегия, что из нее не говно, а хризантемы лезут, подумать только, так и сказал, а Рената, ее родная дочь, которая не постеснялась подглядеть, как она в спальне по-большому делает, поддакивает, что горшок и вправду из золота и весь в гербах, но внутри — говно как говно, дерьмо человечесь, да еще и похуже, потому как это — дерьмо изуверское — подумайте, родная-то дочь! — вот почему и знает она, Фернанда, цену истинную членам этой семьи, но в любом случае имеет право ожидать от мужа большего уважения, ибо, худо-бедно, он ей муж, Богом посланный, сам ее взявший в дом, законный насильник, который по своей и Божьей воле возложил на себя ответственность за то, что лишил ее родительского крова, где жила она без забот и хлопот, плела венки погребальные ради собственного удовольствия, ибо ее крестный отец прислал письмо за собственной подписью и скрепленное сургучной печатью своего перстня, дабы предупредить, что ручки его крестницы не предназначены для дел мира сего, разве только для игры на клавинодах, тем не менее ее муж-болван увез ее из дому, невзирая на все предупреждения и предостережения, и бросил в этот адский котел, где от жары задохнуться впору, и, не дождавшись конца Великого поста, уже отправился из дому со своими бродячими сундуками и со своим дурацким аккордеоном прелюбодействовать с потаскухой, у которой достаточно посмотреть на задницу — ладно, слово уже сказано, — достаточно посмотреть, как

она вертит своим кобылым крупом, чтобы узнать, кто она такая: полная противоположность ей, Фернанде, которая всегда остается дамой — во дворце или в свинарнике, за столом или в постели, прирожденной дамой, почитающей Господа своего, послушной его законам и покорной его предначертаниям, и с которой, конечно, нельзя вытворять всякие гнусные штучки, привычные для той, другой, готовой на все мерзости, как французские шлюхи, да она похуже и этих, ведь, если подумать, они все же имеют совесть вешать красный фонарик на двери, а подобные гадости, будьте уверены, не сделаешь с единственной возлюбленной дочерью доньи Ренаты Арготе и дона Фернандо дель Карпио, и прежде всего этого гранда, этого, без сомнения, святого человека, истинного христианина, кавалера ордена Гроба Господня, из тех, кто Божьей милостью не подвержен тлену могильному и чья кожа сохраняет свежесть и лоск атласного платья невесты, а глаза — живость и чистоту изумрудов...\*

- <sup>2</sup> Более подробно этот аргумент обсуждается в книге «Сингулярность уже близка»: глава 6, раздел «[The Impact... ] on the Intelligent Destiny of the Cosmos: Why We Are Probably Alone in the Universe»; Ray Kurzweil, *The Singularity Is Near*, 2005, Viking.
- <sup>3</sup> James D. Watson, *Discovering the Brain*, National Academy Press, 1992.
- <sup>4</sup> Sebastian Seung, *Connectome, How the Brain's Wiring Makes Us Who We Are*, Houghton Mifflin Harcourt, 2012; в русском переводе: Сеунг С. Коннектом: Как мозг делает нас тем, что мы есть. М. : Бином. Лаборатория знаний, 2014.
- <sup>5</sup> Множество Мандельброта, Zoom-видео: <http://www.youtube.com/watch?v=gEw8xpb1aRA> и [http://www.youtube.com/watch?v=G\\_GBwuYuOOs](http://www.youtube.com/watch?v=G_GBwuYuOOs)

## Глава первая.

### Мысленные эксперименты надмиром

- <sup>1</sup> Charles Darwin, *The Origin of Species*, P. F. Collier & Son, 1909, 185/95–96.
- <sup>2</sup> Charles Darwin, *On the Origin of Species*, 751. Peckham's Variorum edition. Morse Peckham (ed.) (1959), *The Origin of Species By Charles Darwin: A Variorum Text* (Philadelphia).

---

\* Перевод с испанского М. Былинкиной.

- <sup>3</sup> Dahm R., *Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Hum. Genet.*, 2008,122 (6): 565–81.
- <sup>4</sup> Valery N. Soyfer, *The consequences of political dictatorship for Russian science. Nature Reviews Genetics*, 2001, 2 (9): 723–729.
- <sup>5</sup> Watson J.D. and Crick F.H.C. *A Structure for Deoxyribose Nucleic Acid. Nature*, 1953, 171 (4356): 737–738; <http://www.nature.com/nature/dna50/watsoncrick.pdf>.  
Double Helix: 50 Years of DNA, Nature Archives, <http://www.nature.com/nature/dna50/archive.html>.
- <sup>6</sup> Франклин умерла в 1958 г., а Нобелевскую премию за открытие ДНК вручали в 1962 г. Существуют разные мнения относительно того, получила бы ее Франклин или нет, если бы была жива.
- <sup>7</sup> Альберт Эйнштейн, «К электродинамике движущихся тел», 1905. В этой статье была сформулирована Специальная теория относительности. См. Lindsay, Robert Bruce; Margenau, Henry (1981). *Foundations of physics*. Ox Bow Press. p. 330.
- <sup>8</sup> См. [https://ru.wikipedia.org/wiki/Радиометр\\_Крукса](https://ru.wikipedia.org/wiki/Радиометр_Крукса).
- <sup>9</sup> Заметим, что часть импульса фотонов передается молекулам воздуха (поскольку в колбе находится разреженный воздух, а не абсолютный вакуум), а затем от нагретых молекул воздуха — к лопастям.
- <sup>10</sup> Альберт Эйнштейн, «Зависит ли инерция тела от содержащейся в нем энергии?» В этой статье была представлена знаменитая формула Эйнштейна  $E = mc^2$ .
- <sup>11</sup> Письма Альберта Эйнштейна президенту Франклину Делано Рузвельту, см. <http://hypertextbook.com/eworld/einstein.shtml>.

### **Глава третья. Модель новой коры; теория мысленного распознавания образов**

- <sup>1</sup> Считается, что некоторые не относящиеся к млекопитающим животные, такие как вороны, попугаи и осьминоги, тоже в определенной степени способны размышлять, однако эта их способность неразвита и недостаточна для того, чтобы создавать орудия. У этих животных произошла адаптация других отделов мозга, осуществляющих некоторые этапы иерархического мышления, но для неограниченного иерархического мышления, на которое способен человек, нужна новая кора.
- <sup>2</sup> Mountcastle, V. B. (1978), *An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System*, in Gerald M. Edelman and Vernon B. Mountcastle, *The Mindful Brain*, MIT Press, 1982.

- <sup>3</sup> Herbert A. Simon, *The Organization of Complex Systems. Hierarchy Theory, The Challenge of Complex Systems*, Edited by Howard H. Pattee, George Braziller, Inc., New York, 1973; <http://blog.santafe.edu/wp-content/uploads/2009/03/simon1973.pdf>.
- <sup>4</sup> Marc D. Hauser, Noam Chomsky and W. Tecumseh Fitch. *The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?* *Science*. November, 2002, Volume 298, 1569–1579; [www.wjh.harvard.edu/~mnkylab/publications/language/hauser,Chomsky,Fitch.pdf](http://www.wjh.harvard.edu/~mnkylab/publications/language/hauser_chomsky_fitch.pdf).
- <sup>5</sup> Ниже представлен фрагмент книги Рэя Курцвейла и Терри Гроссмана «*Transcend, Nine Steps to Living Well Forever*», Rodale, 2009, в котором подробнее описана данная техника.

Я придумал способ решать задачи во сне. Я несколько десятилетий отработывал этот метод и обнаружил некоторые детали, которые позволяют ему функционировать лучше.

Начинаю я с того, что перед тем, как идти спать, формулирую для себя интересующую меня проблему. Это может быть проблема любого рода: математическая задача, элементы моих разработок, деловые вопросы и даже отношения с людьми.

Я думаю об этом несколько минут, но не пытаюсь искать решение. Это только прервет творческий процесс поиска. Я пытаюсь обдумывать. Что мне действительно известно? В какой форме должен быть ответ? А потом я иду спать. Эта процедура заставляет мое подсознание работать в заданном направлении.

**Терри:** Зигмунд Фрейд писал, что, когда мы видим сны, многие сенсорные центры в нашем мозге расслабляются, поэтому во сне мы можем видеть вещи, в реальности являющиеся социальными, культурными и даже сексуальными табу. Нам могут сниться таинственные вещи, о которых мы не можем позволить себе думать днем. Это одна из причин странности снов.

**Рэй:** Существуют также профессиональные ограничения, мешающие творческому движению мысли, многие из которых сформировались еще в процессе обучения, такие блоки как: «проблему обработки сигнала нельзя решить этим способом» или «лингвистика не допускает использования этих правил». Данные постулаты во сне тоже ослабевают, а потому я могу думать о новых путях решения проблемы, не будучи ограничен «дневными рамками».

**Терри:** Кроме того, когда мы спим, отключается часть нашего мозга, ответственная за оценку разумности наших идей. Это еще одна

причина, почему во сне с нами происходят таинственные и фантастические вещи. Если слон проходит через стену, мы не удивляемся, а просто отмечаем про себя: «О'кей, слон проходит через стену, ничего особенного». Если я просыпаюсь среди ночи, я часто обнаруживаю, что обдумывал стоящую передо мной проблему в очень странном и нестандартном ракурсе.

**Рэй:** Следующий этап наступает утром, на границе между сном и бодрствованием. Эту фазу сна часто называют *осознанными сновидениями*. В этом состоянии я по-прежнему сохраняю ощущения и образы из моего сна, но теперь уже пользуюсь рациональным мышлением. Я, например, понимаю, что лежу в постели, и могу сформулировать рациональную мысль о том, что у меня много дел и поэтому мне лучше бы встать. Но это было бы ошибкой. Если я могу, я остаюсь в постели и сохраняю состояние осознанных сновидений, поскольку именно в нем состоит суть творческого метода решения проблем. Кстати, если звонит будильник, метод не работает.

**Читатель:** Звучит прекрасно.

**Рэй:** Так и есть. Я по-прежнему имею доступ к снам о тех проблемах, на которые настроил себя накануне. Но при этом нахожусь в сознании и способен рационально мыслить, чтобы оценить новые творческие идеи, пришедшие мне на ум ночью. Я могу определить, какие из них осмыслены. Примерно через 20 минут я начинаю видеть свою проблему по-другому.

Этот метод помог мне сделать несколько изобретений (и потратить остаток дня на написание патентных заявок), спланировать материал для книг и выработать полезные идеи для решения самых разных проблем. Если мне нужно принять важное решение, я всегда прибегаю к этому способу рассуждений и весьма доверяю принятым решениям.

Важнейший момент здесь — дать развиваться вашей мысли, не судить ее преждевременно и не думать о том, хорошо ли работает метод. Этот метод противоположен методу ментальной дисциплины. Подумайте о проблеме, но затем во сне позвольте идеям захватить вас. А утром вновь включите разум и пересмотрите пришедшие вам во сне странные идеи. Для меня лично это бесценный метод использования творческого начала из моих снов.

**Читатель:** Отлично, такие трудоголики, как мы, теперь могут работать и во сне. Однако я не уверен, что моей жене это понравится.

**Рэй:** А вы можете считать, что теперь ваши сны делают за вас вашу работу.

## Глава четвертая. Биологическая новая кора

- <sup>1</sup> Steven Pinker, *How the Mind Works*, Norton Press, 1997, 152–153.
- <sup>2</sup> D. O. Hebb, (1949). *The organization of behavior*. New York: Wiley & Sons.
- <sup>3</sup> Henry Markram and Rodrigo Perrin, Innate Neural Assemblies for Lego Memory, *Frontiers in Neural Circuits*, 2011: 5: 6.
- <sup>4</sup> Электронное сообщение от Генри Маркрама, 19 февраля 2012 г.
- <sup>5</sup> Van J. Wedeen, Douglas L. Rosene, Ruopeng Wang, Guangping Dai, Farzad Mortazavi, Patric Hagmann, Jon H. Kaas, Wen-Yih I. Tseng, The Geometric Structure of the Brain Fiber Pathways. *Science*, March 30, 2012. Volume 335.
- <sup>6</sup> Tai Sing Lee, Computations in the early visual cortex, *Journal of Physiology — Paris* 97 (2003) 121–139.
- <sup>7</sup> Список статей можно найти здесь: [http://cbcl.mit.edu/people/poggio/tpcv\\_short\\_pubs.pdf](http://cbcl.mit.edu/people/poggio/tpcv_short_pubs.pdf).
- <sup>8</sup> Daniel J. Felleman and David C. Van Essen, Distributed Hierarchical Processing in the Primate Cerebral Cortex, *Cerebral Cortex*, January/February 1991, 1: 1–47.  
Подробный анализ байесовской математики восходящих и нисходящих контактов в новой коре представлен в статье Tai Sing Lee, Hierarchical Bayesian inference in the visual cortex, *Journal of the Optical Society of America*, 2003.
- <sup>9</sup> Uri Hasson, Eunice Yang, Ignacio Vallines, David J. Heeger, and Nava Rubin, A Hierarchy of Temporal Receptive Windows in Human Cortex, *The Journal of Neuroscience*, March 5, 2008.
- <sup>10</sup> Marina Bedny, Alvaro Pascual-Leone, David Dodell-Feder, Evelina Fedorenko, and Rebecca Saxe, Language Processing in the occipital cortex of congenitally blind adults, *Proceedings of the National Academy of Sciences*, 2011, Volume 108, no. 11, 4429–4434.
- <sup>11</sup> Daniel E. Feldman, Synaptic Mechanisms for Plasticity in Neocortex, *Annual Review of Neuroscience* 2009, 32: 33–55.
- <sup>12</sup> Aaron C. Koralek, Xin Jin, John D. Long II, Rui M. Costa, and Jose M. Carmena, Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills, *Nature*, 2012, 483, 331–335.
- <sup>13</sup> Электронное сообщение, январь 2012 г.
- <sup>14</sup> Min Fu, Xinzhu Yu, Ju Lu, Yi Zuo. Repetitive motor learning induces coordinated formation of clustered dendritic spines in vivo. *Nature*, 2012; *Nature* 483, 92–95.
- <sup>15</sup> Dario Bonanomi et al., Ret Is a Multifunctional Coreceptor that Integrates Diffusible- and Contact-Axon Guidance Signals, *Cell*, 2012,148(3): 568–582.

## Глава пятая. Старый мозг

- <sup>1</sup> Vernon B. Mountcastle, *Johns Hopkins Medical Journal*, 1975, 136: 131.
- <sup>2</sup> B. Roska and F. Werblin, Vertical Interactions Across Ten Parallel, Stacked Representations in the Mammalian Retina, *Nature* 2001, 410: 583–87; Журнал Университета Калифорнии, Беркли, Eye Strips Images of All But Bare Essentials Before Sending Visual Information to Brain, UC Berkeley Research Shows, March 28, 2001, [www.berkeley.edu/news/media/releases/2001/03/28\\_wers1.html](http://www.berkeley.edu/news/media/releases/2001/03/28_wers1.html).
- <sup>3</sup> Lloyd Watts, Reverse-Engineering the Human Auditory Pathway. J. Liu et. Al. (eds.), 47–59, 2012, Springer-Verlag, Berlin.  
Lloyd Watts, Real-Time, High-Resolution Simulation of the Auditory Pathway, with Application to Cell-Phone Noise Reduction. 2010, IEEE, 3821–3824.  
Другие статьи можно найти на сайте <http://www.lloydwatts.com/publications.html>.
- <sup>4</sup> Sandra Blakeslee, Humanity? Maybe It's All in the Wiring, *The New York Times*, December 11, 2003, <http://www.nytimes.com/2003/12/09/science/09BRAI.html>.
- <sup>5</sup> T. E. J. Behrens, H. Johnsen-Berg, M. W. Woolrich, S. M. Smith, C. A. M. Wheeler-Kingshott, P. A. Boulby, G. J. Barker, E. L. Sillery, K. Sheehan, O. Ciccarelli, A. J. Thompson, J. M. Brady and P. M. Matthews, Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, vol. 6 (7), 2003.
- <sup>6</sup> Timothy J. Buschman, Markus Siegel, Jefferson E. Roy, Earl K. Miller. Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences*, 2011, vol. 108 (27): 11252–11255.
- <sup>7</sup> Theodore W. Berger, Robert E. Hampson, Don Song, Anushka Goonawardena, Vasilis Z. Marmarelis and Sam A. Deadwyler, A cortical neural prosthesis for restoring and enhancing memory, *Journal of Neural Engineering*, 2011, vol. 8 (4).
- <sup>8</sup> Базисные функции — это нелинейные функции, которые можно комбинировать линейным способом (путем суммирования нескольких функций) для аппроксимации любой нелинейной функции.  
Pouget and Snyder, Computational approaches to sensorimotor transformations, *Nature Neuroscience* 2000, 3 (11): 1192–1198.
- <sup>9</sup> Bloedel, J. R. Functional heterogeneity with structural homogeneity: How does the cerebellum operate? *Behav. Brain Sci.* 1992, vol. 15, 666–678.
- <sup>10</sup> S. Grossberg and R. W. Paine, A Neural Model of Cortico-Cerebellar Interactions During Attentive Imitation and Predictive Learning of

- Sequential Handwriting Movements, *Neural Networks*, 2000, 13(8–9): 999–1046.
- <sup>11</sup> Javier F. Medina and Michael D. Mauk, Computer simulation of cerebellar information processing. *Nature Neuroscience Supplement*. Volume 3, November 2000.
- <sup>12</sup> Olds, James, Pleasure centers in the brain. *Scientific American*. 1956, 105–116.
- Routtenberg, Aryeh, The reward system of the brain. *Scientific American*. 1978, 154–164.
- Berridge, K. C., Kringelbach, M. L. Affective neuroscience of pleasure: Reward in humans and other animals. *Psychopharmacology* 2008, 199, 457–480.
- Kringelbach, M. L. *The pleasure center: Trust Your Animal Instincts*. Oxford University Press, 2009.
- <sup>13</sup> Liebowitz, Michael, R. *The Chemistry of Love*. Boston: Little, Brown, & Co. 1983.
- Whitters, W. L. & Jones-Whitter, P. *Human Sexuality — A Biological Perspective*. New York: Van Nostrand. 1980.

## Глава шестая. Трансцендентные способности

- <sup>1</sup> Nielsen, Michael. *Reinventing discovery: the new era of networked science*. Princeton, N. J.: Princeton University Press. 2012, pp. 1–3.
- Gowers, T., Nielsen, M. Massively collaborative mathematics. *Nature* 2009, 461: 879–881.
- <http://gowers.wordpress.com/2009/02/01/a-combinatorial-approach-to-density-hales-jewett/>.
- Nielsen, Michael. The Polymath project: scope of participation (2009-03-20) <http://michaelnielsen.org/blog/?p=584>.
- Rehmyer, Julie SIAM: Massively Collaborative Mathematics. *siam.org*. 1 April 2010. <http://www.siam.org/news/news.php?id=1731>.
- <sup>2</sup> Dayan P., Huys Q. J. M. Serotonin, inhibition, and negative mood. *PLoS Comput Biol* 2008, 4(2).

## Глава седьмая. Цифровая кора — модель биологической коры

- <sup>1</sup> Gary Cziko, *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*, The MIT Press, 1955.
- <sup>2</sup> Дэвид Делримпл, которому сейчас уже больше 20 лет, был моим воспитанником с 1999 г., когда ему было восемь лет. Подробнее

о его проекте можно узнать здесь: <http://esp.mit.edu/learn/teachers/davidad/bio.html> и <http://www.brainsciences.org/Research-Team/mr-david-dalrymple.html>.

<sup>3</sup> Jonathan Fildes (22 July 2009). Artificial brain «10 years away», *BBC News*. <http://news.bbc.co.uk/2/hi/8164060.stm>.

См. также передачу про Генри Марккрама:

<http://www.kurzweilai.net/henry-markram-simulating-the-brain-next-decisive-years> и [http://www.youtube.com/watch?feature=player\\_embedded&v=cn\\_gEoQ8nQI#!](http://www.youtube.com/watch?feature=player_embedded&v=cn_gEoQ8nQI#!).

<sup>4</sup> M. Mitchell Waldrop, Computer modelling: Brain in a box, *Nature News*, 22 February 2012, <http://www.nature.com/news/computer-modelling-brain-in-a-box-1.10066>

<sup>5</sup> Jonah Lehrer, Can a Thinking, Remembering, Decision-Making Biologically Accurate Brain Be Built From a Supercomputer? *Seed Magazine*, 2012.

<sup>6</sup> Jonathan Fildes (22 July 2009). Artificial brain «10 years away». *BBC News*. <http://news.bbc.co.uk/2/hi/8164060.stm>.

<sup>7</sup> См. сайт <http://www.humanconnectomeproject.org/>.

<sup>8</sup> Anders Sandberg and Nick Bostrom (2008): *Whole Brain Emulation: A Roadmap*. Technical Report 2008 (3), Future of Humanity Institute, Oxford University. URL: [www.fhi.ox.ac.uk/reports/2008-3.pdf](http://www.fhi.ox.ac.uk/reports/2008-3.pdf).

<sup>9</sup> Ниже представлена базовая схема алгоритма функционирования сети нейронов. Возможно множество вариаций алгоритма, и создателю программы необходимо знать некоторые важнейшие параметры и методы, которые перечислены ниже. Создание сети нейронов для решения проблемы происходит в несколько этапов.

Определить входные данные.

Определить топологию сети (например, слои нейронов и связи между нейронами).

Оработать систему на нескольких примерах.

Опубликовать/запатентовать систему.

Эти этапы (кроме последнего) подробнее описаны ниже.

Входные данные.

Входные данные на сеть нейронов представляют собой несколько наборов чисел. Вот некоторые примеры:

— для системы распознавания визуальных образов — это двумерный ряд чисел, соответствующих пикселям изображения;

— для системы распознавания звуковой информации (например, речи) — это двумерный набор чисел, соответствующих звуку:

## Комментарии

первое измерение — параметры звука (например, частота), а второе измерение — различные моменты времени;  
— в системе распознавания произвольных образов — это  $n$ -мерный набор чисел, соответствующих входному сигналу.

Определение топологии сети.

Архитектура каждого нейрона определяется следующими параметрами.

Множество входных сигналов, каждый из которых «соединен» либо с выходным сигналом другого нейрона, либо с одним из чисел, определяющих входной сигнал.

Обычно один выходной сигнал, связанный либо с входным сигналом другого нейрона (обычно более высокого уровня), либо с результирующим сигналом.

Установка нейронов первого уровня.

Создать  $N_0$  нейронов первого уровня. Для каждого из них осуществить «связь» каждого из многочисленных входных сигналов с «точками» (например, численными значениями) на входе в систему. Эти связи могут быть установлены случайным образом или с помощью эволюционного алгоритма (см. ниже).

Присвоить начальное значение «синаптического потенциала» каждой созданной связи. Эти параметры сначала могут быть одинаковыми, могут быть выбраны случайным образом или определены каким-то другим путем (см. ниже).

Создать следующие уровни нейронов.

Создать все  $M$  уровней нейронов. На каждом уровне определить все нейроны.

Для уровня  $i$ .

Создать  $N_i$  нейронов уровня  $i$ . Для каждого нейрона осуществить «связь» каждого из множества входных сигналов с выходными сигналами нейронов уровня  $i-1$  (ниже см. возможные варианты).

Присвоить начальное значение «синаптического потенциала» каждой созданной связи.

Эти параметры сначала могут быть одинаковыми, могут быть выбраны случайным образом или определены каким-то другим путем (см. ниже).

Выходные сигналы нейронов уровня  $M$  представляют собой результирующие сигналы на выходе из сети нейронов (ниже см. возможные варианты).

Тест.

Как работает каждый нейрон.

Когда нейрон определен, в каждом тесте он выполняет следующую функцию.

Каждый поступающий на нейрон взвешенный сигнал обчисляется путем умножения выходного сигнала от другого нейрона (или исходного входного сигнала), с которым соединен данный нейрон, на величину синаптического потенциала данной связи.

Все взвешенные сигналы, поступающие на данный нейрон, суммируются.

Если эта сумма выше порогового значения возбуждения нейрона, данный нейрон считается возбужденным, и его выходному сигналу присваивается значение 1. В противном случае выходной сигнал получает значение 0 (см. варианты ниже).

В каждом тесте проводится следующая процедура.

Для каждого уровня — от 0 до  $M$ :

Для каждого нейрона данного уровня:

Суммируются взвешенные входные сигналы (каждый взвешенный входной сигнал = выходной сигнал другого нейрона (или исходный входной сигнал), с которым связан данный нейрон,  $x$  синаптический потенциал связи).

Если эта сумма взвешенных сигналов выше порогового значения возбуждения нейрона, данный нейрон считается возбужденным, и его выходному сигналу присваивается значение 1. В противном случае выходной сигнал получает значение 0.

Тестирование сети нейронов.

Проводят серию повторных тестов для простой задачи.

После каждого теста показатели синаптических потенциалов всех связей настраивают таким образом, чтобы улучшить производительность сети нейронов в данной задаче (ниже обсуждается, как это сделать).

Продолжают тестирование до тех пор, пока точность распознавания не перестанет повышаться (то есть выйдет на асимптоту).

Ключевые решения.

В рамках описанной выше простой схемы создатель алгоритма должен прежде всего определить следующие параметры:

- чему соответствуют значения входных параметров;
- количество уровней нейронов;
- количество нейронов на каждом уровне (на разных уровнях не обязательно должно содержаться одинаковое количество нейронов);
- количество входных сигналов, приходящих на каждый нейрон каждого уровня. Количество входных сигналов (то есть связей

между нейронами) может быть различным для разных нейронов и разных уровней;

— список «возбужденных» связей. Для каждого нейрона каждого уровня создается список других нейронов, выходные сигналы которых являются входными сигналами на данный нейрон. Это ключевой момент в создании программы. Сделать это можно несколькими способами:

- 1) возбужденные связи выбираются случайным образом; или
- 2) для оптимального возбуждения применяется эволюционный алгоритм (см. ниже); или
- 3) для определения возбужденных связей используется лучшее конструктивное решение.

Определение исходного синаптического потенциала (то есть веса) каждой связи. Сделать это можно несколькими способами:

- 1) всем синаптическим потенциалам присваивается одинаковое значение; или
- 2) синаптическим потенциалам присваиваются разные случайные значения; или
- 3) для определения оптимального набора синаптических потенциалов применяется эволюционный алгоритм; или
- 4) для определения исходных значений синаптических потенциалов используется лучшее конструктивное решение.

Порог возбуждения каждого нейрона.

Определение выходного сигнала. Выходной сигнал может быть:

- 1) выходными сигналами после уровня  $M$ ; или
- 2) выходным сигналом единственного нейрона, входным сигналом на который являются выходные сигналы после уровня  $M$ ; или
- 3) функцией (например, суммой) выходных сигналов после уровня  $M$ ; или
- 4) другой функцией выходных сигналов нейронов разных уровней.

Определение степени подгонки синаптических потенциалов всех связей в процессе тестирования сети нейронов. Это очень важный элемент подготовки программы, который активно исследуется и широко обсуждается. Для осуществления этой процедуры существует несколько возможностей:

- 1) в каждом тесте повышают или понижают величину синаптического потенциала на небольшую фиксированную величину, чтобы выходной сигнал нервной сети можно было ближе подходить к правильному результату. Один из способов — пытаться и увеличивать, и уменьшать значение, чтобы выяснить, какое действие приближает систему

к желаемому результату. Однако для этого может потребоваться много времени, так что существуют другие методы, позволяющие определить, в каком направлении (уменьшения или увеличения потенциала) следует двигаться;

2) существуют другие статистические методы изменения синаптического потенциала после каждого теста, чтобы добиться лучшего совпадения получаемых результатов с истинным значением.

Заметим, что тренировка системы будет происходить даже в тех случаях, когда не все результаты испытаний близки к истинным значениям. Это позволяет использовать результаты тестов, даже содержащих ошибку. Ключ к успеху при создании системы распознавания на основе сетей нейронов заключается в использовании большого количества тренировочных данных. Обычно для получения удовлетворительных результатов требуется весьма большой объем данных. Для системы на основе сетей нейронов, как и для студентов, процесс обучения — важнейший фактор в достижении высокой эффективности.

Вариации алгоритма.

Возможно множество вариаций.

Существуют разные способы определения топологии. В частности, возбуждение связей между нейронами можно задавать случайным образом, а можно с помощью эволюционного алгоритма.

Существуют разные способы определения исходных значений синаптического потенциала.

Входные сигналы на нейроны уровня  $i$  не обязательно являются выходными сигналами от нейронов уровня  $i-1$ . Входные сигналы на нейроны любого уровня могут приходиться от нейронов любого уровня.

Существуют разные пути определения конечного выходного сигнала.

Описанный выше метод приводит к возбуждению по принципу «все или ничего» (1 или 0) — это так называемая нелинейная функция. Но могут использоваться и другие нелинейные функции. Обычно применяют функцию, изменяющуюся в пределах от 0 до 1 достаточно быстро, но все же ступенчато. Кроме того, выходным сигналам могут присваиваться другие значения, отличные от 0 и 1.

Выбор метода подгонки величины синаптического потенциала в процессе наладки программы — очень важный момент. Описанная выше схема относится к «синхронным» сетям нейронов, в которых каждое испытание происходит путем вычислений выходов на каждом уровне нейронов, начиная от уровня 0 и заканчивая уровнем  $M$ . В истинно параллельных системах, где каждый нейрон действует

независимо от остальных, нейроны могут функционировать «асинхронно». При асинхронном подходе каждый нейрон постоянно сканирует свои входные сигналы и возбуждается в том случае, если сумма взвешенных входных сигналов превышает определенное пороговое значение (или при определенной специфике его выходного сигнала).

<sup>10</sup> Ниже представлена схема генетического (эволюционного) алгоритма. Существует множество вариаций этого алгоритма, и создателям системы необходимо подобрать некоторые критические параметры и методы (см. ниже).

Генетический алгоритм.

Создать  $N$  «сущест» (решений). Каждое из них характеризуется генетическим кодом — последовательностью чисел, характеризующих возможное решение проблемы. Эти числа могут описывать критические параметры, стадии решения, правила и др.

Процедуру осуществляют для каждого поколения в ходе эволюции.

Процедуру осуществляют для каждого из  $N$  сущест (решений).

Используют данное решение (представленное генетическим кодом существа) для решения проблемы или симуляции. Оценивают решение.

Выбирают  $L$  сущест с максимальным рейтингом выживания в будущем поколении.

Отбрасывают  $(N-L)$  оставшихся не способных выжить сущест.

Создают  $(N-L)$  новых сущест из  $L$  выживших сущест следующим образом:

1) делают копии  $L$  выживших сущест. В каждую копию вводят небольшие случайные вариации; или

2) создают дополнительных сущест путем комбинаций генетического кода («половое» воспроизведение или комбинация отдельных частей хромосом)  $L$  выживших сущест; или

3) выполняют комбинацию методов 1 и 2.

Принимают решение о необходимости дальнейшей эволюции.

Улучшение = (высший рейтинг в данном поколении) – (высший рейтинг в предыдущем поколении).

Если Улучшение < Предел улучшения, значит, процедура закончена.

Существо с наивысшим рейтингом из последнего поколения является наилучшим решением.

Используют это решение, определяемое его генетическим кодом, для решения проблемы.

Ключевые решения.

При реализации приведенной выше простой схемы создатель программы должен сначала установить следующие показатели.

Ключевые параметры:

N

L

Предел улучшения.

Чему соответствуют числа генетического кода, и как на основе генетического кода рассчитывается решение.

Метод выбора N существ в первом поколении. Обычно это лишь некая «разумная» попытка найти решение. Если разброс решений первого поколения слишком велик, эволюционный алгоритм может не найти хорошего решения. Часто имеет смысл выбрать исходные существа таким образом, чтобы они отличались разумным разнообразием. Это позволит избежать обнаружения оптимального решения «локального» значения.

Способ оценки решений.

Способ воспроизводства выживших существ.

Вариации.

Возможно множество вариаций предыдущего сценария. Например...

В каждом поколении не обязательно должно оставаться фиксированное количество выживших существ (L). Правила выживания могут позволять выживать разному количеству существ.

В каждом поколении не обязательно должно появляться фиксированное количество новых существ (N-L). Правила воспроизводства могут быть независимыми от размера популяции. Воспроизводство может коррелировать с выживанием, в результате чего наиболее успешные существа лучше воспроизводятся.

Решение о продолжении эволюции может приниматься по-разному. Оно может учитывать не только лучших существ из последних поколений, но и другую информацию. Решение также может учитывать тенденцию, простирающуюся за пределы двух последних поколений.

<sup>11</sup> Диссертационная работа Дайлипа Джорджа: Dileep George, How the brain might work: a hierarchical and temporal model for learning and recognition, 2008.

- <sup>12</sup> Turing, A. M. Computing Machinery and intelligence. *Mind*. 1950, vol. 59(236): 433–460.
- <sup>13</sup> Хью Лебнер учредил ежегодно вручаемую «Премия Лебнера». По замыслу Лебнера, серебряную медаль получает создатель компьютера, способного пройти стандартный (текстовый) тест Тьюринга. Золотой медалью награждают создателей машины, способной пройти вариант теста, подразумевающий аудио- и видеоварианты вопросов и ответов. На мой взгляд, пройти аудио- и видеoversию теста не сложнее, чем стандартную версию.
- <sup>14</sup> Умный помощник, который обучается и организует, SRI International, <http://www.ai.sri.com/project/CALO>.
- <sup>15</sup> Dragon Go!, Nuance Communications, Inc., <http://www.nuance.com/products/dragon-go-in-action/index.htm>.
- <sup>16</sup> Блог Стивена Вольфрама, 17 апреля 2012. «Overcoming Artificial Stupidity»: <http://blog.wolframalpha.com/author/stephenwolfram/>.

## Глава восьмая. Мозг как компьютер

- <sup>1</sup> S. Bochner. *A Biographical Memoire of John von Neumann*. National Academy of Sciences. Washington, D.C. 1958.
- <sup>2</sup> Turing, A. M. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 1936, 2 (42): 230–65.  
<http://www.comlab.ox.ac.uk/activities/ieg/e-library/sources/tp2-ie.pdf>.
- Turing, A.M. On Computable Numbers, with an Application to the Entscheidungsproblem: A correction. *Proceedings of the London Mathematical Society*, 1938, 43: 544–546.
- <sup>3</sup> Von Neumann, John. First Draft of a Report on the EDVAC. Moore School of Electrical Engineering, University of Pennsylvania, 1945.  
Von Neumann, John. A Mathematical Theory of Communication by John Von Neumann in the *Bell System Technical Journal*, 1948.
- <sup>4</sup> Jeremy Bernstein, *The Analytical Engine: Computers — Past, Present, and Future*, William Morrow & Co; Revised edition, 1981.
- <sup>5</sup> Japan's K Computer Tops 10 Petaflop/s to Stay Atop TOP500 List, *Top 500*, Nov. 11, 2011, <http://top500.org/lists/2011/11/press-release>.
- <sup>6</sup> Carver Mead, *Analog VLSI and Neural Systems*, Reading, Mass.: Addison-Wesley, 1986.
- <sup>7</sup> IBM Unveils Cognitive Computing Chips, IBM, Aug 18, 2011, <http://www-03.ibm.com/press/us/en/pressrelease/35251.wss>

## Глава девятая. Мысленные эксперименты по анализу разума

- <sup>1</sup> John R. Searle, I Married a Computer, in Richards et al., *Are We Spiritual Machines?*
- <sup>2</sup> Stuart Hameroff, *Ultimate Computing, Biomolecular Consciousness and NanoTechnology*. Elsevier Science Publishers B.V., 1987.
- <sup>3</sup> Sebel P. S. et al., The incidence of awareness during anesthesia: A multicenter United States study, *Anesth Analg* 2004; 99: 833–839.
- <sup>4</sup> Stuart Sutherland, in *The International Dictionary of Psychology*, Macmillan, 1990.
- <sup>5</sup> David Cockburn. *JSTOR Philosophy*. 1994, Vol. 69 (268): 135–150.
- <sup>6</sup> Из лекции Ивана Петровича Павлова (1913), *Lectures on conditioned reflexes. Twenty-five year of objective study of the higher nervous activity [behavior] of animals*, London: Martin Lawrence, 1928, p. 222.
- <sup>7</sup> Roger W. Sperry, from *James Arthur Lecture on the Evolution of the Human Brain*, 1964, p. 2.
- <sup>8</sup> H. Maudsley, The Double Brain, *Mind*, 1889, XIV (54): 161–187.
- <sup>9</sup> Susan Curtiss, Stella de Bode, Language after hemispherectomy, *Brain Cogn.*, 2000, 43(1–3): 135–138.
- <sup>10</sup> Vining E. P., Freeman J. M., Pillas D. J., Uematsu S., Carson B. S., Brandt J., Boatman D., Pulsifer M. B., Zuckerman A. Why would you remove half a brain? The outcome of 58 children after hemispherectomy — the Johns Hopkins experience: 1968 to 1996. *Pediatrics*. 1997, 100(2 Pt 1): 163–171. Pulsifer M. B., Brandt J., Salorio C. F., Vining E. P., Carson B. S., Freeman J. M. The cognitive outcome of hemispherectomy in 71 children. *Epilepsia*. 2004, 45(3): 243–254.
- <sup>11</sup> McClelland S. 3rd, Maxwell R. E. Hemispherectomy for intractable epilepsy in adults: the first reported series. *Ann Neurol*. 2007, 61(4): 372–376.
- <sup>12</sup> Lars Muckli, Marcus J. Naumerd and Wolf Singer, Bilateral visual field maps in a patient with only one hemisphere *PNAS*, 2009 vol. 106 (31): 13034–13039.
- <sup>13</sup> Marvin Minsky, *The Society of Mind*, Simon and Schuster, 1988.
- <sup>14</sup> F. Fay Evans-Martin, *The Nervous System*, Chelsea House Publications, 2005.  
<http://www.scribd.com/doc/5012597/The-Nervous-System>.
- <sup>15</sup> Benjamin Libet, *Mind time: The temporal factor in consciousness, Perspectives in Cognitive Neuroscience*; <http://www.amazon.com/Mind-Time-Consciousness-Perspectives-Neuroscience/dp/067401846X>.

- <sup>16</sup> Daniel C. Dennett, *Freedom Evolves*, New York: Viking, 2003.
- <sup>17</sup> Michael S. Gazzaniga, *Who's in Charge? Free Will and the Science of the Brain* Ecco/HarperCollins, 2011.
- <sup>18</sup> Hume, D. *An Enquiry Concerning Human Understanding* (1765) Indianapolis: Hackett Publishing Co. Second edition, 1993.
- <sup>19</sup> Schopenhauer, Arthur, *The Wisdom of Life*, p. 147.
- <sup>20</sup> Schopenhauer, Arthur, *On the Freedom of the Will*, Oxford: Basil Blackwell.
- <sup>21</sup> Raymond Smullyan, *5000 B.C. and Other Philosophical Fantasies*. St. Martin's Press, 1983.
- <sup>22</sup> How Do You Persist When Your Molecules Don't? *Science and Consciousness Review*, 1.1 (2004); <http://www.sci-con.org/articles/20040601.html>.

## **Глава десятая. Закон ускорения отдачи в применении к мозгу**

- <sup>1</sup> Секвенирование ДНК: <http://www.dnasequencing.org/history-of-dna>.
- <sup>2</sup> Закон Купера: Arraycomm, <http://www.arraycomm.com/technology/coopers-law>
- <sup>3</sup> Without driver or map, vans go from Italy to China, *The Sydney Morning Herald*, October 29, 2010; <http://www.smh.com.au/technology/technology-news/without-driver-or-map-vans-go-from-italy-to-china-20101029-176ja.html>.

## **Глава одиннадцатая. Критика**

- <sup>1</sup> Paul G. Allen and Mark Graves, Paul Allen: The Singularity Isn't Near, *Technology Review Magazine*, October 12, 2011; <http://www.technologyreview.com/blog/guest/27206/>.
- <sup>2</sup> Версия плана ITRS 2009 г.: *ITRS 2011 Edition*; <http://www.itrs.net/Links/2011ITRS/Home2011.htm>.
- <sup>3</sup> Ray Kurzweil, *The Singularity Is Near*, Viking, 2005 (глава 2).
- <sup>4</sup> Комментарий 2 в статье «Сингулярность еще не близка» Пола Аллена и Марка Гривса звучит следующим образом: «Мы уже начинаем создавать компьютеры, которые необходимы нам для подобных симуляций функции мозга. Петафлопсные компьютеры (такие как BlueGene/P компании IBM, который был использован для создания Ватсона) теперь уже есть в продаже. Уже проектируются эксафлопсные суперкомпьютеры. Эти системы, возможно, будут достаточно мощными, чтобы симулировать возбуждение

всех нейронов мозга, хотя пока это происходит во много раз медленнее, чем в настоящем мозге».

<sup>5</sup> Ray Kurzweil, *The Singularity Is Near*, глава 9, с. 435–442.

<sup>6</sup> Ray Kurzweil, *The Singularity Is Near*, глава 9.

<sup>7</sup> Хотя из-за повторов пар оснований нельзя точно оценить информационное содержимое генома, оно очевидно намного меньше всего объема данных до компрессии. Ниже представлены два способа оценки информационного содержания генома после компрессии. В обоих случаях верхним пределом является значение от 30 до 100 млн байт.

1. Без сжатия генетический код человека содержит 3 млрд оснований ДНК, каждое кодируется двумя битами информации (поскольку в каждой позиции может находиться один из четырех нуклеотидов). Таким образом, до компрессии в геноме человека содержится около 800 млн байт информации. Некодирующую ДНК иногда называют «мусорной» ДНК, однако сейчас уже ясно, что она играет важную роль в экспрессии генов. Но закодирована она очень неэффективно. В частности, имеет место чрезвычайная избыточность информации (например, последовательность «ALU» повторяется сотни тысяч раз), что можно использовать при компрессии.

Ввиду активного расширения объема информации в генетических базах данных существует повышенный интерес к разработке способов сжатия генетических данных. С помощью стандартных алгоритмов компрессии генетические данные удастся сжать примерно на 90% по объему: Hisahiko Sato et al., DNA Data Compression in the Post Genome Era, *Genome Informatics* 12 (2001): 512–514.

Таким образом, удастся сжать геном до объема около 80 млн байт без потери информации (это означает, что на основании этих данных мы прекрасно можем воссоздать исходные 800 млн байт генетической информации).

Теперь учтем, что свыше 98% генома не кодирует белки. Даже после применения стандартного алгоритма компрессии (в результате чего устраняется избыточность и используется стандартный поиск известных последовательностей) алгоритмическое содержание не кодирующих областей оказывается достаточно низким, что означает, что мы можем дополнительно сократить объем текста без потери информации. Однако, поскольку мы все еще находимся в начале пути обратного проектирования генома, мы не можем достаточно надежно оценить последствия дальнейшего сокращения объема на основе

функционально эквивалентного алгоритма. Таким образом, мне кажется разумным остановиться на компрессии до 30–100 млн байт. Верхний предел этого диапазона соответствует только компрессии данных, без алгоритмического упрощения.

Только часть (хотя и значительная) этой информации относится к строению головного мозга.

2. Второй способ рассуждений следующий. Поскольку в геноме человека содержится около 3 млрд оснований, лишь небольшая часть отвечает за кодирование белков. По современным оценкам, в геноме человека существует около 26 тыс. генов белков. Если принять, что эти гены в среднем содержат около 3000 оснований полезной информации, получим около 78 млн оснований. Информационное содержание одного основания ДНК составляет только два бита, так что в сумме все эти основания составляют около 20 млн байт информации (78 млн оснований поделить на четыре). В кодирующей белок последовательности гена каждое слово (кодон), состоящее из трех оснований ДНК, транслируется в одну аминокислоту. Таким образом, существует  $4^3 = 64$  кодона, состоящих из трех оснований ДНК. Однако в белках присутствует лишь 20 аминокислот плюс стоп-кодон (не кодирующий никакой аминокислоты). Оставшиеся 43 кодона — это синонимы первых 21. Для кодирования 64 возможных комбинаций нуклеотидов нужно 6 бит, а для кодирования 21 комбинации — только 4,4 бита ( $\log_2 21$ ), что позволяет сохранить 1,6 бит из 6 (около 27%) и приводит нас к общему значению 15 млн байт. Кроме того, возможно произвести стандартную компрессию повторяющихся последовательностей, хотя в кодирующих последовательностях возможностей для сжатия значительно меньше, чем в участках «мусорной» ДНК, отличающейся высокой степенью избыточности. Так мы можем спуститься до значения около 12 млн байт. Однако теперь мы должны добавить информацию относительно некодирующих участков ДНК, которые контролируют экспрессию генов. Хотя эти последовательности ДНК составляют преобладающую часть генома, их информационное содержание достаточно низкое, а уровень избыточности высокий. Учитывая наличие примерно 12 млн байт информации в кодирующих областях ДНК, вновь возвращаемся к значению около 24 млн байт. Это значительно ниже, чем полученная выше оценка от 30 до 100 млн байт.

<sup>8</sup> S. Modha, et al., *Communications of the ACM*, 2011, Vol. 54 (8): 62–71; <http://cacm.acm.org/magazines/2011/8/114944-cognitive-computing/fulltext>.

<sup>9</sup> Ray Kurzweil, *The Singularity Is Near*, глава 9, с. 458–469.

- <sup>10</sup> Micheal Denton, *Organism and Machine: The Flawed Analogy*, in *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI*, Discovery Institute, 2002, Discovery Institute, 2001.
- <sup>11</sup> Hans Moravec, *Mind Children*, Harvard University Press, 1988.

## Эпилог

- <sup>1</sup> <http://www.gallup.com/poll/147350/optimism-future-youth-reaches-time-low.aspx>.
- <sup>2</sup> James C. Riley, *Rising Life Expectancy: A Global History*, Cambridge University Press, 2001.
- <sup>3</sup> Martine Rothblatt, *Transgender to Transhuman*, 2011. Автор объясняет, как изменится наше отношение к «трансчеловеку» — например, обладателю небологического, но убедительно сознательного разума.
- <sup>4</sup> В приведенном ниже абзаце из третьей главы книги «Сингулярность уже близка» (с. 133–135) обсуждаются пределы возможности компьютерных вычислений, основанные на законах физики.

Возможности компьютеров весьма велики. На базе работ профессора Ганса Бремерманна и профессора нанотехнологии Роберта Фрейтаса из Университета Калифорнии в Беркли профессор Сет Ллойд из Массачусетского технологического института в соответствии с известными физическими законами оценил возможности «последнего» компьютера массой один килограмм и объемом один литр (что соответствует размеру и весу современного портативного компьютера):

Seth Lloyd, *Ultimate Physical Limits to Computation*, *Nature* 406 (2000): 1047–1054.

Первые оценки возможностей компьютерных вычислений были даны Гансом Бремерманном в 1962 г.:

Hans J. Bremermann, *Optimization Through Evolution and Recombination*, in M. C. Yovits, C. T. Jacobi, C. D. Goldstein (eds.), *Self-Organizing Systems*, Spartan Books, Washington, 1962, pp. 93–106.

В 1984 г. Роберт Фрейтас младший продолжил исследования Бремерманна:

Robert A. Freitas Jr., *Xenopsychology*, *Analog*, 1984, 104: 41–53, <http://www.rfreitas.com/Astro/Xenopsychology.htm#SentienceQuotient>.

Возможности компьютерных вычислений зависят от доступности энергии. Энергия, заключенная в определенном количестве вещества, связана с атомами и субатомными частицами этого вещества.

Чем больше в веществе атомов, тем больше энергия. Как обсуждалось выше, для вычислений в принципе могут использоваться любые атомы. Поэтому чем больше атомов — тем больше возможности вычислений. Энергия атомов и частиц увеличивается с повышением частоты их движений — чем активнее движение, тем больше энергия. То же соотношение верно и для возможности компьютерных вычислений: чем выше частота движения, тем больше вычислений может совершить каждый элемент системы (например, атом). Именно это правило реализуется в современных микросхемах: чем выше частота чипа, тем больше скорость вычислений.

Таким образом, способность объекта производить вычисления пропорциональна его энергии. В соответствии с уравнением Эйнштейна ( $E = mc^2$ ) потенциальная энергия килограмма вещества чрезвычайно велика. Квадрат скорости света — это очень большая величина, примерно  $10^{17} \text{ м}^2/\text{с}^2$ . Вычислительный потенциал вещества также определяется постоянной Планка, которая имеет чрезвычайно малую величину —  $6,6 \times 10^{-34} \text{ Дж}\cdot\text{с}$ . Теоретический предел способности объекта осуществлять вычисления можно получить путем деления общей энергии объекта (средней энергии каждого атома или частицы этого объекта, умноженной на число таких атомов или частиц) на постоянную Планка.

Ллойд показывает, что потенциальная вычислительная способность килограмма вещества равна энергии, поделенной на постоянную Планка и умноженной на число  $\pi$ . И поскольку энергия килограмма вещества столь велика, а постоянная Планка столь мала, данное уравнение приводит нас к чрезвычайно большой величине — около  $5 \times 10^{50}$  операций в секунду:

$$(\pi \times 10^{17} \text{ м}^2/\text{с}^2)/(6,6 \times 10^{-34} \text{ Дж}\cdot\text{с}) \sim 5 \times 10^{50} \text{ операций/с.}$$

Если мы сравним это значение с самыми консервативными оценками способностей человеческого мозга ( $10^{19}$  операций/с и около  $10^{10}$  жителей планеты), мы приходим к выводу, что операционная способность килограмма вещества в пять миллиардов триллионов раз выше соответствующей способности человеческой цивилизации:

$5 \times 10^{50}$  операций/с  $\sim 5 \times 10^{21}$  (пять миллиардов триллионов)  $\times$  способность человеческой цивилизации ( $10^{29}$  операций/с).

Если, как я считаю, для функциональной симуляции человеческого разума будет достаточно операционной способности  $10^{16}$  операций/с, вычислительная способность «последнего» компьютера будет эквивалентна вычислительной способности разума пяти триллионов триллионов человеческих цивилизаций:

$10^{10}$  человек  $\times$   $10^{16}$  операций/с =  $10^{26}$  — такова вычислительная способность человеческой цивилизации;  $5 \times 10^{50}$  операций/с  $\sim 5 \times 10^{24}$  (5 триллионов триллионов)  $\times 10^{26}$ .

Объем вычислений, проделанных таким компьютером за одну тысячную наносекунды, эквивалентен объему вычислений, произведенных всей человеческой цивилизацией за последние 10 тыс. лет. Это оценочное значение основано на допущении, что за последние 10 тыс. лет на Земле жило 10 млрд человек, что, конечно же, неверно. В реальности население Земли очень сильно выросло за последнее время и достигло примерно 6,1 млрд в 2000 г. В году  $3 \times 10^7$  секунд, а в десяти тысячах лет —  $3 \times 10^{11}$  секунды. Таким образом, используя оценочное значение для вычислительной способности человеческой цивилизации  $10^{26}$  операций/с, можно сказать, что за 10 тыс. лет суммарный разум всей человеческой цивилизации совершил не более  $3 \times 10^{37}$  вычислений. «Последний» компьютер осуществляет  $5 \times 10^{50}$  вычислений в секунду. Поэтому симуляция мыслительной активности 10 млрд человек за 10 тыс. лет займет примерно  $10^{-13}$  с, что составляет одну тысячную наносекунды.

Нужно сделать еще несколько замечаний. Если мы превратим в энергию массу всего компьютера, это будет сравнимо с термоядерным взрывом. Конечно же, мы не стремимся к тому, чтобы компьютер взорвался, а хотим, чтобы он сохранял свой объем. Поэтому эта система должна быть определенным образом упакована. Анализируя максимальную энтропию такой системы (степень свободы, определяемую состоянием всех составляющих ее частиц), Ллойд пришел к выводу, что подобный компьютер теоретически может иметь объем памяти около  $10^{31}$  бит. Трудно представить себе технологию, которая сможет достичь такого предела, но можно попытаться к нему приблизиться. Как показывают результаты проекта, выполнявшегося в Университете Оклахомы, мы уже способны сохранять примерно 50 бит информации в расчете на один атом (хотя пока еще для небольшого количества атомов). Таким образом, можно представить себе хранение  $10^{27}$  бит памяти в  $10^{25}$  атомах, содержащихся в одном килограмме вещества.

Однако, поскольку для хранения информации используются различные свойства атомов (такие как их точное положение, спиновое число и квантовое состояние всех составляющих атом частиц), возможно, нам удастся превзойти этот показатель. Нейробиолог Андерс Сэндберг оценивает потенциальную емкость атома водорода примерно в 4 млн бит информации (Anders Sandberg, *The Physics of the Information Processing Superobjects: Daily Life Among the Jupiter Brains*,

## Комментарии

*Journal of Evolution & Technology* Vol. 5 (1999); <http://www.transhumanist.com/volume5/Brains2.pdf>). Однако пока такие возможности еще не реализованы, и мы придерживаемся более консервативных оценок.

Как отмечалось выше, скорости  $10^{42}$  операций в секунду можно достичь без значительного разогрева системы. Использование всех возможностей техники обратимых вычислений, систем с низким количеством ошибок и небольшими потерями энергии позволит нам достичь значения в пределах от  $10^{42}$  до  $10^{50}$  операций в секунду.

Я не буду здесь рассматривать технические вопросы, возникающие при попытках повышения производительности систем от  $10^{42}$  до  $10^{50}$  операций в секунду. Скажу только, что вряд ли мы начнем от предельной величины  $10^{50}$  и будем двигаться назад в соответствии с различными практическими соображениями. Скорее всего, технология будет постепенно продвигаться вперед, каждый раз основываясь на самых последних достижениях технологии предыдущего поколения. Так что, когда мы подберемся к значению  $10^{42}$  (на каждый килограмм массы), ученые и инженеры той эпохи используют свой в значительной мере небиологический разум, чтобы продвинуться к значению  $10^{43}$ , потом  $10^{44}$  и т. д. По моему ощущению, мы сможем очень близко подойти к теоретическим пределам.

Даже при производительности  $10^{42}$  операций/с килограммовый «последний» компьютер будет способен за 10 микросекунд осуществить вычисления, эквивалентные всем мыслительным операциям человечества за 10 тыс. лет ( $10^{42}$  в  $10^8$  меньше, чем  $10^{50}$ , так что одна десятитысячная наносекунды превращается в десять микросекунд).

Учитывая экспоненциальный рост вычислительного потенциала компьютеров (вторая глава), я считаю, что компьютеры стоимостью в тысячу долларов будут способны осуществлять подобные операции примерно к 2080 г.

# КОГДА ВЫ ДАРИТЕ КНИГУ, ВЫ ДАРИТЕ ЦЕЛЫЙ МИР

## ХОТИТЕ ЗНАТЬ БОЛЬШЕ?

Заходите на сайт:

<https://eksmo.ru/b2b/>

Звоните по телефону:

+7 495 411-68-59, доб. 2261



ВАШ ЛОГОТИП  
НА ОБЛОЖКЕ

ВАШ ЛОГОТИП НА КОРЕШКЕ

ОБРАЩЕНИЕ  
К КЛИЕНТАМ  
НА ОБЛОЖКЕ

**РЭЙ КУРЦВЕЙЛ** – знаковая фигура в науке, один из самых известных профессиональных футурологов современности, успешный американский изобретатель. Его научные технологические прогнозы по поводу искусственного интеллекта и продления человеческой жизни известны во всем мире. Рэю Курцвейлу принадлежит понятие «технологическая сингулярность». Его называют «последней мыслящей машиной» и «достойным наследником Томаса Эдисона». Автор нескольких научных бестселлеров.

Центральная идея **«ЭВОЛЮЦИИ РАЗУМА»** – **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**, который со временем будет властвовать во всех сферах жизни людей. Воссоздание человеческого интеллекта станет той точкой невозврата, после которой люди окончательно потеряют возможность осознавать прогресс. Вы узнаете о бесконечном потенциале возможностей в сфере обратного проектирования человеческого мозга. К **2029** году, по уверению автора, исчезнет пропасть, разделяющая людей и компьютеры. Мозг будет подключаться к компьютеру напрямую.

## **И НАСТУПИТ НОВАЯ ЭПОХА СУЩЕСТВОВАНИЯ МИРА...**

«Если вы когда-нибудь задумывались о том, как работает мозг, прочтите эту книгу. Курцвейл раскрывает секреты человеческого мышления и говорит о возможности его воссоздания. Это убедительная книга, которая заставляет думать».

**ДИН КЕЙМЕН**, физик, создатель  
портативного дозатора инсулина



### **БОМБОРА**

Бомбора — это новое название Эксмо Non-fiction, лидера на рынке полезных и вдохновляющих книг. Мы любим книги и создаем их, чтобы вы могли творить, открывать мир, пробовать новое, расти. Быть счастливыми. Быть на волне.

**f vk @ bomborabooks**  
[www.bombora.ru](http://www.bombora.ru)