

Файл предназначен только для личного ознакомления.
Тиражирование, копирование, распространение запрещено.

БЫТЬ СОБОЙ

НОВАЯ ТЕОРИЯ СОЗНАНИЯ



АНИЛ СЕТ

МЕЖДУНАРОДНЫЙ БЕСТСЕЛЛЕР

Пролог

Пять лет назад я в третий раз в своей жизни перестал существовать. Мне делали небольшую операцию, и в мозг поступал анестетик. Я помню, как погружался в темноту, отключался, распадался на части...

Общий наркоз совершенно не похож на сон. Так и должно быть: если бы мы просто засыпали, то спрыгивали бы с операционного стола при первом же прикосновении скальпеля. Глубокая анестезия больше напоминает такие катастрофические состояния, как кома или вегетативная жизнь, когда сознание отсутствует полностью. При глубоком наркозе электрическая активность мозга буквально сходит на нет, чего никогда не случается в обычной жизни — ни во сне, ни во время бодрствования. Это одно из подлинных чудес современной медицины, когда анестезиолог в штатном режиме вмешивается в работу человеческого мозга, погружая его в глубоко бессознательное состояние, а затем выводя оттуда. Преобразование едва ли не магическое: наркоз — это искусство превращения человека в неодушевленный объект.

Конечно, потом этот объект возвращают в прежнее состояние. Вернулся и я, одурманенный, дезориентированный, но определенно в себе. Времени, казалось, не прошло и секунды. Пробуждаясь от крепкого сна, иногда не понимаешь, который сейчас час, но ощущение, что хотя бы сколько-то времени прошло, всегда имеется, всегда есть некий мостик между осознанием себя «тогда» и «сейчас». Под общим наркозом все иначе. Я мог быть в отключке пять минут, пять часов, пять лет, а то и все 50, кто знает. Да и «отключка» не то слово. Наркоз — это самое настоящее небытие, предвестник полного беспомыслия, наступающего со смертью и дарующего загадочное умиротворение.

Общий наркоз действует не просто на мозг, он действует на разум. На сознание. В результате изменений тонкого электрохимического баланса в нейронной сети мозга упраздняется, пусть и временно, та основополагающая структура, в которой и заключено наше «бытие». Этот процесс таит в себе одну из величайших тайн науки — и философии, — до сих пор остающихся неразгаданными.

Каким-то образом в человеческом мозге совокупная деятельность миллиардов нейронов, каждый из которых представляет собой крошечный биологический механизм, порождает сознательный опыт, причем именно ваш, именно здесь, именно сейчас. Как это происходит? Почему у нас есть возможность проживать жизнь «от первого лица»?

Когда-то в далеком детстве, в очередной раз глянув в зеркало в ванной, я вдруг отчетливо осознал, что мое ощущение в этот конкретный момент — ощущение себя собой — рано или поздно закончится и это «я» умрет. Мне было лет восемь или девять, так что воспоминание это, как и все детские воспоминания, нельзя назвать надежным. Но, возможно, тогда же я осознал и другое: если мое сознание прекратится, наверное, это как-то связано с «тестом», из которого я слеплен, то есть с физической материальностью моего тела и мозга. Очень может статься, что с тех самых пор я и бьюсь над загадкой сознания.

В начале 1990-х гг. меня, кембриджского студента, пылавшего юношеской любовью к физике и философии, захватили психология и нейробиология, хотя в те времена в этих науках принято было обходить стороной — и даже объявлять вне закона — любые упоминания о сознании. Работа над диссертацией увела меня в неожиданные области искусственного интеллекта и робототехники — надолго, но с неожиданной пользой, после чего я шесть лет провел в Институте нейробиологии в Сан-Диего, на Тихоокеанском побережье, получив, наконец, возможность исследовать мозговые основы сознания напрямую. Я работал там с нобелевским лауреатом Джеральдом Эдельманом — одним из тех, кто больше других сделал для того, чтобы загадка сознания вновь оказалась в фокусе внимания науки.

И вот уже больше десяти лет я руковожу Центром изучения сознания имени Артура Саклера при Сассексском университете, который уютно расположился среди зеленых холмов Саут-Даунс неподалеку от морского курорта Брайтон. Наш центр объединяет нейроученых, психологов, психиатров, специалистов по нейровизуализации, разработчиков виртуальной реальности, математиков и философов — и все мы совместными усилиями ищем способы проникнуть в скрытую в мозге основу сознательного опыта.

Ученый вы или обычный человек, тайна сознания не может не волновать вас. Сознательный опыт — это все, что у нас есть. А без него нет ничего: ни окружающего мира, ни «себя», ни внутреннего, ни внешнего.

Представьте, что некий будущий я, возможно из самого ближайшего будущего, предлагает вам сделку века. Я заменю ваш мозг машиной, целиком ему соответствующей и ни в чем ему не уступающей, так что со стороны никто не заметит разницы. У этой машины много преимуществ — она не подвержена старению, поэтому, возможно, позволит вам жить вечно.

Но есть и подвох. Поскольку будущий я не знаю точно, как в настоящем мозге возникает сознание, я не могу гарантировать, что в случае согласия на эту сделку у вас останется сознательный опыт. Может быть, останется, если сознание обусловлено только функциональными способностями, мощностью и сложностью сетей мозга, а может быть, и нет, если сознание зависит от определенного биологического материала, например нейронов. Разумеется, поскольку машинный мозг будет обеспечивать поведение, полностью «идентичное натуральному», на мой вопрос, есть ли у вас сознание, «новый» вы все равно ответите утвердительно. Но что, если вопреки этому ответу вы уже не будете проживать жизнь от первого лица?

Подозреваю, что на сделку вы не согласитесь, ведь без сознания вам будет уже неважно, проживете вы еще пять лет или 500. Все это время вы в любом случае никак не будете ощущать себя собой.

Если отвлечься от философских материй, окажется, что изучение мозговых основ сознания имеет и практическую значимость, которую трудно переоценить. Помимо общего наркоза, по праву считающегося одним из величайших изобретений в истории, существуют и другие нарушения сознания. Они сопровождают черепно-мозговые травмы и психические заболевания у все большего числа людей (включая меня). Кроме того, у каждого из нас сознательное восприятие меняется по мере того, как мы переходим с одного жизненного этапа на другой: от сумбура детства к кажущейся, но, скорее всего, иллюзорной и уж точно не полной ясности во взрослом возрасте, за которой следует постепенное (а у кого-то предательски стремительное) скольжение под откос, в распад личности, по мере развития нейродегенеративных процессов на склоне лет. Вы существуете на каждом из этих этапов, но полагать, что все это время у вас имеется единственное и неповторимое сознательное «я» (душа?), может быть в корне неверно. Природа личности — это, собственно, один из самых интригующих моментов тайны сознания. Возможно ли сознание без осознания себя, а если да, будет ли оно по-прежнему настолько для нас значимо?

От ответов на подобные непростые вопросы во многом зависит, как мы будем воспринимать окружающий мир и его обитателей. На каком этапе развития живого организма возникает сознание? Появляется ли оно при рождении или закладывается еще в утробе? А что с сознанием у других животных помимо человека? Причем не только у приматов и прочих млекопитающих, но и у таких фантастических созданий, как осьминог, или даже у таких простых организмов, как нематоды или бактерии? Каково это — быть бактерией *Escherichia coli* или морским окунем? А как насчет будущих машин? Нам нужно задумываться не только о той власти, которую новые

формы искусственного интеллекта получают над нами, но и о том, надо ли нам (и если надо, то когда) занимать по отношению к ним ту или иную этическую позицию. Меня эти вопросы настраивают на такое же необъяснимое сопереживание, которое я испытывал, глядя, как в фильме «2001 год: Космическая одиссея» Дейв Боумен уничтожает личность суперкомпьютера HAL, просто деактивируя один за другим его банки памяти. А исходя из того, что незавидному положению репликантов в фильме Ридли Скотта «Бегущий по лезвию» мы сопереживаем еще больше, можно заключить, насколько важна для ощущения себя сознательным существом природа человека как живой машины.

Моя книга посвящена нейронауке сознания — это попытка понять, как внутренняя вселенная субъективного опыта соотносится с биологическими и физическими процессами, протекающими в мозге и организме, и как она может объясняться в терминах этих процессов. Эта задача занимала меня на протяжении всего моего научного пути, и вот теперь передо мной, кажется, забрезжили проблески ответов.

Эти проблески уже меняют, причем самым драматическим образом, наши представления о сознательном опыте, касающемся окружающего нас мира и нашего места в нем. Наши представления о сознании затрагивают все составляющие нашего бытия. Наука о сознании — это не что иное, как представление о том, кто мы такие, что значит быть мной, быть вами и почему вообще существует подобное «бытие».

История, которую я вам изложу, — это личный взгляд, сформировавшийся за годы исследований, размышлений и обсуждений. Мне представляется, что сознание не получится «расшифровать», как геном человека, или доказать, как глобальное потепление. И внезапного прорыва в его тайны — с триумфальным воплем «Эврика!» — вопреки красивому, но обычно далекому от реальности мифу о том, как происходит прогресс в науке, тоже ждать не стоит.

С моей точки зрения, теория сознания должна объяснять, как различные его свойства обуславливаются операциями нейронных «биомощностей» в нашем мозге и соотносятся с ними. В задачи теории сознания (по крайней мере, первостепенные) не должен входить поиск причин, по которым оно в принципе существует во вселенной. Не ее прерогатива выяснять, как устроен и работает весь наш сложный мозг, откладывая тайну сознания в долгий ящик. Я рассчитываю показать вам, что, объяснив свойства сознания с точки зрения механизмов мозга и организма, мы потихоньку будем проливать свет и на относящиеся к сознанию «почему» и «как» глубоко метафизического свойства.

Говоря «биомощности», я хочу подчеркнуть, что мозг — это не компьютер из плоти. Это в равной степени химическое устройство и электрическая сеть. Каждый когда-либо существовавший мозг — это часть живого организма, встроенного в свою среду и взаимодействующего с этой средой, в которую во многих случаях входят и другие экземпляры мозга, заключенные каждый в свое тело. Чтобы объяснять свойства сознания в терминах биофизических механизмов, необходимо рассматривать мозг — и сознательный разум — как встроенную и включенную в тело систему.

Мне хочется, чтобы после прочтения этой книги у вас появилось новое понимание себя — той составляющей сознания, которая, наверное, для каждого из нас является самой значимой. Авторитетная научная традиция, ведущая отсчет как минимум с написанных в XVII в. трудов Декарта, утверждала, что прочие животные, помимо человека, сознательным «я» не обладают, поскольку у них нет рационального разума, управляющего их поведением. Они считались «животными-машинами» — автоматами из плоти, не способными осмыслить собственное существование.

Я с этим не согласен. На мой взгляд, сознание связано не с интеллектом, а с жизнью. Мы являем собой сознательные «я» именно потому, что мы животные-машины. Я

докажу, что опыт бытия вами или бытия мной возникает из того, как мозг прогнозирует и контролирует внутреннее состояние организма. Суть собственного «я» — это и не рациональный разум, и не бесплотная душа. Это глубоко укорененный в организме биологический процесс — тот самый процесс, на котором базируется простое ощущение себя живым, в свою очередь составляющее основу любого нашего восприятия себя и, собственно, любого сознательного опыта в принципе. «Бытие собой» имеет самое прямое и буквальное отношение к существованию организма.

Моя книга состоит из четырех разделов. В первом я объясняю свой подход к научному исследованию сознания. Там же рассматривается вопрос сознательного «уровня» (насколько сознательным может быть кто-то или что-то), а также подвижки в попытках «измерить» сознание. Во втором разделе поднимается тема «содержания» сознания — что именно мы создаем, когда создаем. В третьем мы обратимся внутрь, к личности и ко всему разнообразию опыта, которое предполагает сознательное «я». А в четвертом, и последнем, посвященном «другим», постараемся выяснить, что этот новый подход к пониманию сознания позволяет сказать о прочих животных (помимо человека) и о возможности существования разумных машин. К концу книги вы поймете, что наше сознательное восприятие мира и самих себя является собой формы возникающего в мозге предвидения — «контролируемых галлюцинаций», которые являются результатом жизнедеятельности организма.

Зигмунд Фрейд, пусть в представлении современных нейроученых он порядком дискредитировал себя, на самом деле во многом был прав. Рассмотрев предшествующую историю науки, он отметил три «удара» по пьедесталу, на который вознесло себя человечество, — три крупных научных открытия, принятых современниками в штыки. Первый удар нанес Коперник, доказавший в своей гелиоцентрической теории, что Земля вращается вокруг Солнца, а не наоборот. Так пришло осознание, что мы не центр вселенной, а просто искра в необъятной кромешной тьме, голубая точка, висящая в бездне. Следующим был Дарвин, установивший родство человека со всеми остальными живыми существами, — это осознание, как ни поразительно, в некоторых уголках мира отвергают до сих пор. Третьим ударом по человеческой исключительности Фрейд нескромно назвал собственную теорию бессознательного, поставившую под сомнение идею, что наша психическая жизнь находится под сознательным контролем разума. И хотя в частностях Фрейд несколько ошибся, он был абсолютно прав, называя естественно-научное объяснение разума и сознания еще одной, возможно завершающей, вехой в свержении человечества с пьедестала.

Эти сдвиги в нашем самовосприятии можно только приветствовать. С каждым новым шагом вперед по пути понимания приходит ощущение чуда и осознание себя не отделенными от остальной природы, а в большей степени включенными в нее.

Наш сознательный опыт — это такая же часть живой природы, как наш организм и окружающий нас мир. Сознание заканчивается тогда, когда заканчивается жизнь. Эта мысль возвращает меня к моему ощущению — неощущению — общего наркоза. К беспмятству небытия, пусть даже умиротворяющему, но все же беспмятству. Лучше всего об этом сказал писатель Джулиан Барнс в своих размышлениях о смерти¹. Когда сознанию приходит конец, нам действительно становится нечего — совсем нечего — бояться.



Глава 1

Настоящая проблема

Что такое сознание?

У существа, обладающего сознанием, имеется некое ощущение того, что значит быть этим существом. Быть мной, быть вами, возможно, быть овцой или дельфином. У каждого из таких существ имеются субъективные впечатления, субъективный опыт. Быть мной — это некое определенное ощущение, аналогичного которому почти наверняка не найдется у бактерии, травинки или игрушечного робота. У этих объектов никакого субъективного опыта (предположительно) быть не может: у них нет внутреннего мира, восприятия окружающего, сознания.

Этот подход ассоциируется в первую очередь с теорией философа Томаса Нагеля, который в 1974 г. опубликовал теперь уже легендарную статью, озаглавленную «Что значит быть летучей мышью?». В статье он доказывал: даже если человек никогда не сможет испытать те же ощущения и получить тот же опыт, что и летучая мышь, у самой мыши некое ощущение себя летучей мышью все же присутствует^[2]. Я всегда выступал за этот подход, поскольку он делает упор на феноменологию — субъективные свойства сознательного опыта (например, почему визуальное впечатление имеет именно такую форму, структуру и качества в сравнении с субъективными свойствами эмоционального впечатления или обонятельного). В философии эти свойства иногда называются «квалиа» — краснота красного цвета, жгучесть ревности, пронзительность или тупое нытье зубной боли и тому подобное.

Сознательным будет считаться организм, обладающий той или иной собственной феноменологией. В зачет идет любой опыт — любое феноменологическое качество. Где есть опыт, есть феноменология, а где есть феноменология, есть сознание. Существо, которое возникает лишь на миг, будет сознательным ровно столько, сколько длится опыт бытия этим существом, даже если он сведется к секундному ощущению боли или удовольствия.

Феноменологические свойства сознания вполне отличимы от его функциональных или поведенческих свойств, относящихся к ролям, которые сознание может играть в операциях нашего разума и мозга, и к поведению, на которое способно живое существо вследствие обладания сознательным опытом. И хотя связанные с сознанием функции и поведение — тема важная, обращаться к ним за определениями вряд ли стоит. Сознание — это, в первую и самую главную очередь, субъективный опыт, а именно феноменология.

Эта очевидная, казалось бы, идея была очевидной не всегда. В прошлом обладание сознанием зачастую путали с наличием языка, умственных способностей или демонстрацией определенного поведения. Однако сознание не зависит от внешнего поведения, как нетрудно убедиться на примере спящих или полностью парализованных. Заявлять, что для обладания сознанием необходимо владение речью, — значит отказывать в наличии сознания еще не разговаривающим

младенцам, утратившим речевые способности взрослым, а также большинству, если не всем, животным помимо человека. Сложное абстрактное мышление — это лишь малая часть обладания сознанием (хотя и не исключено, что сугубо человеческая).

Ряд авторитетных теорий в области науки о сознании по-прежнему ставит функции и поведение превыше феноменологии. В первую очередь это «теория глобального рабочего пространства»¹, которую в течение многих лет разрабатывали психолог Бернард Баарс, нейробиолог Станислас Деан и прочие. Согласно этой теории психическое содержание (восприятие, мысли, эмоции и так далее) становится частью сознания, только получая доступ к «рабочему пространству», которое, если рассматривать анатомически, распределено между лобными и теменными долями коры. (Кора — это отличающаяся сильной складчатостью внешняя поверхность мозга, образованная плотно скомпонованными нейронами.)^[3] Когда психическое содержание транслируется в рабочее пространство коры, мы его осознаем, и оно получает возможность руководить нашим поведением гораздо более гибко, чем бессознательное восприятие. Допустим, я отдаю себе отчет в том, что на столе передо мной находится стакан воды. Я могу взять его и выпить, выплеснуть в монитор своего компьютера (искушение велико), написать о нем стихотворение или, наконец, осознав, что он стоит тут уже не первый день, отнести его обратно на кухню. Бессознательное восприятие такой степени поведенческой гибкости не допускает².

Другая известная теория — так называемая теория сознания высшего порядка³ — предполагает, что психическое содержание становится сознательным, когда его делает таковым тот или иной ориентированный на него когнитивный процесс «высшего порядка». Согласно этой теории сознание тесно связано с процессами вроде метасознания (то есть представляющими собой «осознание осознания»), и, следовательно, как и теория глобального рабочего пространства, данная теория перевозносит (хотя и не так сильно) функциональные свойства над феноменологией. И точно так же, как теория глобального рабочего пространства, ключевую роль в наличии сознания она отводит лобным долям.

Ничуть не умаляя значимости и интереса, который представляют эти теории, больше я в своей книге к ним возвращаться не буду по той причине, что обе они выводят на передний план функциональные и поведенческие аспекты сознания, тогда как мой подход отталкивается от феноменологии — опыта как такового — и от нее уже перебрасывает мостик к функциям и поведению.

Определение сознания как «любого субъективного опыта во всех его видах» просто до банальности, но этим и ценно. Когда сложное явление понимается не до конца, преждевременно точные определения могут сковывать или увлекать нас в неверном направлении. История науки раз за разом демонстрировала, что полезные определения вырабатываются по мере появления научного понимания, выступая для научного прогресса подмостками, опорой, а не отправной или конечной точкой. В генетике, например, определение гена успело существенно измениться в ходе развития молекулярной биологии⁴. Точно так же, по мере того как будет углубляться и шириться наше понимание сознания, будет уточняться и его определение — или определения. Если сейчас мы условимся, что сознание — это прежде всего феноменология, можно будет переходить к следующему вопросу.

Как возникает сознание? Как связан сознательный опыт с биофизической машинерией нашего мозга и организма? Как он связан с круговертью атомов, или кварков, или суперструн, или из чего там в конечном итоге состоит наша вселенная?

В классической формулировке этот вопрос известен как «трудная проблема сознания». Впервые о ней заговорил австралийский философ Дэвид Чалмерс в начале

1990-х гг., и с тех пор именно она задает направление развития немалой части науки о сознании. Вот как описывает эту проблему сам Чалмерс:

Некоторые живые существа безусловно являются субъектами опыта. Но вопрос о том, как получается, что эти организмы становятся субъектами опыта, вызывает замешательство. Почему в том случае, когда наши когнитивные системы занимаются обработкой зрительной и слуховой информации, мы получаем зрительные или слуховые впечатления, воспринимая качество цвета (насыщенный синий) или звука (до первой октавы)? Как объяснить, почему существует возникновение мысленных образов или переживание эмоций? Общеизвестно, что опыт возникает на физической основе, но у нас нет внятного объяснения, почему и как он на ней возникает. С какой стати из физической обработки должны проистекать насыщенные внутренние переживания? Никакой объективной логике это не поддается, однако дело обстоит именно так⁵.

Чалмерс противопоставляет эту трудную проблему сознания так называемой легкой проблеме, точнее, легким проблемам⁶, связанным с необходимостью объяснить, как физической системе (мозгу) удастся генерировать те или иные функциональные и поведенческие свойства. К числу этих свойств относится обработка сенсорных сигналов, выбор действия и контроль поведения, сосредоточение внимания, речь и так далее. Круг легких проблем охватывает все способности таких существ, как мы, а также все, что можно обозначить как функцию (преобразование входящей информации в исходящий результат) или поведение.

Разумеется, эти «легкие» проблемы вовсе не так просты. На их решение у нейрочеловека уйдет не одно десятилетие, а может, и столетие. Чалмерс имел в виду, что они в принципе решаются легко — чего нельзя сказать о трудной проблеме. Точнее, Чалмерс не видит никаких концептуальных препятствий к тому, чтобы легкие проблемы в конце концов получили объяснение в терминах физических механизмов, тогда как трудной проблеме, судя по всему, подобные объяснения ничего не дадут. (Уточню, что под механизмом в данном случае подразумевается система причинно-взаимодействующих частей, вызывающих следствие.)⁷ И когда все легкие проблемы будут одна за другой вычеркнуты из списка нерешенных, к трудной проблеме по-прежнему будет не подступиться. «Даже когда мы объясним выполнение всех функций, которыми окружен чувственный опыт, — перцептивное различение, категоризацию, внутренний доступ, вербальный отчет, — вопрос о том, почему выполнение этих функций сопровождается чувственным опытом, все равно грозит остаться открытым»⁸.

Хотя зачатки трудной проблемы можно отыскать в древнегреческой философии или даже раньше, особенно отчетливо она проявляется в XVII в. у Рене Декарта, разделившего мир на духовный — *res cogitans* («вещь мыслящую») и материальный — *res extensa* («вещь протяженную»). Это разделение положило начало философии дуализма и сильно усложнило и запутало все последующие дискуссии о сознании. Самое наглядное свидетельство этой путаницы — изобилие разных философских концепций, в рамках которых ведутся рассуждения о сознании.

Запаситесь терпением, сейчас пойдут «-измы».

Моя любимая философская позиция и платформа, на которой по умолчанию стоят многие нейрочеловеки, — физикализм. Его основная идея заключается в том, что вселенная образована физической материей и состояния сознания либо идентичны определенным конфигурациям этой материи, либо каким-то образом из них проистекают. Некоторые философы предпочитают пользоваться термином «материализм», но мы будем считать физикализм и материализм синонимами⁹.

Диаметральной противоположностью физикализма выступает идеализм. Эта концепция (основоположником которой часто называют жившего в XVIII в. епископа

Джорджа Беркли) утверждает, что конечным источником действительности выступает сознание или разум, а не физическая составляющая или материя. Поэтому спрашивать нужно не о том, как из материи возникает разум, а о том, как из разума возникает материя.

Пристроившиеся между этими двумя полюсами дуалисты во главе с Декартом полагают, что сознание (разум) и физическая материя — это отдельные субстанции или режимы существования, взаимодействие которых — отдельный, довольно щекотливый вопрос. Сегодня мало кто из философов или ученых готов во всеуслышание объявить себя дуалистом. Но многих людей, по крайней мере на Западе, дуализм по-прежнему очаровывает. Из заманчивого интуитивного предположения, что сознательный опыт вроде бы должен быть нефизическим, рождается «наивный дуализм», в котором представление о том, как обстоит дело в действительности, диктуется таким вот «вроде бы». Как мы еще не раз увидим в последующих главах, попытки судить об устройстве мира по тому, каким он кажется, могут завести совсем не туда, куда надо.

Одно из особенно влиятельных течений физикализма носит название «функционализм». Как и физикализм, функционализм — распространенная и часто даже не требующая отдельных оговорок платформа для значительного числа нейроученых. Многие, кто считает физикализм само собой разумеющимся, также считают само собой разумеющимся и функционализм¹⁰. Я же, однако, по отношению к нему склонен к скептицизму и настороженности.

Главный постулат функционализма — сознание зависит не от физического строения системы, а от того, что она делает, какие функции выполняет, от того, как она преобразует входящие сигналы в исходящий результат. Функционализмом движет интуитивная идея, что разум и сознание — это виды обработки информации, которые могут обеспечиваться мозгом, но при этом биологический мозг для них не строго необходим.

Обратите внимание, как ловко и незаметно в это объяснение проник термин «обработка информации» (точно так же, как несколькими страницами ранее он появился в цитате из Чалмерса). Этот термин встречается в обсуждениях разума, мозга и сознания так часто, что ему ничего не стоит проскользнуть безотчетно, а этого лучше не допускать, поскольку за фразой «мозг обрабатывает информацию» скрываются довольно зыбкие предположения. В зависимости от того, кто их делает, они могут варьироваться от уподобления мозга компьютеру, в котором разум (и сознание) выступает программным обеспечением (или «разумным обеспечением»), до предположений о том, что, собственно, представляет собой информация как таковая. Все эти предположения опасны. Мозг сильно отличается от компьютера, по крайней мере от привычного нам¹¹. А вопрос, что такое информация, как мы еще убедимся в последующих главах, почти такой же скользкий, как и вопрос о том, что такое сознание. Именно поэтому я так настороженно отношусь к функционализму.

Если, по примеру многих, принимать функционализм некритически, можно увериться, будто сознание удастся каким-то образом симулировать на компьютере. Мы ведь помним, что, с точки зрения функционалистов, сознание зависит только от того, что делает система, а не от того, чем она образована. А значит, чтобы получить сознание, достаточно будет правильно задать функциональные отношения, то есть гарантировать правильное «отображение ввода-вывода». Иными словами, для функционалистов симуляция тождественна инстанцированию (воплощению, появлению в действительности).

Насколько это оправдано? В некоторых случаях симуляцию действительно можно расценивать как инстанцирование. Компьютер, играющий в го (как чемпион мира AlphaGo Zero, созданный британской компанией DeepMind, которая занимается

разработками в области искусственного интеллекта), действительно играет в го¹². Однако во многих ситуациях знак равенства между этими процессами поставить будет нельзя. Возьмем прогнозы погоды. Компьютерная симуляция метеосистемы, даже самая точная и подробная, не промокнет от дождя и не заколышется от ветра. На что больше похоже сознание — на го или на погоду? Ответа не ждите — его нет, по крайней мере пока. Достаточно и того, что у нас есть правильный вопрос¹³. Вот, собственно, откуда мой скептический настрой по отношению к функционализму.

Держитесь, мы почти закончили, осталось всего два «-изма».

Первый из них — панпсихизм. Его идея состоит в том, что сознание — это основополагающее свойство вселенной наряду с другими основополагающими свойствами, такими как масса/энергия и заряд, и оно присутствует в той или иной мере повсюду и во всем. Иногда панпсихизм высмеивают как якобы утверждающий, будто камень или ложка обладают сознанием, которое ни в чем не уступает нашему с вами, однако обычно это намеренное упрощение, призванное выставить концепцию в неприглядном свете. У нее есть гораздо более сложные версии (с некоторыми из них мы ознакомимся в следующих главах), но основные недостатки панпсихизма связаны не с тем, что все это слегка отдаст безумием, — в конце концов и безумные идеи иногда оказываются верными или по крайней мере полезными. Главный его недостаток в том, что он ничего, по сути, не объясняет и не выдвигает проверяемых гипотез. Это всего лишь отписка от загадки, которую нам задает трудная проблема, и, воспользовавшись этой отпиской, наука окажется в эмпирическом тупике¹⁴.

Замыкает список «-измов» мистерианизм, или мистерианство, — концепция, связанная с именем философа Колина Макгинна¹⁵. Мистерианство предполагает, что исчерпывающее физическое объяснение сознания — окончательное решение трудной проблемы Чалмерса, — может быть, и существует, но нам, людям, просто не хватает (и никогда не хватит) ума, чтобы это решение отыскать или хотя бы распознать, если его нам преподнесут на блюдечке сверхумные инопланетяне. Физическое понимание сознания существует, но оно настолько же непостижимо для нас, как для лягушек непостижима криптовалюта. Оно для нас когнитивно недостижимо в силу видоспецифичных умственных ограничений.

Что можно сказать о мистерианстве? Явлений, которых мы в силу ограниченности нашего мозга и разума не поймем никогда, в мире хватает. В частности, никто не понимает до конца, как летает Airbus A380 (однако это не мешает нам удобно устраиваться в его салоне, как довелось и мне во время возвращения из Дубая). Что-то безусловно остается когнитивно неподвластным подавляющему большинству обычных людей, даже перестав быть загадкой для человечества в целом (например, тонкости теории струн в физике). Поскольку ресурсы мозга как физической системы конечны и всегда найдется то, что не каждому мозгу доступно, неизбежно будут явления, которые при всей своей реальности никогда не поддадутся пониманию человека¹⁶. Однако было бы неоправданным пессимизмом заранее включать сознание в эту неизведанную область видоспецифичного непознаваемого.

Одно из самых замечательных свойств научного метода — его кумулятивный и поступательный характер. Сегодня многие из нас неплохо разбираются в том, что казалось совершенно непостижимым нашим предкам, а иногда и ученым с философами, работавшим каких-нибудь несколько десятилетий назад. Со временем одна тайна за другой раскрываются под воздействием систематической умственной работы и эксперимента. Если брать на вооружение мистерианство, можно сразу расписываться в собственном бессилии и расходиться. Поэтому брать его мы не будем.

Все эти «-измы» предлагают нам разные подходы к осмыслению взаимосвязей между сознанием и вселенной в целом. Взвешивая их достоинства и недостатки, следует

отдавать себе отчет, что важнее всего не «правильность» подхода с точки зрения доказуемой истинности, а то, насколько он полезен для нашего продвижения в понимании сознания. Именно поэтому мне больше всего импонирует функциональная агностичность физикализма. На мой взгляд, это самый прагматичный и продуктивный способ мышления применительно к науке о сознании. А еще, насколько я могу судить, самый честный в интеллектуальном отношении¹⁷.

Но несмотря на всю привлекательность физикализма, его адептами выступают далеко не все исследователи сознания. Чаще всего для критики этого подхода прибегают к так называемому мысленному эксперименту с «зомби». Зомби в данном случае — это не пожирающая мозг нежить из фильмов, это другие зомби, «философские». Но с ними точно так же придется бороться, поскольку иначе естественное физикалистское объяснение сознания погибнет на корню.

Философский зомби — это существо, неотличимое от сознательного, но лишённое сознания. То есть зомби Анил Сет будет выглядеть, как я, действовать, как я, двигаться и говорить, как я, но у него не будет представления о собственном бытии, не будет внутреннего мира и чувственного опыта. Спросите зомби Анилу, обладает ли он сознанием, и он ответит: «Да, обладаю». У него даже найдется несколько разных статей о нейронауке сознания, в том числе содержащих некоторые соображения о недоказанной связи философских зомби с этим вопросом. Однако ничто из этого совершенно не предполагает никакого сознательного опыта¹⁸.

Почему идея зомби выступает доводом против физикалистского объяснения сознания? Вот почему: если вы способны представить себе зомби, значит, способны вообразить мир, неотличимый от нашего, но лишённый сознания. А если вы можете вообразить такой мир, значит, физическим явлением сознание быть не может.

А вот почему эта логика не работает. Аргумент зомби, как и многие мысленные эксперименты, направленные против физикализма, — это аргумент представимости, а такие аргументы по природе своей слабы. Как и у многих из них, его вероятность обратно пропорциональна объему знаний, которым мы обладаем.

Вы можете вообразить Airbus A380, летящий задом наперед? Конечно, можете. Вам ничего не стоит мысленно нарисовать в небе огромный самолет, который движется словно в обратной перематке. А можете представить, чтобы такое происходило на самом деле? Чем лучше вы разбираетесь в аэродинамике и авиастроении, тем менее представимым будет для вас такой полет. В данном случае даже минимальных знаний вам хватит, чтобы заявить: самолеты хвостом вперед не летают. Это просто неосуществимо⁴.

То же самое и с зомби. С одной стороны, представить себе философского зомби проще простого. Вообразить себя самого без каких бы то ни было осознанных переживаний, вот и все. А на самом деле? Насколько это представимо? Ведь в действительности мне предлагается осмыслить весь объем и пределы возможностей огромной сети из многих миллиардов нейронов¹⁹ и тьмы-тьмущей синапсов (контактов между нейронами), не говоря уже о глиальных клетках, градиентах нейромедиаторов и других нейробиологических финтифлюшках, и все это внутри тела, которое взаимодействует с внешним миром, включающим в себя другие тела с другими такими же мозгами. Могу я такое представить? Кто-нибудь может? Сомневаюсь. Точно так же, как с аэробусом, чем больше мы знаем о мозге и его связи с сознательным опытом и поведением, тем труднее нам представить зомби⁵.

О представимости чего бы то ни было мы зачастую судим на основании психологических наблюдений за представляющим, а не постижения природы действительности. В этом слабость идеи зомби. Нас просят вообразить

невообразимое, а потом на основе этого иллюзорного осмысления делают выводы о пределах физикалистского подхода²⁰.

Теперь мы готовы познакомиться с настоящей проблемой сознания, как я ее называю. Этот подход к исследованию сознания формировался у меня на протяжении многих лет, прирастая открытиями и идеями многих других ученых²¹. Мне кажется, что именно обращение к настоящей проблеме позволит науке о сознании добиться успеха.

С точки зрения настоящей проблемы первоочередные задачи науки о сознании — объяснить, спрогнозировать и взять под контроль феноменологические свойства сознательного опыта. Это значит объяснить, почему тот или иной сознательный опыт именно таков — почему он обладает именно такими феноменологическими свойствами — в терминах физических механизмов и процессов мозга и организма. Эти объяснения позволят нам спрогнозировать моменты, в которые будет возникать определенный субъективный опыт, а также дадут возможность контролировать этот опыт за счет вмешательства в лежащие в его основе механизмы. Иными словами, для обращения к настоящей проблеме требуется не просто установить, что определенный паттерн активности мозга — или другого физического процесса — соответствует определенному виду сознательного опыта, но и объяснить, почему.

Настоящая проблема отличается от трудной проблемы тем, что не ставит целью (первоочередной по крайней мере) объяснить, почему и каким образом сознание вообще существует как часть нашей вселенной. Она не охотится за особым соусом, под которым можно каким-то чудом произвести сознание из простого механизма (или наоборот). Отличается она и от легкой проблемы (или проблем), поскольку в центре ее внимания находится феноменология, а не функции и поведение. Она не замечает субъективные составляющие сознания под ковер. И благодаря основному упору на механизмы и процессы настоящая проблема естественным образом согласуется с физикалистским взглядом на взаимоотношения между материей и разумом.

Чтобы подробнее разобраться в этих различиях, давайте посмотрим, как в рамках вышеупомянутых подходов можно было бы попытаться объяснить такое субъективное переживание, как красный цвет.

С точки зрения легкой проблемы наша задача — объяснить все механистические, функциональные и поведенческие свойства, связанные с восприятием красного: как активируют зрительную систему определенные длины волн спектра, при каких условиях мы говорим «этот предмет красный», как выглядит типичный отклик на сигналы светофора, каким образом красный может вызывать определенные эмоциональные реакции и так далее.

Этот подход по самой сути своей не подразумевает объяснения, почему и как эти функциональные, механистические и поведенческие свойства должны сопровождаться какой бы то ни было феноменологией — в данном случае феноменологией «красного». Объяснять существование субъективного опыта в противовес отсутствию опыта — это прерогатива трудной проблемы. Каким бы объемом механистической информации вы ни располагали, всегда будет резонно спросить: «Хорошо, но почему этот механизм связан с сознательным опытом?» Если проникнуться трудной проблемой всерьез, вы всегда будете подозревать разрыв между механистическими объяснениями и субъективным опытом «восприятия красного».

Настоящая проблема признает существование сознательного опыта и сосредоточивается в первую очередь на феноменологических свойствах. Например: восприятие красного зрительное; оно обычно, но не всегда привязано к предметам;

судя по всему, оно относится к поверхностям; оно может отличаться уровнем насыщенности; оно выделяется в особую категорию в цветовом восприятии, при этом может плавно варьироваться внутри этой категории — и так далее. Важно, что всё это свойства самого восприятия, самого чувственного опыта, а не функциональных качеств или поведения, связанного с этим восприятием, по крайней мере не в первую очередь. Основная задача настоящей проблемы — объяснить, спрогнозировать и взять под контроль эти феноменологические свойства с точки зрения происходящего в мозге и организме. Нам хотелось бы знать, что в тех или иных паттернах активности мозга (например, в сложной закольцованной активности зрительной коры[6]) объясняет (а также прогнозирует и контролирует) причины, по которым чувственный опыт, в частности восприятие красного, именно таков, каков он есть, и не является иным. Почему он отличен от синего, от зубной боли, от ревности.

Объяснение, прогнозирование и контроль. Вот критерии, по которым оценивается большинство других научных проектов независимо от того, насколько загадочными представляются поначалу объекты их исследования. Физикам удалось невероятно далеко продвинуться в распутывании тайн вселенной — объяснении, прогнозировании и взятии под контроль ее свойств, — хотя вопросы о том, из чего состоит вселенная или почему она существует, по-прежнему ставят их в тупик. Точно так же и наука о сознании может очень далеко продвинуться в попытках пролить свет на свойства и природу сознательного опыта без необходимости объяснять, как и почему он составляет часть вселенной, в которой мы обитаем.

Кроме того, не нужно думать, что научное объяснение непременно должно совпадать с интуитивными представлениями. Все мы знаем, насколько контринтуитивна квантовая механика в физике, но это не мешает признавать ее лучшим на данный момент объяснением природы физической действительности. Поэтому не исключено, что развитая наука о сознании точно так же позволит нам объяснять, прогнозировать и контролировать феноменологические свойства, не вызывая при этом торжества интуитивной догадки — «да, точно, конечно, именно так и должно было быть!».

Важно помнить, что настоящая проблема сознания — это не признание поражения перед лицом трудной проблемы. Настоящая проблема нацелена на трудную проблему — пусть не напрямую, но нацелена. Чтобы понять, почему это так, давайте познакомимся с «нейронными коррелятами сознания».

Меня по-прежнему поражает, какой парией была наука о сознании всего каких-нибудь 30 лет назад. В 1989 г., за год до того, как я поступил в Кембридж, психолог Стюарт Сазерленд написал: «Сознание — это феномен завораживающий, но призрачный. Невозможно определить, что это такое, что оно делает, почему оно развилось. Ничего стоящего внимания о нем не написано»²². Это уничижительное резюме, помещенное не где-нибудь, а в «Международном психологическом словаре», отражает то самое отношение к сознанию, с которым я часто сталкивался, когда делал свои первые шаги в науке.

Однако кое-где, вдали от Кембриджа, дела обстояли лучше, но я тогда об этом не подозревал. Фрэнсис Крик (открывший вместе с Розалинд Франклин и Джеймсом Уотсоном молекулярную структуру ДНК) и его коллега Кристоф Кох, трудившиеся тогда в калифорнийском Сан-Диего, разрабатывали метод, который в дальнейшем станет главным в становлении науки о сознании²³. Назовут его поиском нейронных коррелятов сознания.

Согласно эталонному определению, нейронный коррелят сознания, или НКС, — это «наименьшая совокупность нейрональных механизмов, которой будет достаточно для того или иного сознательного опыта перцепции»²⁴. Этот подход предполагает, что для каждого чувственного опыта, например «восприятия красного», существует

свой определенный паттерн нейронной активности. Когда эта активность присутствует, возникает чувственный опыт красного, а когда ее нет — не возникает. Великое достоинство подхода НКС состоит в том, что он дает практический рецепт проведения исследований. Чтобы идентифицировать НКС, достаточно придумать ситуацию, в которой люди будут в каких-то случаях испытывать тот или иной сознательный опыт, а в каких-то не будут, но по всем остальным параметрам эти ситуации должны совпадать как можно точнее. Создав такую ситуацию, вы затем сравниваете активность в мозге при двух различных условиях с использованием таких методов, как функциональная магнитно-резонансная томография (фМРТ) или электроэнцефалография (ЭЭГ)^[7]. Активность мозга, характерная для «сознательного» условия, будет содержать НКС для данного конкретного чувственного опыта²⁵.

Хорошим примером такой ситуации может служить «бинокулярное соперничество». В таком тесте каждому глазу предъявляется свое изображение: допустим, левому — изображение лица, а правому — дома. Однако сознательное восприятие не рождает в этом случае нелепую химеру «лицо — дом». Оно переключается туда-обратно от дома к лицу, задерживаясь по несколько секунд на каждом. Сперва вы видите дом, потом лицо, потом снова дом и так далее. Здесь важно, что сознательное восприятие меняется, хотя входящая сенсорная информация остается неизменной. Поэтому, глядя на происходящее в это время в мозге, можно отделить активность, которая сопровождает сознательное восприятие, от активности, сопровождающей ввод какой бы то ни было сенсорной информации. Активностью мозга, которая присутствует при сознательном восприятии, идентифицируется НКС для данного восприятия²⁶.

Стратегия НКС уже много лет приносит впечатляющие плоды, обеспечивая массу потрясающих открытий, но ее недостатки тоже проявляются все отчетливее. Один из них состоит в том, что очень трудно (а в конечном итоге, вероятно, невозможно) отделить «подлинный» НКС от ряда потенциально искажающих результаты факторов, самые важные из которых — те нейронные события, которые выступают либо предпосылкой, либо следствием самого НКС²⁷. В случае бинокулярного соперничества активность мозга, сопровождающая сознательное восприятие, может сопровождать и предшествующие (служащие предпосылками) процессы, такие как «обращение внимания», и, наоборот, последующие, такие как вербальный «отчет» — когда вы сообщаете, что видите лицо или дом. Хотя нейронные механизмы, отвечающие за внимание и вербальный отчет (или другие предпосылки и следствия), связаны с потоком сознательного восприятия, их не следует путать с теми, что отвечают за сознательное восприятие как таковое²⁸.

Еще более серьезная трудность заключается в том, что корреляция — это не объяснение. Мы все прекрасно знаем, что соответствие само по себе не означает причинно-следственной связи, но не стоит забывать, что и до объяснения оно тоже недотягивает. Какие бы хитроумные эксперименты ни разрабатывались и какие бы чудеса ни демонстрировала нам нейровизуализация, назначить корреляцию объяснением не получится. В этом смысле стратегия НКС и трудная проблема — два сапога пара. Если мы ограничимся сбором корреляций между происходящим в мозге и происходящим в нашем чувственном опыте, объяснительный разрыв между физическим и феноменологическим так никуда и не денется. Если же помимо установки корреляций мы будем искать объяснения, увязывающие свойства нейронных механизмов со свойствами субъективного опыта, как призывает настоящая проблема, этот разрыв сократится и, возможно, когда-нибудь исчезнет совсем. Когда мы сможем делать прогнозы, почему опыт восприятия красного будет именно таков и не похож на восприятие синего или на ревность (а также объяснять и контролировать), — загадка возникновения чувственного опыта красного будет уже не такой загадочной, а может, и утратит загадочность окончательно.

Честолюбивый замысел настоящей проблемы как научного подхода состоит в том, что наведение все более прочных и надежных объяснительных мостов между физическим и феноменологическим будет постепенно расщеплять интуитивное ядро трудной проблемы (идею о том, что сознание никогда не удастся объяснить с физической точки зрения), пока в конце концов оно не рассыплется и не исчезнет в клубах метафизического дыма. И тогда у нас в руках окажется удовлетворительная, полностью отвечающая всем нашим требованиям теория сознательного опыта²⁹.

На чем базируется этот замысел? Чем он оправдан? Давайте посмотрим, насколько выросло и окрепло за последние пару столетий научное понимание жизни.

Совсем недавно жизнь представлялась такой же загадкой, какой сегодня кажется сознание. Ученые и философы сомневались, что физическими или химическими механизмами в принципе удастся когда-либо объяснить такое свойство, как быть живым. Разрыв между жизнью и ее отсутствием, между одушевленным и неодушевленным, выглядел настолько фундаментальным, что считалось просто невыполнимым преодолеть его механистическими объяснениями.

Расцвет философии витализма пришелся на XIX в. Имея среди приверженцев выдающихся биологов, таких как Иоганн Мюллер и Луи Пастер, она не сдавала позиций и в XX столетии. Виталисты считали, что возникновение жизни можно объяснить, только ссылаясь на некую внешнюю силу — «искру жизни», *élan vital*³⁰. Но как мы сегодня знаем, этого внешнего вмешательства не требуется. Сейчас витализм в научных кругах не рассматривается. И хотя в том, что касается жизни, остается еще много неизвестного (как работает клетка, например), идея, что для появления жизни требуется некая сверхъестественная составляющая, полностью себя дискредитировала. Роковой ошибкой витализма было воспринимать недостаток воображения как указующую неизбежность. Эта же ошибка вкралась в основу аргумента зомби.

Наука о жизни сумела преодолеть близорукость витализма благодаря тому, что сосредоточилась на практическом прогрессе, то есть сделала упор на «настоящие проблемы» того, что означает быть живым. Биологи, не поддаваясь виталистическому пессимизму, продолжали описывать свойства живых систем, а затем объяснять (а также прогнозировать и контролировать) каждое из этих свойств в терминах физических и химических механизмов. Размножение, обмен веществ, рост, самовосстановление, развитие, гомеостатическое саморегулирование — все это вместе и по отдельности поддавалось механистическим объяснениям. По мере того как заполнялись пробелы и выяснялись подробности — этот процесс продолжается и сейчас, — не только сдавала позиции основная загадка того, «что такое жизнь», но и само понятие жизни разветвилось так, что «быть живым» уже не считается единым свойством, предполагающим «все или ничего». Выявились серые зоны, самая известная из которых — вирусы, но кроме них еще и синтетические организмы и даже скопления капель минерального масла, и каждая из этих систем обладает пусть не всеми, но некоторыми характерными свойствами живой. Жизнь освободили от элементов сверхъестественного, и от этого она стала только еще более невероятной. Эта параллель не только внушает нам оптимизм, но и подсказывает практическую стратегию работы с настоящей проблемой сознания.

Сегодняшние исследователи сознания могут быть сейчас в таком же положении, в котором находились биологи, изучавшие природу жизни, всего несколько поколений назад. Это воодушевляет. То, что кажется неразрешимой загадкой сейчас, не обязательно будет казаться таким всегда. По мере того, как мы будем продвигаться в объяснении различных свойств сознания в терминах механизмов, находящихся в их

основе, фундаментальная загадка «возникновения сознания», возможно, начнет проясняться точно так же, как постепенно прояснялась «тайна жизни».

Конечно, параллель между жизнью и сознанием нельзя считать абсолютной. Самое явное различие между ними в том, что свойства жизни поддаются объективному описанию, а предмет науки о сознании субъективен — он может быть описан только от первого лица. Однако это препятствие нельзя назвать непреодолимым, оно означает главным образом, что нужные данные в силу их субъективности труднее будет собрать.

Практическую стратегию нам дает понимание, что сознание, как и жизнь, — это не единое явление. Перестав рассматривать жизнь как одну огромную и страшную загадку, биологи стали менее склонны желать или требовать некоей эврики, гениального озарения, которое ответит сразу на все вопросы³¹. Вместо этого они разделили «проблему» жизни на ряд связанных между собой, но все же отдельных процессов. Воспользовавшись этой же стратегией применительно к сознанию, я в своей книге сосредоточусь на уровне, содержании и самости (нашем «я») как ключевых свойствах того, в чем состоит наше бытие собой. И в результате у нас появится удовлетворительная картина всего сознательного опыта.

Под уровнем сознания понимается «насколько мы сознательны» — на шкале от полного отсутствия сознательного опыта, как в коме или при смерти мозга, до ярких состояний осознанности, сопровождающих нормальное бодрствование.

К содержанию сознания относится то, что мы, собственно, осознаём, — зрительные образы, звуки, запахи, эмоции, настроение, мысли и убеждения, которые и составляют наш внутренний мир. Содержание сознания — это восприятие, интерпретация мозгом сенсорных сигналов, которые в совокупности образуют наш сознательный опыт. (Восприятие, как мы еще увидим, может быть как сознательным, так и бессознательным.)

Затем идет сознательная личность — определенный опыт бытия собой, ведущая тема этой книги. Опыт «бытия собой» — одна из составляющих содержания сознания, включающая в себя чувство обладания определенным телом, точку зрения от первого лица, набор уникальных воспоминаний, а также настроения, эмоции и ощущение «свободы воли». Наше «я», самость, — это, пожалуй, тот элемент сознания, за который мы цепляемся крепче всего, настолько крепко, что иногда велик соблазн подменить самоосознанием (опытом бытия собой) сознание как таковое (наличие любого субъективного опыта, любой, какой бы то ни было феноменологии).

Обособляя эти составляющие сознания, я вовсе не хочу сказать, что они существуют в отрыве друг от друга. Они взаимосвязаны, и выяснить, как именно, — это еще одна серьезная задача для науки о сознании.

Однако в нарезке настоящей проблемы сознания на эти обширные пласты есть множество преимуществ. Наметив отдельные объекты для исследования, мы облегчаем поиск механизмов, способных выполнять функции объяснения, прогноза и контроля. Немаловажно также, что в этом случае мы отмежевываемся от ограничивающего представления о сознании как о «едином» — единой, внушающей трепет загадке, которая может вовсе не поддаваться научному объяснению. Вместо этого мы посмотрим, как разные свойства сознания различным образом объединяются у разных видов и даже у разных людей. Способов быть сознательным существует столько же, сколько различных сознательных живых существ.

В конце концов может дрогнуть сама трудная проблема сознания, и мы начнем понимать сознание без отрыва от остальной природы, отменяя необходимость изобретать разные «-измы» с их взятыми с потолка догмами о том, как связаны феноменология и физика.

Вот какие перспективы рисует нам настоящая проблема. В дальнейших главах мы посмотрим, насколько она позволит нам продвинуться по намеченному пути.

Глава 1

Глава 2

Измерить сознание

Насколько вы сознательны? Сейчас, в данный момент? В чем состоит разница между наличием хотя бы какого-то сознания и существованием в виде простого куска живой плоти или безжизненного кремния без какого бы то ни было внутреннего мира? Новые теории и технологии впервые дают ученым возможность измерить уровень сознания. Чтобы лучше разобраться в этом новом направлении исследований, давайте начнем с его истоков.

Был в Париже XVII в. некий глубокий, темный и холодный подвал — прямо под обсерваторией на левом берегу Сены. Этому подвалу довелось сыграть в истории науки неожиданную роль, демонстрирующую важность измерений для научного прогресса.

Тогдашние философы и ученые (их, впрочем, еще даже учеными не называли) наперегонки разрабатывали надежный термометр, который позволил бы дать физическое объяснение природы теплоты. От популярных «калорических» теорий, представляющих теплоту субстанцией, способной вытекать из тел или втекать в них, постепенно отказывались. Но чтобы пересмотреть эти теории, необходимы были точные эксперименты, в которых можно было бы систематически оценивать, насколько предметы горячи или холодны. Для таких экспериментов требовалось и средство измерения «теплоты», чем бы она ни была, и шкала, на которой можно было бы сравнивать результаты. Поэтому и началось состязание — кто быстрее разработает надежный термометр и температурную шкалу. Но как проверять надежность термометра, если нет откалиброванной шкалы? И как разрабатывать шкалу, если нет надежного термометра?

Первым шагом к решению этой задачи было найти точку отсчета — некий эталон, обладающий неизменной температурой. Даже это оказалось нелегко. Перспективные кандидатуры (например, температура кипения воды) зависели от сторонних факторов, таких как давление, которое менялось в зависимости от высоты и погоды, и даже от таких нюансов, как гладкость поверхности стеклянного сосуда. И тогда на фоне всех этих неудач парижский подвал — с его вроде бы постоянным ровным холодом — какое-то время представлялся разумной фиксированной точкой температурного отсчета. (Это было не единственное необычное предложение. Самое, пожалуй, странное поступило от некоего Жоашена Далансе¹ — принять за отправную точку температуру плавления сливочного масла.)

В конце концов надежный и точный ртутный термометр все-таки был разработан, и тогда на смену калорическим теориям пришла новая наука термодинамика — этот переворот связан с такими легендарными фигурами, как Людвиг Больцман и лорд Кельвин. В термодинамике температура — это крупномасштабное свойство движения молекул внутри вещества, точнее, средняя кинетическая энергия молекул. Чем быстрее движение, тем выше температура. Под «теплотой» стали понимать энергию, передающуюся в процессе теплообмена между двумя системами, имеющими разную температуру. Важно отметить, что термодинамика не просто установила корреляцию средней кинетической энергии с температурой, а предположила, что именно это и есть температура. Теперь вооруженные новой теорией ученые могли говорить о температуре на поверхности Солнца и даже определить «абсолютный ноль», при котором теоретически прекращается всякое

молекулярное движение. Первые температурные шкалы, созданные на основе измерения характеристик определенных веществ (шкалы Цельсия и Фаренгейта), заменили шкалой, базирующейся на основополагающих физических свойствах (шкалой Кельвина, по имени лорда, открывшего законы термодинамики). Физическая основа температуры и теплоты перестала быть загадкой.

Обо всем этом я узнал из книги «Изобретение температуры» (Inventing Temperature: Measurement and Scientific Progress), написанной преподавателем Университетского колледжа Лондона историком Хасок Чаном². До тех пор я не представлял в полной мере, насколько научный прогресс зависит от измерений. История термометрии и ее влияние на наше понимание теплоты — яркий пример того, как возможность проводить подробные количественные измерения на шкале с фиксированными отметками превращает загадочное и таинственное в понятное и объяснимое.

Можно ли применить аналогичный подход к изучению сознания?³

Философы заговаривают порой о гипотетическом «счетчике сознания», который мог бы определять, имеется ли сознание у объекта, будь то другой человек, животное или машина. В 1990-х гг., во времена расцвета теории трудной проблемы, на конференции Дэвид Чалмерс приставил к голове старый фен, показывая, насколько полезен был бы такой прибор-счетчик, если бы он и в самом деле существовал. Наводите счетчик сознания на предмет и считываете результат. И не нужно больше гадать, насколько широк заколдованный круг сознания.

Между тем, как показывает история температуры, ценность измерения заключается не только в том, чтобы дать однозначный ответ о наличии или отсутствии свойства, но и в обеспечении возможности количественных экспериментов, которые могут в перспективе перевернуть научное понимание⁴.

Если сознание окажется чем-то вроде температуры, то есть если существует единый основополагающий физический процесс, тождественный «обладанию сознанием», мы совершим гигантский прорыв. Мы не только получим возможность определять «сознательность» того или иного объекта, мы сможем предметно говорить о конкретных «уровнях» и «степенях» сознания, а также о разновидностях сознания, далеко отстоящих от нашего скромного человеческого примера.

Но даже если история сознания будет развиваться иначе — не как с температурой, а, скорее, как с жизнью (что, на мой взгляд, более вероятно), — возможность выполнять точные измерения по-прежнему будет существенным шагом вперед в объяснении, прогнозировании и контроле природы субъективного опыта. При любом сценарии измерение превращает качественное в количественное, расплывчатое в точное.

Существует и практическая потребность в измерении сознания. Общая анестезия, которой сейчас подвергаются четыре с лишним миллиона пациентов в день⁵, призвана погружать человека во временное небытие, не пересекая грань, за которой сон превратится в вечный. Чтобы гарантированно удержаться на этой тонкой грани, очень пригодился бы надежный и точный счетчик сознания, тем более что к анестетику часто добавляют миорелаксанты — вещества, блокирующие нервно-мышечную передачу и вызывающие временную парализацию, чтобы мышечные рефлексы не препятствовали манипуляциям хирургов. И, как мы вскоре убедимся, не менее важно научиться определять, сохраняется ли сознание после серьезных повреждений мозга, при которых пациентам ставят страшные диагнозы «вегетативное состояние» или «состояние минимального сознания».

Собственно, мониторинг сознания по параметрам мозговой деятельности применяется в операционных уже много лет. Самый распространенный метод — мониторинг биспектрального индекса⁶. Подробности метода защищены патентом и

не раскрываются, но общий принцип его состоит в том, что на основе ряда электроэнцефалографических (ЭЭГ) показателей выводится постоянно обновляющееся число, на которое и ориентируется анестезиолог во время операции. Идея прекрасная, однако полностью на мониторы биспектрального индекса положиться не получается⁷, в том числе и потому, что были прецеденты, когда их показания не совпадали с поведенческими признаками сознания (в частности, пациенты открывали глаза или помнили, что говорили хирурги во время операции). Более глубокая проблема, касающаяся науки о сознании, заключается в том, что за биспектральным индексом не стоит никакая обоснованная теория.

В последние несколько лет начало формироваться новое поколение счетчиков сознания — уже не в операционных, а в нейробиологических лабораториях. В отличие от своих предшественников эти новые инструменты тесно связаны с зарождающимся теоретическим пониманием нейронной основы наличия сознания, и все они уже сейчас демонстрируют практическую ценность.

Измерять уровень сознания у человека — это совсем не то же самое, что определять, спит он или бодрствует. Уровень сознания не равен физиологической активности центральной нервной системы. И хотя они зачастую тесно коррелируют, сознание (осознанность) и бодрствование (физиологическая активность ЦНС) во многом различаются, отчетливо демонстрируя тем самым, что в их основе не могут находиться одни и те же биологические процессы. Когда вам снятся сны, вы по определению спите, но при этом получаете богатый и разнообразный сознательный опыт. На другом полюсе этой шкалы находятся катастрофические состояния, такие как вегетативное (называемое также синдромом безответного бодрствования), в котором человек по-прежнему проходит циклы сна и бодрствования, но не подает поведенческих признаков осознанности: свет в окнах периодически загорается, но дома никого нет. На схеме ниже показана зависимость между осознанностью и бодрствованием в самых разных условиях — как нормальных, так и патологических.

Чтобы отследить уровень сознания, нам нужно задаться вопросом: что в мозге обуславливает осознанность в противоположность простому бодрствованию? Может быть, все дело просто в числе задействованных нейронов? Судя по всему, нет. В мозжечке («маленький мозг», помещающийся в задней части коры) находится примерно в четыре раза больше нейронов⁸, чем во всем остальном мозге, однако в обеспечении осознанности они почти не участвуют. При таком редком заболевании, как недоразвитие мозжечка, человек, у которого мозжечок в результате далек от нормы, все равно может по большому счету вести вполне нормальную жизнь. И конечно, сомневаться в наличии у него сознания не приходится⁹.

Может быть, тогда дело в общей степени нейронной активности? Наверное, в сознательном состоянии мозг должен быть более активен, чем в бессознательном? В какой-то мере, пожалуй, да, — но ненамного. И хотя потребление энергии мозгом зависит от уровня сознания¹⁰, разница эта довольно незначительна и мозг точно не прекращает работу, когда сознание угасает.

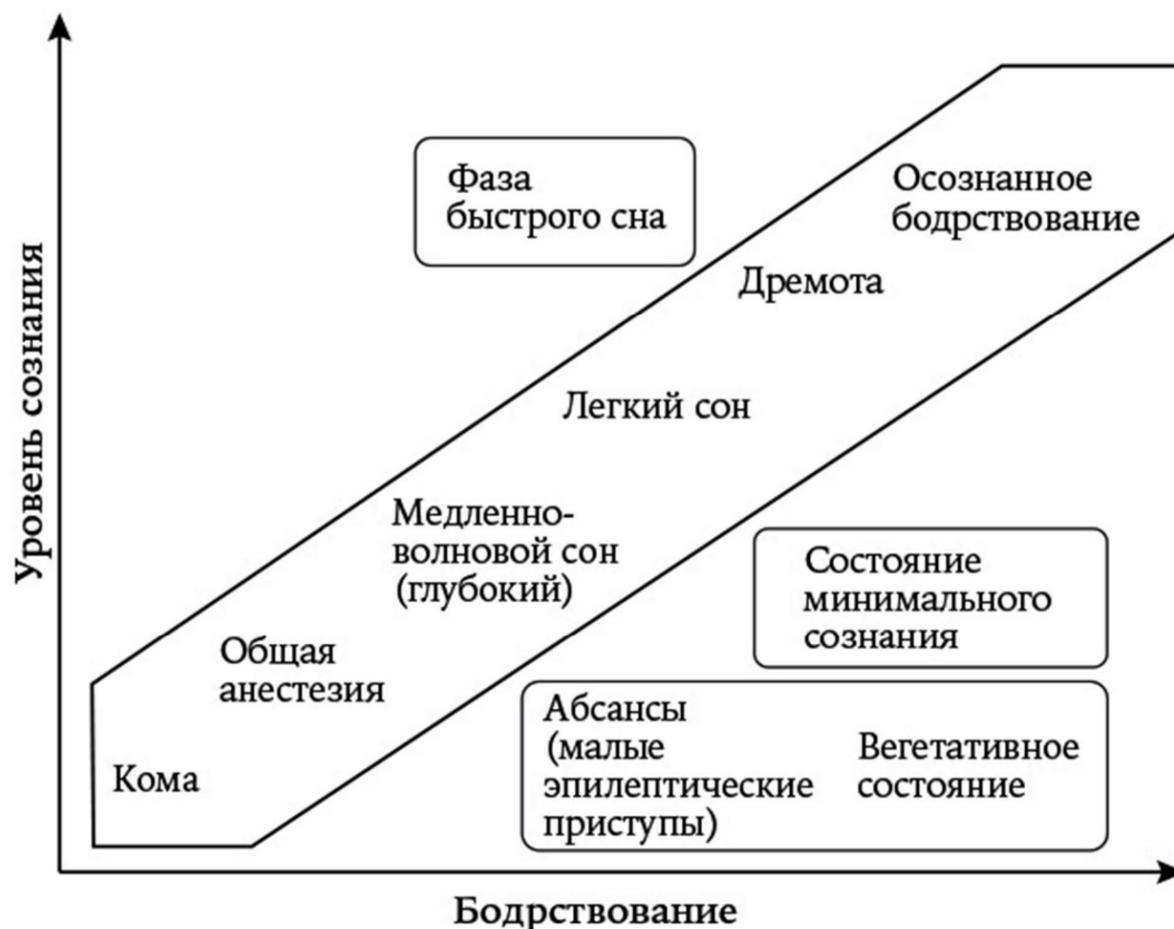


Рис. 1. Зависимость между уровнем сознания (осознанности) и бодрствованием (физиологической активностью ЦНС)

Судя по всему, сознание зависит не от деятельности мозга в целом, а от коммуникации разных его частей между собой. Значимыми выглядят паттерны активности таламокортикальной системы, объединяющей кору головного мозга и таламус (комплекс овальных структур мозга — «ядер», располагающихся прямо под корой и соединенных с ней замысловатой сетью связей). Новейшие и самые потрясающие подходы к измерению уровня сознания и поисков способа отличить его от бодрствования основываются на отслеживании и количественной оценке этого взаимодействия. Самая амбициозная вариация на тему этой идеи выдает единое число, обозначающее градус сознания у человека. Совсем как термометр.

Основоположником этого нового подхода стал итальянский нейробиолог Марчелло Массимини, работавший сначала с известным исследователем сознания Джулио Тонони в Висконсинском университете в Мадисоне, а затем со своей собственной группой в Миланском университете. Решение они нашли простое и элегантное. Чтобы проверить, как коммуницируют друг с другом разные отделы коры, они стимулировали активность в одной ее области и фиксировали пространственно-временное распространение этой активности в другие. Для этого они применяли две технологии — ЭЭГ и транскраниальную магнитную стимуляцию (ТМС). С помощью аппарата ТМС — точно контролируемого электромагнита — исследователь посылает короткий и резкий импульс непосредственно в мозг сквозь череп, а ЭЭГ позволяет зафиксировать реакцию мозга на этот импульс. Это примерно как ударить по мозгу электрическим молотком и слушать эхо.

Как ни удивительно, сам этот разряд ТМС испытуемый не чувствует, если только тот не провоцирует какую-нибудь реакцию, которую невозможно не заметить, например двигательную (когда магнит располагают над моторной корой, контролирующей движения) или мгновенное зрительное ощущение — «фосфен», случающееся, когда возбуждается зрительная кора. И если разряд вызовет спазм мышц лица или скальпа, вы этот болезненный укол заметите. Но в большинстве случаев сильная встряска мозговой активности, вызываемая ТМС, никаких изменений в сознательном опыте не вызывает. Впрочем, может, ничего удивительного здесь и нет. Просто свидетельство того, что мы не отдаем себе отчет в действиях наших нейронов — собственно, зачем бы нам его отдавать?

Но хотя самих импульсов ТМС мы не чувствуем, их электрическое эхо, как выяснили Массимини и Тонони, можно использовать для различения уровней сознания. В бессознательном состоянии, таком как сон без сновидений или общий наркоз, ответный импульс очень прост. Сначала в той части мозга, на которую направлен разряд, возникает сильный отклик, но он быстро пропадает, словно круги от камня, брошенного в стоячий пруд. В сознательном состоянии реакция выглядит совсем иной: типичное «эхо» раскатывается по всей поверхности коры, то пропадая, то появляясь вновь. Сложный пространственно-временной рисунок этих появлений и исчезновений¹¹ означает, что разные области мозга, в частности таламокортикальная система, коммуницируют друг с другом в сознательном состоянии гораздо более замысловатым образом, чем в бессознательном.

Хотя при сопоставлении данных разницу между двумя состояниями часто нетрудно увидеть невооруженным глазом, по-настоящему в этом подходе поражает именно то, что сложность ответного импульса поддается количественной оценке. То есть величину сложности можно обозначить неким числом. Этот метод называется zip-and-zip («разряд и сжатие») — ТМС воздействует разрядом (zip) на кору мозга, затем компьютерный алгоритм обрабатывает полученную реакцию, электрическое эхо, и сжимает ее (zip) в одно-единственное число.

При сжатии используется тот же алгоритм, что и при сжатии электронных фотографий, когда уменьшается объем файла. Любой рисунок, будь то фото из летнего отпуска или прокатывающееся по мозгу электрическое эхо, имеющее протяженность и длительность, можно представить как последовательность единиц и нулей. Для любой неслучайной последовательности есть возможность создать и сжатую версию¹², более короткую цепочку цифр, по которой затем несложно восстановить оригинал. Длина кратчайшего из возможных сжатых выражений называется алгоритмической сложностью последовательности. Самая низкая алгоритмическая сложность будет у полностью предсказуемой последовательности (состоящей из одних нулей или одних единиц), самая высокая — у совершенно случайной, и где-то посередине окажется сложность частично предсказуемых последовательностей. Установить алгоритмическую сложность для любой заданной последовательности позволяет алгоритм сжатия, вычисляющий так называемую сложность Лемпеля — Зива — Велча (или, если кратко, сложность ЛЗВ).

Массимини и его команда назвали способ замера ответного импульса, фиксируемого в ходе экспериментов, индексом сложности возмущений (ИСВ). Он дает возможность с помощью алгоритма ЛЗВ высчитать величину (индекс) алгоритмической сложности реакции мозга на возмущение, то есть на импульс ТМС.

Сперва группа Массимини подтвердила действенность своего метода, показав, что значения ИСВ в бессознательном состоянии, таком как сон без сновидений и общий наркоз, устойчиво ниже, чем при исходном сознательном состоянии бодрствования в покое. Это обнадеживает, но главная сила метода ИСВ заключается в том, что он дает непрерывную шкалу, на которую можно нанести более точную и подробную градуировку. В основополагающем исследовании 2013 г.¹³ группа Массимини измерила ИСВ у большой выборки пациентов с повреждениями мозга, страдающих расстройствами сознания. В результате было обнаружено, что величина ИСВ тесно коррелирует с уровнем нарушения функций мозга, независимо диагностированным неврологами. В частности, у пациентов в вегетативном состоянии, при котором сознание считается отсутствующим при сохранении бодрствования, показатели ИСВ получались ниже, чем у пациентов в состоянии минимального сознания, при котором поведенческие признаки сознания появляются и исчезают. Ученым удалось даже провести границу между значениями ИСВ, указывающими на наличие и на отсутствие сознания.

Мы с моей исследовательской группой в Сассексском университете работаем над схожими методами оценки уровня сознания. Но вместо того, чтобы посылать энергетический импульс в кору с помощью ТМС, мы измеряем алгоритмическую сложность имеющейся, естественной — спонтанной, как мы ее обозначаем, — мозговой активности. То есть вместо «запинга» получается «зиппинг». В ходе серии исследований, проведенных моим коллегой Адамом Барреттом и нашим бывшим докторантом Майклом Шартнером¹⁴, мы обнаружили, что сложность спонтанной активности коры, по измерениям ЭЭГ, устойчиво падает как при начальной стадии сна, так и при анестезии. Кроме того, мы выяснили, что во время фазы быстрого сна (фазы быстрого движения глаз) сложность спонтанной активности примерно такая же, как при нормальном сознательном бодрствовании, — вполне логично, поскольку именно в фазу быстрого сна наиболее вероятны сновидения, а они сознательны. Такую же картину выявила с помощью своего метода ИСВ группа Массимини¹⁵, еще надежнее подкрепив тем самым утверждение, что таким образом мы измеряем уровень сознания, а не бодрствования.

Возможность измерять уровень сознания независимо от бодрствования важна не только для науки, она способна в перспективе круто изменить положение дел и для неврологов и их пациентов. Исследование Массимини 2013 г. уже демонстрирует, что метод ИСВ способен различать вегетативное состояние и состояние минимального сознания. Главное преимущество таких замеров, как ИСВ, в этом контексте заключается в том, что они опираются не на внешние поведенческие проявления. Бодрствование — физиологическая активность центральной нервной системы — определяется через поведение. О бодрствовании неврологи в больнице обычно судят по наличию отклика на сенсорную стимуляцию — громкий звук или щипок руки. Сознание же определяется в терминах внутреннего субъективного опыта, поэтому связать его с внешними проявлениями можно лишь косвенно.

Стандартные клинические подходы к определению сознательного статуса у пациентов с повреждением мозга по-прежнему основываются на поведении¹⁶. Невролог, как правило, оценивает не только откликается ли пациент на сенсорную стимуляцию — маркер физиологической активности ЦНС, — но и способен ли взаимодействовать со средой, реагируя на команды или демонстрируя волевые действия. Когда пациент может выполнить двухсоставную просьбу и четко назвать свое имя и сегодняшнюю дату, мы говорим о полном сознании. Недостаток этого подхода в том, что некоторые пациенты могут при сохранении внутренней психической жизни не иметь возможности ее выразить. Поэтому в таких случаях при попытке судить на основании одного только поведения сознание будет ошибочно признано отсутствующим.

Самый яркий пример такого рода — синдром «запертого человека», когда при параличе всего тела полностью сохраняется сознание. Это редкое расстройство¹⁷ может возникать вследствие повреждения ствола мозга (отдела, расположенного в основании полушарий и продолжающего спинной мозг), который, кроме всего прочего, отвечает за контроль лицевых мышц и мускулатуры тела. Из-за некоторых причуд анатомии у «запертых» в своем парализованном теле пациентов остается способность выполнять ограниченный набор движений глазами, открывающая узкий и легко упускаемый из виду поведенческий канал для диагностики и коммуникации. Бывший редактор журнала Elle Жан-Доминик Боби, впавший в такое состояние в 1995 г. после геморрагического инсульта, написал с помощью сигналов, передаваемых морганием, целую книгу — «Скафандр и бабочка»¹⁸. У так называемых полностью запертых пациентов нет даже таких коммуникативных каналов, поэтому диагностировать у них наличие сознания еще труднее. Если судить только по поведению, ничего не стоит принять синдром

«запертого человека» за полное и постоянное отсутствие сознания. Но если поместить такого пациента, как Боби, в томограф, сразу станет видно, что общая активность мозга почти соответствует норме. В исследовании Массимини 2013 г. значения ИСВ у «запертых» пациентов не отличались от показателей у здоровых контрольных испытуемых аналогичного возраста, свидетельствуя о полностью сохранном сознании.

Еще труднее приходится врачам, имеющим дело с серой зоной между жизнью и смертью при таких состояниях, как вегетативное и минимальное состояние сознания. В этих пограничных областях поведенческие признаки сознания могут отсутствовать или быть непоследовательными, а повреждение мозга настолько обширным, что и нейровизуализация не гарантирует однозначного ответа. И вот здесь такие методы, как ИСВ, могли бы кардинально изменить положение дел. Когда показатели ИСВ у пациента указывают на наличие сознания, даже если все остальные признаки говорят об обратном, имеет смысл поискать это сознание еще раз.

Недавно Марчелло Массимини привел мне пример, когда все решил именно ИСВ. В миланскую больницу поступил молодой мужчина с тяжелыми травмами головы. Поскольку на простые вопросы и команды он не реагировал, у него диагностировали вегетативное состояние. Однако ИСВ у него оказался как у здорового человека в полном сознании, что вызвало у медиков полное замешательство, ведь синдром «запертого» у него не прослеживался. В конце концов персонал больницы отыскал дядю этого пациента, и тот приехал в Италию из Северной Африки, откуда молодой человек был родом. Когда дядя начал общаться с племянником на арабском, тот откликнулся мгновенно — улыбался шуткам и даже показал большой палец, когда ему включили фильм. Все это время он был в сознании — просто не отзывался на итальянский. Почему так произошло — трудно сказать. Массимини подозревает здесь необычный случай «культурного отчуждения»: итальянский мир словно утратил для пациента смысл¹⁹. Как бы то ни было, если бы не красноречивое электрическое эхо, измеренное с помощью ИСВ, история этого молодого человека могла бы закончиться совсем иначе.

Сейчас диагностирование остаточного сознания у пациентов с повреждениями мозга стремительно развивается. Помимо предлагаемого Массимини ИСВ существует несколько других методов, перебирающихся из лабораторий в клиники. Тот, которому отдаю предпочтение я, основывается на знаменитом в неврологических кругах эксперименте с «домом и теннисом», проведенном нейробиологом Адрианом Оуэном и его группой в 2006 г. В этом эксперименте пострадавшую в автомобильной аварии 23-летнюю пациентку, которая не подавала поведенческого отклика, уложили в аппарат фМРТ и выдали ряд вербальных заданий. То ее просили представить, что она играет в теннис, то предлагали вообразить, как она ходит по комнатам своего дома. На первый взгляд странный прием, ведь такие пациенты вообще ни на что не реагируют, что уж говорить о сложных устных инструкциях. Однако исследования с участием здоровых людей показали, что области мозга, участвующие в воображении плавных, перетекающих друг в друга движений (как при игре в теннис), сильно отличаются от активизирующихся при воображении перемещения в пространстве^[8]. И пациентка Оуэна продемонстрировала именно такую картину отклика в виде активности мозга, а это значит, что она тоже полностью выполняла задания, создавая высокоспецифичные мысленные образы. Невозможно представить, чтобы человек был способен на такое, находясь совершенно без сознания, поэтому Оуэн пришел к выводу, что поведенческий диагноз вегетативного состояния был ошибочным и на самом деле сознание у девушки наличествует²⁰. По сути, Оуэн с командой превратили аппарат МРТ в инструмент, позволяющий пациентке взаимодействовать с окружающей средой посредством мозга, а не тела.

Эта идея получила развитие в дальнейших исследованиях: метод Оуэна использовали не только для диагностики, но и для коммуникации. В проекте 2010 г. под руководством Мартина Монти пациент, поступивший с диагнозом «вегетативное состояние», смог отвечать на вопросы, представляя себе игру в теннис, чтобы дать ответ «да», и обход своего дома, чтобы дать ответ «нет». Трудоемкая коммуникация, что и говорить²¹, но для тех, у кого не осталось другого способа связаться с внешним миром, это настоящая революция.

Сколько еще таких людей, не откликающихся, но сохраняющих сознание, томится в неизвестности в неврологических отделениях и домах инвалидов? Трудно сказать. Метод Оуэна, появившийся раньше, чем ИСВ, исследовался чаще, и недавний его анализ²² позволяет предположить, что ту или иную форму скрытого сознания могут сохранять от 10 до 20% пациентов в вегетативном состоянии, что в пересчете на мировые масштабы будет означать несколько тысяч человек. И это, возможно, еще заниженные цифры. Чтобы пройти тест Оуэна, пациенту нужно понимать язык и быть в состоянии подолгу создавать мысленные образы, а на это, несмотря на наличие сознания, оказываются способны не все. И вот здесь новые методы вроде ИСВ особенно важны, поскольку гипотетически они позволяют обнаруживать остаточное сознание, не требуя от пациента вообще никаких действий. Именно так, как и положено подлинному счетчику сознания.

Пока я использовал термин «уровень» лишь применительно к довольно заметным изменениям у конкретного человека, таким как разница между обычным бодрствованием и общим наркозом или нахождением в вегетативном состоянии. Однако можно посмотреть на уровень сознания и под другим углом. Менее ли сознателен младенец, чем взрослый? И насколько сознательна по сравнению с ними обоими черепаха?

Конечно, в таком взгляде на проблему существует опасность: подобные рассуждения исподволь настраивают нас на то, что любая форма сознания, отличающаяся от сознания здорового взрослого человека, чем-то ниже или хуже. Такие установки симптоматичны для возведения человека на пьедестал и придания ему исключительности, от чего неоднократно страдала биология, поскольку подобными представлениями омрачена вся история человеческой мысли. У сознания не одно свойство, а значит, неверно принимать выражение определенного набора свойств, типичного для взрослого человека, за сущность сознания во всех его разновидностях, а затем именно это выражение принимать за точку отсчета для одномерной шкалы. Сознательный опыт, вне всякого сомнения, возникает не сразу — будь то при индивидуальном развитии (неважно, человека либо иного живого существа) или в грандиозных эволюционных масштабах. Однако было бы очень большой натяжкой рассматривать и тот и другой процесс как разворачивающийся линейно или завершающийся воплощенным во взрослом человеке идеалом того, что значит быть вами или мной. Это один из возможных недостатков и ограничений аналогии между сознанием и температурой²³ — той самой аналогии, с которой я начал эту главу.

Смежный вопрос: представляет ли собой сознание «все или ничего» («свет» либо горит, либо погас), или же четкой границы между наличием сознания и его отсутствием не существует? Этот вопрос относится как к появлению сознания в эволюционном или онтогенетическом смысле, так и к его появлению после возвращения из беспомощности общей анестезии или сна без сновидений. Вопрос, конечно, увлекательный, но поставлен он, на мой взгляд, неверно. В данном случае не стоит прибегать к противопоставлению «или — или». Применительно и к эволюции, и к индивидуальному онтогенетическому развитию, и к повседневной жизни, и к нахождению в медицинском учреждении я предпочитаю мыслить в терминах довольно резких переходов от полного отсутствия сознания к присутствию

по крайней мере какого-то сознательного опыта, который затем будет проявляться в разной степени или, возможно, разными гранями, но только после того, как «в окнах дома» появятся хотя бы какие-то проблески света.

Возьмем типичного взрослого человека. Будет его уровень сознания выше (или ниже) во сне или когда он сидит за рабочим столом слегка осоловелый после сытного обеда и мысли почти не ворочаются? Прямо и однозначно на такой вопрос ответить не получится. Сон может быть «более осознанным» в одном смысле (например, с точки зрения яркости и живости воспринимаемой феноменологии) и «менее осознанным» в других (например, с точки зрения рефлексивного осмысления происходящего^[9]).

Одно из важных следствий отказа от восприятия уровня сознания как одномерного — исчезновение резких различий между уровнем и содержанием сознания. В этом случае становится бессмысленно полностью отделять то, насколько мы сознательны, от того, в чем наше сознание состоит. Возможно, мерить всех одной меркой — а именно так предполагается по аналогии с температурой, если брать эту аналогию совсем буквально, — это не выход.

Наглядным примером тесной взаимосвязи между уровнем и содержанием сознания может служить проведенное нами несколько лет назад исследование активности мозга в психоделическом состоянии. Психоделические препараты помимо прочих своих многочисленных функций служат и науке о сознании, открывая перед ней уникальные перспективы, поскольку вызывают глубокие изменения в содержании сознания, не требуя при этом ничего, кроме простого фармакологического вмешательства в мозг.

Представление о том, насколько драматичными могут быть эти изменения, можно почерпнуть из записок изобретателя диэтиламида лизергиновой кислоты (ЛСД) — швейцарского химика Альберта Хофмана — о своей поездке домой из лаборатории расположенной в Базеле фармацевтической компании «Сандоз» 19 апреля 1943 г. В этот день, который вошел в анналы как «День велосипеда», он решил принять небольшую дозу своего свежего изобретения и вскоре начал ощущать себя несколько необычно, поэтому отправился на велосипеде домой. Каким-то чудом добравшись, несмотря на всевозможные шокирующие переживания, и уже уверившись, что сходит с ума, он улегся на диван и закрыл глаза.

...Понемногу я начал получать удовольствие от невиданных цветов и игры форм, упорно продолжающейся под моими сомкнутыми веками. Меня крутило в калейдоскопическом водовороте фантастических образов, которые чередовались, прорастали друг в друга, раскрывались кругами и спиралями, потом сжимались, взрывались цветными фонтанами, перетасовывались и соединялись, и все это без остановки, в постоянном течении...²⁴

В психоделическом состоянии яркие перцептивные галлюцинации часто сопровождаются необычным восприятием себя самого, часто описываемым как «растворение эго» — когда границы между собой, окружающим миром и другими людьми словно смещаются или растворяются. Эти отклонения от «нормального» сознательного опыта настолько глубоки и всеохватны, что, вполне возможно, психоделическое состояние представляет собой изменение не только содержания сознания, но и общего его уровня. Это предположение мы и взяли проверить совместно с Робинот Кархартом-Харрисом из Имперского колледжа Лондона и Сурешем Мутхукумарасвами из Оклендского университета.

В апреле 2016 г. мы с Робинот принимали участие в конференции, проходившей в аризонском Тусоне, прямо у подножия Каталинских гор. Нас обоих пригласили выступить с докладами о наших исследованиях, и мы воспользовались этой возможностью, чтобы выяснить, насколько наши с ним интересы в изучении сознания могут пересечься в контексте психоделии. Тогда как раз после нескольких

десятилетий опалы и забвения возобновились научные и медицинские исследования ЛСД, а также других психоделических соединений, таких как псилоцибин (активный компонент галлюциногенных грибов). За экспериментом Хофмана на самом себе последовал всплеск исследований, направленных на изучение потенциала ЛСД как лекарства от ряда психологических расстройств, включая алкоголизм и другие зависимости. Результаты выглядели многообещающими, но к концу 1960-х, когда ЛСД с легкой руки Тимоти Лири и прочих превратился в рекреационный наркотик и символ бунтарства, почти все эти исследования были закрыты. Сколько-нибудь существенные научные проекты начали появляться вновь только в 2000-х, после почти полувекового простоя²⁵.

На нейрохимическом уровне классические психоделики — ЛСД, псилоцибин, мескалин и диметилтриптамин (ДМТ, действующее вещество южноамериканского галлюциногенного отвара аяуаска) — работают преимущественно за счет воздействия на серотониновую систему мозга. Серотонин — один из главных нейромедиаторов, то есть тех химических веществ, которые обеспечивают общение нейронов между собой в нейронных сетях. Психоделические вещества влияют на серотониновую систему, связываясь с определенным рецептором к серотонину — 5-НТ_{2а}, широко распространенным по всему мозгу. Одна из главных задач исследования психоделических веществ — понять, каким образом фармакологическое вмешательство на низком уровне преобразует глобальные паттерны активности мозга, вызывая такие глубокие изменения в сознательном опыте.

Как удалось еще раньше обнаружить научной группе Робина, психоделическое состояние включает в себя заметные изменения в динамике мозга по сравнению с контрольным состоянием при получении плацебо²⁶. Нейронные связи между теми областями мозга, активность которых обычно скоординирована (так называемые сети состояния покоя), разъединяются, и подключаются другие области, в обычное время более или менее независимые. Иными словами, мы имеем дело с нарушением схем соединений, характерных для мозга в обычных условиях. Робин предположил, что этими нарушениями и объясняются типичные особенности психоделического состояния, такие как растворение границ между собой и миром и смешение чувств.

Мы с Робинем поняли, что собранные им данные как нельзя лучше подходят для проведения анализа алгоритмической сложности, который мы с моей научной группой в Сассексе применяли к снам и анестезии. Так, например, часть нейровизуализационных исследований его группа выполняла с использованием магнитоэнцефалографии (МЭГ), дающей высокое временное разрешение и охват всего мозга целиком²⁷, — именно то, что нам требовалось. С помощью МЭГ группа Робина измеряла активность мозга у добровольцев, принимавших псилоцибин, ЛСД или кетамин в малых дозах. (Если в больших дозах кетамин действует как анестетик, то в малых вызывает скорее галлюциногенный эффект.) И теперь, воспользовавшись этими данными, мы могли попытаться ответить на вопрос: что происходит с уровнем сознания, когда его содержание меняется так резко, как при психоделическом трипе?

Оставшиеся в Сассексе Майкл Шартнер и Адам Барретт просчитали изменение алгоритмической сложности сигнала МЭГ в различных областях мозга для всех трех психоделических состояний. Результаты получились совершенно четкими и абсолютно неожиданными²⁸: и псилоцибин, и ЛСД, и кетамин давали повышение уровня сознания по сравнению с контрольным состоянием при принятии плацебо. Такое было впервые: до тех пор еще никто не наблюдал повышения уровня сознания по сравнению с исходным уровнем бодрствования в покое. Во всех предшествующих сопоставлениях — с состоянием сознания во время сна, при анестезии, при расстройствах сознания — отмечалось только понижение уровня по сравнению с исходным.

Что же означают эти результаты? Как уже говорилось, измерение алгоритмической сложности, которое мы используем, лучше всего представлять как измерение случайности, или «многообразия», сигналов мозга, к которым этот метод применяется. Полностью случайная последовательность будет иметь высочайшую алгоритмическую сложность из возможных, то есть величайшее возможное разнообразие сигналов. Таким образом, наши результаты дополнили предшествующее исследование Робина²⁹, показывая, что активность мозга в психоделическом состоянии со временем становится более произвольной — в полном соответствии с теми беспорядочными изменениями перцептивного опыта, о котором так часто рассказывают пережившие трип. Помимо этого наши результаты проливают чуть больше света на взаимосвязь уровня и содержания сознания. Перед нами пример измерения уровня сознания при обширных изменениях содержания сознания, которыми характеризуется психоделия. Восприимчивость уровня сознания к изменению его содержания не оставляет сомнений в том, что эти две составляющие не независимы друг от друга.

Итоги нашего анализа психоделических состояний заставляли задуматься. Ведет ли максимально произвольная (согласно показателям алгоритмической сложности) активность мозга к максимальному психоделическому опыту? Или она дает другой (отличный от исходного) «уровень» сознания какого-то иного рода? Такая экстраполяция кажется маловероятной. Скорее всего, мозг со всеми его вспыхивающими как попало нейронами просто перестанет выдавать сознательный опыт точно так же, как свободный джаз в какой-то момент перестает быть музыкой. Однако алгоритмическая сложность — это не совсем то, что обычно понимают под «сложностью». Сложность в нашем интуитивном понимании никак не связывается с хаотичностью, тогда как на самом деле ее лучше представлять как нечто среднее между порядком и беспорядком, то есть не предельное выражение беспорядка. Это Нина Симон и Телониус Монк, а не Bonzo Dog Doo-Dah Band^[10]. Что произойдет, если руководствоваться вот таким, более научным пониманием сложности?

Ответ дают авторы статьи, вышедшей в 1998 г. в журнале Science, — Джулио Тонони и мой ныне покойный руководитель Джеральд Эдельман³⁰. Я и сейчас, 20 лет спустя, помню, как прочитал ее впервые. Она перевернула мое представление о сознании и, можно сказать, привела в Институт нейробиологии в Сан-Диего.

Вместо того чтобы, как предполагал подход «нейронных коррелятов сознания» (НКС), сосредоточиться на одном эталонном сознательном опыте, например «восприятию красного», Тонони и Эдельман задались вопросом: что характерно для сознательного опыта в целом? И пришли к простому, но глубокому выводу: сознательный опыт — любой сознательный опыт — отличается одновременно информативностью и интегративностью. Отталкиваясь от этого, они и делали дальнейшие утверждения о нейронной основе всего сознательного опыта, а не только конкретных переживаний вроде восприятия красного, чувства зависти или зубной боли.

Давайте немного проясним, что скрывается за формулировкой «одновременная информативность и интегративность сознания».

Начнем с информации. Что мы имеем в виду, называя сознательный опыт информативным? Эдельман и Тонони употребляли этот термин не в том смысле, в каком информативным может быть чтение газеты, а в том, который при кажущейся банальности таит в себе большое богатство. Сознательный опыт информативен, поскольку отличается от любого другого сознательного опыта, который у вас когда-либо был, будет или мог бы быть.

Я скольжу взглядом по своему рабочему столу и смотрю в расположенное за ним окно — и впервые получаю чувственный опыт именно этого сочетания кофейных чашек, компьютерных мониторов и облаков, которому сообщают еще большую уникальность все прочие перцепции, эмоции и мысли, присутствующие в данный момент в глубине моего внутреннего мира. В каждый конкретный момент мы получаем один-единственный сознательный опыт из огромного множества возможных. Таким образом, каждый сознательный опыт значительно снижает неопределенность, поскольку реализуется именно этот опыт, а не тот и не другой и так далее. А снижение неопределенности — это как раз информативность (в математическом смысле).

Информативность того или иного сознательного опыта не зависит от того, насколько он насыщен или подробен, или от того, насколько он познавателен для обладателя. Слушая Нину Симон за поеданием клубники на американских горках, мы исключаем ровно столько же альтернативных впечатлений, сколько исключим, сидя с закрытыми глазами в тихой пустой комнате и не испытывая почти ничего. Всякий опыт снижает неопределенность по отношению к множеству возможных впечатлений ровно настолько же, насколько любой другой.

С этой точки зрения пресловутое «что значит быть тем-то» любого конкретного сознательного опыта определяется не столько тем, чем он является, сколько всеми нереализованными, однако вероятными вариантами, которыми он не является. Впечатление чистого красного создается именно таким, какое оно есть, не в силу какого-то неотъемлемого свойства красного, а в силу того, что красный — это не синий, не зеленый и не любой другой цвет, не запах, не мысль, не чувство сожаления и вообще не какая бы то ни было другая форма психического содержания. Красный — это красный в силу всего, чем он не является, и то же самое относится к любому другому сознательному опыту.

Но одной лишь высокой информативности недостаточно. Сознательный опыт не только высокоинформативен, он еще и интегративен. Единого мнения о том, что именно понимать под «интегративностью» сознания, пока нет, но, по сути, это значит, что любой сознательный опыт возникает как единое, целостное впечатление. Мы не отделяем цвет от формы, а предмет от фона. Множество различных составляющих того сознательного опыта, который я испытываю в данный момент, — компьютеры и кофейные чашки, звук закрывающейся двери в коридоре, мои мысли насчет того, что сейчас написать, — фундаментально и необходимо связаны друг с другом как части единого всеобъемлющего сознательного впечатления³¹.

Ключевым шагом, который сделали Тонони и Эдельман, было следующее предположение: если каждый сознательный опыт является одновременно информативным и единым на феноменологическом уровне, то оба этих свойства должны быть присущи и нейронным механизмам в основе сознательного опыта. И одновременное выражение этих свойств закономерно предполагает, что нейронные механизмы не просто коррелируют с центральными феноменологическими чертами любого сознательного опыта, но, по сути, обуславливают эти черты.

Что означает для механизма эта одновременная интегративность и информативность? Давайте ненадолго отвлекусь от мозга и рассмотрим некую абстрактную систему, состоящую из большого количества взаимодействующих элементов (неважно каких). Как показывает иллюстрация ниже, для любой такой системы можно создать шкалу с двумя полюсами. На одном полюсе (левом) все элементы ведут себя независимо и случайно, как молекулы газа. Такая система будет максимально информативной — максимально случайной, но абсолютно не интегративной, поскольку все ее элементы независимы друг от друга.

На другом полюсе (правом) все элементы системы ведут себя совершенно одинаково, поэтому состояние каждого из них полностью определяется состоянием всех остальных. Никакой случайности. Аналогом будет расположение атомов кристаллической решетки, в которой позиция любого отдельного атома полностью определяется структурой решетки, образуемой позициями других атомов. Такая организация будет максимально интегративной, но почти неинформативной, поскольку состояний, в которых эта система может пребывать, окажется крайне мало.

В промежутке располагаются системы, в которых отдельные элементы могут действовать по-разному, но между собой они в той или иной степени скоординированы, поэтому система так или иначе ведет себя как «единое целое». Именно здесь мы обнаруживаем одновременно и интегративность, и информативность. И в этом случае возникает область пересечения порядка и беспорядка, к которой принадлежат системы, обычно называемые сложными.

А теперь перенесем эти принципы на устройство мозга и увидим, что они действительно проливают некоторый свет на нейронные основы сознания.

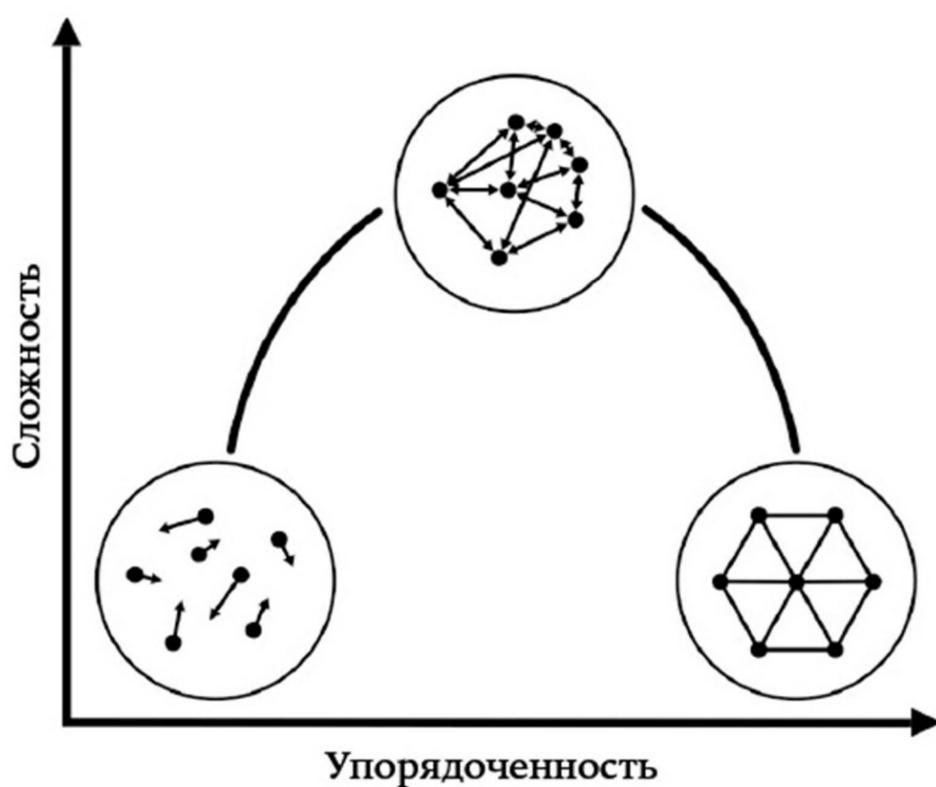


Рис. 2. Зависимость между сложностью и упорядоченностью

В максимально информативном мозге все нейроны действовали бы независимо друг от друга и выстреливали беспорядочно, словно абсолютно ни с чем не связанные. В таком мозге показатель алгоритмической сложности, допустим ЛЗВ, вышел бы очень высоким. Однако никаких состояний сознания этот мозг, обладающий высокой информативностью, но полностью лишенный интегративности, обеспечить бы не смог. На другом полюсе располагается максимально упорядоченный мозг, в котором все нейроны будут делать абсолютно одно и то же — возможно, выстреливать синхронно, единым залпом (примерно так и происходит во время эпилептического приступа). Алгоритмическая сложность здесь окажется очень низкой. Как и в предшествующем варианте, сознания в максимально упорядоченном мозге тоже не будет, но по прямо противоположной причине — высокая интегративность, полное отсутствие информативности.

Поэтому соответствующее своей задаче измерение уровня сознания должно отслеживать не информативность как таковую, а скорее совместное проявление информативности и интегративности. Такие замеры — измерение сложности в ее подлинном смысле — могли бы служить примером подхода к сознанию с позиций настоящей проблемы, напрямую связывая свойства механизма со свойствами опыта.

Как мы уже видели, аппроксимации алгоритмической сложности, такие как метод ЛЗВ, для этого не особенно годятся. Они много сообщают нам об информативности и почти ничего об интегративности. Метод ИСВ справляется несколько лучше. Чтобы

отметка на шкале ИСВ была высокой, энергетический импульс от аппарата ТМС должен вызвать такой паттерн активности мозга, который будет трудно сжать, то есть указывающий на высокую информативность. При этом импульс должен прокатиться по всей поверхности коры, порождая «эхо», сжимаемость которого тоже можно будет оценить. Но хотя это распространение по коре и говорит об интегративности, оно все-таки не удовлетворяет требованиям, которые мы предъявляем к измерению такого рода в идеале. Для измерения ИСВ активность мозга должна быть хотя бы сколько-то интегративной — иначе никакого эха не возникнет, — однако количественные замеры интегративности (в отличие от замеров информативности) метод ИСВ не производит. А мы ищем методы измерения, позволяющие непосредственно оценить и интегративность, и информативность по одним и тем же данным, одним и тем же способом и одновременно.

Этим критериям, по крайней мере теоретически, отвечают несколько методов. Еще в 1990-х гг. Тонони и Эдельман вместе со своим коллегой Олафом Спорнсом разработали метод измерений, получивший название «метод нейронной сложности»³², а 10 лет спустя я предложил собственный метод, основанный на другом способе вычислений, и назвал его «метод каузальной плотности»³³. На основе этих методов был создан ряд других, все более замысловатых, с частью которых мы познакомимся в следующей главе. Все эти методы представляют собой попытки так или иначе выразить количественно, где именно в той самой области пересечения порядка и беспорядка, к которой относится совокупность информативности и интегративности, располагается оцениваемая система. Проблема, однако, в том, что на реальных данных нейровизуализации все они показывают не самые лучшие результаты.

Создается довольно любопытная ситуация. Вполне резонно ожидать, что методы, крепче увязанные с теорией, на практике будут показывать себя лучше, чем такие, как метод алгоритмической сложности, у которых опора на основополагающую теорию довольно слаба. Однако ожидания эти, как мы видим, не оправдываются; так что же происходит? Можно предположить, что ошибочна сама теория. Однако я склонен считать, что нам просто нужно отшлифовать математическую основу³⁴, чтобы методы работали именно так, как нам хотелось бы, а также усовершенствовать нейровизуализацию, чтобы получать правильные данные для анализа.

Таким образом, поиски подлинного счетчика сознания продолжаются. Надо сказать, что прогресс на данный момент достигнут значительный. Уже общепризнано, что сознательный уровень — это не то же самое, что бодрствование, и у нас уже есть ряд его измерений, которые впечатляюще эффективно отслеживают разные общие состояния сознания, а также позволяют обнаруживать остаточную осознанность у пациентов с повреждениями мозга. Особенное значение здесь имеет метод ИСВ, предложенный Массимини. С одной стороны, он используется в клинических исследованиях, а с другой — твердо опирается на фундаментальные теоретические принципы информативности и интегративности и поэтому эффективно увязывает между собой нейронные механизмы и универсальные свойства сознательного опыта, как и предполагает «настоящая проблема». Между тем постоянно появляются и другие методы³⁵, основанные на иных, но связанных с методом ИСВ принципах, а простые в использовании аппроксимации, такие как оценка алгоритмической сложности данных спонтанной деятельности мозга, выявляют поразительные связи между уровнем и содержанием сознания.

И тем не менее основополагающий вопрос остается открытым. Сознание больше схоже с температурой, то есть сводимо к базовому свойству физического (или информационного) мира и может идентифицироваться через это свойство? Или оно больше похоже на жизнь — совокупность множества различных свойств, каждое из

которых имеет собственное объяснение с точки зрения находящихся в его основе механизмов? Подходы к измерению сознания, с которыми мы знакомимся до сих пор, равняются на историю с температурой, но интуиция подсказывает мне, что в конце концов им все-таки окажется ближе аналогия с жизнью. Я считаю интегративность и информативность главными свойствами большей части, а может быть, и всего сознательного опыта. Но это не значит, что сознание представляет собой интегрированную информацию, как температура представляет собой среднюю кинетическую энергию молекул.

Чтобы узнать, что движет этим интуитивным убеждением, нам придется довести аналогию между сознанием и температурой до предела и посмотреть, насколько она это выдержит. Пора нам познакомиться с «теорией интегрированной информации».

Глава 3

Ф

Июль 2006 г., я в Лас-Вегасе, ем джелато с Джулио Тонони. Мы в отеле «Венецианский», и я слабо понимаю, что происходит. Из Лондона я прилетел накануне, а в глубинах «Венецианского» всегда ранний вечер — фальшивые звезды разгораются на фальшивом лазоревом небе, под которым фальшивые гондолы скользят мимо фальшивых палаццо. Все для того, чтобы постояльцы и гости задерживались подольше и продолжали тратить деньги, пребывая в перманентном состоянии *aperitivo* и не следя за временем. Я слегка в тумане после перелета и слегка навеселе — под эти фальшивые звезды мы перебрались после долгого обеда, — и уже не первый час мы обсуждаем подробности невероятно перспективной «теории интегрированной информации», или ТИИ. Это детище Тонони, и оно замахивается на решение трудной проблемы куда смелее всех других нейронаучных теорий. ТИИ утверждает, что субъективный опыт — это свойство причинно-следственных паттернов; что информация так же реальна, как масса или энергия, и что даже атомы могут обладать некоторым сознанием¹.

Беспристрастности в нашей дискуссии мало. Я большую часть времени отстаиваю свою недавнюю статью², в которой критикую раннюю версию теории Джулио, а он в свою очередь мягко, но настойчиво пытается объяснить, где я ошибаюсь. Не знаю, что тому виной — разница во времени, выпивка или железная логика Джулио, но я уже не так уверен в себе, как во время перелета. На следующее утро я даю себе слово шевелить мозгами пошустрее, вникать глубже, готовиться тщательнее и пить меньше.

Тогда ТИИ меня покорила — и я продолжаю восхищаться ею до сих пор, потому что это образцовая аналогия между сознанием и температурой. Согласно ТИИ, сознание — это и есть интегрированная информация. Постулируя эту идею, теория переворачивает все глубоко укоренившиеся представления о взаимосвязях разума и материи и о том, как сознание вплетено в ткань вселенной.

В 2006 г. о ТИИ еще мало кто знал. Сегодня это одна из самых известных, но при этом вызывающая наиболее жаркую полемику теория в науке о сознании. Ее продвигают, помимо самого Тонони, выдающиеся специалисты в данной области. Кристоф Кох, бывший сторонник подхода нейронных коррелятов сознания, назвал ее «гигантским шагом к окончательному решению давней загадки тела и разума»³. Однако и несогласие эта теория в силу своей известности и амбициозности вызывает сильное. Одно из возражений — слишком большая математическая составляющая и непростительная сложность этого подхода. На самом деле это совсем не обязательно недостаток: никто не говорил, что разгадать загадку сознания будет легко. Другое возражение состоит в том, что постулаты этой теории слишком контринтуитивны, а

значит, она наверняка неверна. Но руководствоваться интуитивными представлениями, имея дело с таким труднопостижимым феноменом, как сознание, тоже рискованно⁴.

На мой взгляд, гораздо более серьезное препятствие для ТИИ заключается в том, что ее экстраординарные постулаты требуют экстраординарных доказательств, однако сама эта амбициозная задача — решить трудную проблему — делает ее основные утверждения непроверяемыми на практике. Необходимое экстраординарное доказательство просто не получится добыть. К счастью, еще не все потеряно. Как я объясню ниже, некоторые прогнозы ТИИ могут оказаться проверяемыми, по крайней мере в принципе. Кроме того, существуют альтернативные интерпретации ТИИ, больше ориентированные на настоящую проблему, а не на трудную, — они стоят за разработкой новых методов измерения уровня сознания, имеющих твердую теоретическую основу, но при этом применимых практически.

Как подсказывает название, в основе ТИИ находятся понятия информативности и интегративности. Теория опирается на те самые идеи об измерении сознательного уровня, с которыми мы уже встречались в предшествующей главе, однако переосмысливает их по-своему.

Ядро ТИИ — это величина «Ф» (греческая «фи»). Проще всего представлять себе Ф как показатель того, насколько система с точки зрения информативности «больше суммы» своих частей. Как система в принципе может быть больше суммы своих частей? В качестве примерной аналогии можно привести стаю птиц — стая кажется больше совокупности составляющих ее особей и как будто бы «живет собственной жизнью». ТИИ переносит эту идею в информационную область. Показатель Ф в ТИИ обозначает количество информации, порожденной системой «в целом» помимо и сверх количества, порождаемого ее частями по отдельности. На этой основе строится основной постулат теории — «система сознательна в той мере, в какой ее целое порождает больше информации, чем ее части»⁵.

Обратите внимание: это не просто утверждение о корреляции и не предположение в духе настоящей проблемы о механистических свойствах системы, которые стоят за феноменологическими свойствами. Это утверждение о тождестве. Согласно ТИИ, уровень Ф имманентен системе (то есть он не зависит от наблюдения извне) и тождествен количеству сознания, связанного с этой системой. Высокая Ф — много сознания. Нулевая Ф — никакого сознания. Поэтому ТИИ можно считать квинтэссенцией температурной аналогии представления о сознании⁶.

Что нужно, чтобы получить высокую Ф? Основная идея знакома нам еще из предыдущей главы, но есть некоторые важные различия, поэтому давайте начнем сначала.

Представьте себе сеть упрощенных искусственных «нейронов», каждый из которых может быть «включен» или «выключен». Чтобы обеспечить высокую Ф, сеть должна удовлетворять двум главным условиям. Во-первых, общее состояние сети — всей сети «как целого» — должно исключать большое количество вероятных альтернативных общих состояний. Это информативность, и в ней отражено феноменологическое наблюдение, что любой сознательный опыт исключает множество возможных альтернативных ему сознательных переживаний. Во-вторых, информативность при рассмотрении системы в целом должна быть больше, чем при делении ее на части (отдельные нейроны или группы нейронов) и рассмотрении всех этих частей независимо друг от друга. Это интеграция, и в ней отражено наблюдение, что весь сознательный опыт объединен и все его впечатления переживаются «как единое целое». Ф — это способ обозначить систему числом, показывающим, сколько она набирает по обоим параметрам.

Низкую Φ система может получить по многим причинам⁷. Одна из них — недостаточно высокая информативность. Самым простым примером тут служит единичный фотодиод — простой светочувствительный датчик, который может быть либо «включен», либо «выключен». У такого фотодиода Φ будет низким или нулевым, поскольку его состояние в любой момент почти никакой информации ни о чем не несет. Любое состояние, в котором он может находиться (единица или ноль, включен или выключен), всегда исключает только одну альтернативу (ноль или единицу). Единичный фотодиод передает, самое большее, один «бит» информации^[11].

Другая причина, по которой Φ у системы может оказаться низкой, — недостаточно высокая интегративность. Представьте большой комплект фотодиодов, допустим как в светочувствительной матрице камеры вашего телефона. Общее состояние системы — это состояние всего комплекта, и оно очень информативно. Большой комплект сенсоров будет менять общее состояние при каждом изменении состояния мира, с которым он взаимодействует, — в этом и есть польза фотокамер. Однако для самих диодов эта общая информация никакого значения не имеет. Отдельные фотодиоды в матрице независимы друг от друга — их состояние определяется только уровнем освещенности, приходящимся на каждый из них. Разделите матрицу на группы независимых фотодиодов, и они будут работать так же исправно⁸. Информация, передаваемая комплектом диодов как целым, не превысит передаваемую всеми диодами независимо друг от друга, то есть информация от всей системы не превысит информацию от входящих в нее частей, а значит, Φ и в этом случае окажется нулевой.

Еще один поучительный пример, когда Φ может равняться нулю, — ситуация так называемого расщепленного мозга. Представьте себе сеть, разделенную на две совершенно независимые половины. Каждая половина может обладать ненулевой Φ , однако у сети как целого Φ будет нулевой, поскольку при данном способе деления надвое целое не превысит сумму частей. Этот пример подчеркивает, насколько Φ зависит от оптимального деления системы, позволяющего минимизировать разницу между тем, что делает целое, и тем, что делают ее части. Это одна из особенностей ТИИ, отличающая ее от методов измерения сложности, описанных в предыдущей главе.

Из этого же примера следует, что в настоящем расщепленном мозге (образующемся после деления полушарий хирургическим путем при некоторых неизлечимых другими способами случаях эпилепсии⁹) могут обитать два независимых друг от друга «сознания», но единой сознательной общности, охватывающей оба полушария, не будет. Точно так же и нас с вами, хотя каждый из нас обладает сознанием, как единой сознательной общности, включающей в себя нас обоих, не существует, поскольку в информационном отношении нашу пару можно без труда разделить надвое.

Настоящим мозгом мы и займемся. Теория интегрированной информации очень четко объясняет уже имеющийся у нас ряд наблюдений об уровне сознания¹⁰. Как мы помним из предшествующей главы, мозжечок, на который приходится около трех четвертей нейронов мозга, судя по всему, не особенно участвует в обеспечении сознания. ТИИ может это объяснить, поскольку анатомия мозжечка сопоставима со светочувствительной матрицей камеры — это большое количество полунезависимых контуров. Кора больших полушарий, в противоположность ему, полным-полна плотно переплетенных между собой нейронных связей, которые, скорее всего, будут давать высокую Φ . Тогда почему же сознание угасает во время сна без сновидений, общей анестезии, комы, если эти связи остаются прежними? Согласно ТИИ, в таких состояниях нарушается способность нейронов коры взаимодействовать с остальными, ведя к исчезновению Φ .

ТИИ демонстрирует «аксиоматический» подход к сознанию. Она отталкивается не от экспериментальных данных, а от теоретических принципов. Аксиомы в логике — это самоочевидные истины, то есть признанные не требующими доказательств. Хороший пример такой аксиомы — евклидовский постулат «совмещающиеся друг с другом [фигуры] равны между собой». ТИИ выдвигает аксиомы, касающиеся сознания¹¹, — в первую очередь гласящую, что сознательный опыт одновременно информативен и интегративен, — и на каркасе этих аксиом выстраивает утверждения о том, какими свойствами должны обладать механизмы, находящиеся в основе этого опыта. Согласно ТИИ, любой механизм, обладающий этими свойствами, будь то мозг или что-то иное, биологическое или небιологическое, получит ненулевую Φ и у него будет иметься сознание.

Довольно о принципах. Как и в случае с любой теорией, выстоит ТИИ или рухнет, зависит от того, насколько проверяемы ее гипотезы. Основной ее постулат заключается в том, что уровень сознания той или иной системы задан ее Φ . Для проверки этого постулата нужно измерить Φ реальных систем — и вот тут возникает проблема. Как выясняется, измерить Φ крайне трудно, а в большинстве случаев почти или совершенно невозможно. Главным образом потому, что ТИИ трактует «информативность» не совсем так, как принято.

В математике информацию рассматривают так, как ее рассматривал в 1950-х гг. Клод Шеннон, — то есть относительно наблюдателя. Информация относительно наблюдателя (внешняя) — это степень снижения неопределенности с точки зрения наблюдателя при наблюдении за системой в конкретном состоянии. Представьте себе, например, что вы много раз бросаете игральную кость. Каждый раз вы наблюдаете реализацию одной из шести вероятностей, исключая тем самым пять остальных. Это соответствует определенной величине снижения неопределенности (величина эта измеряется в битах) и являет собой информацию, существующую «для» наблюдателя.

Чтобы измерить информацию относительно наблюдателя, обычно достаточно посмотреть, как ведет себя система на протяжении какого-то времени. В примере с игральной костью можно просто записывать цифру, выпадающую при каждом броске, — этого хватит, чтобы вычислить, сколько информации порождается при выкидывании каждого конкретного количества очков. Если же в качестве системы у нас нейронная сеть, достаточно будет фиксировать нейронную активность за какой-то период. Внешний наблюдатель может зафиксировать все разнообразие состояний, в которые впадут нейроны, вычислить вероятности для каждого из этих состояний, а затем измерить снижение неопределенности, связанное с пребыванием сети в любом из них¹².

Однако в ТИИ рассматривать информацию с точки зрения наблюдателя не получится. В этой теории информация (интегрированная информация, Φ) — это и есть сознание, и если брать ее относительно наблюдателя, значит, точно таким же, существующим относительно наблюдателя, нужно считать и сознание. Но оно не таково. Наличие или отсутствие у меня сознания не должно зависеть и не зависит от того, как вы или кто бы то ни было еще измерите мой мозг.

Следовательно, информацию в ТИИ нужно расценивать как внутреннюю, имманентно присущую системе, а не существующую относительно внешнего наблюдателя. Ее нужно определять как не зависящую от какого бы то ни было стороннего зрителя. Это должна быть информация «для» самой системы — не для кого-то или чего-то еще. Если это не так, то отношение тождества между Φ и сознанием, лежащее в основе ТИИ, распадается.

Чтобы измерить имманентную информацию, недостаточно просто понаблюдать, как ведет себя система на протяжении какого-то времени. Вам, как ученому, внешнему наблюдателю, нужно знать все различные варианты поведения системы, даже если на самом деле ее поведение реализуется не во всех. Здесь мы разделяем знание о том, как в действительности ведет себя система на протяжении времени (это узнать легко, по крайней мере в принципе, и это знание относительно наблюдателя), и знание о том, на что способна система гипотетически, даже если в действительности она этого не делает (это узнать обычно нелегко, а то и невозможно, но это безотносительное к наблюдателю знание).

На языке информатики разница между этими ситуациями — это разница между «эмпирическим» распределением состояний системы и распределением ее «максимальной энтропии» (название означает, что данное распределение отражает максимальный уровень неопределенности для данной системы). Представьте, что вы несколько раз кинули две игральные кости, и у вас выпало, допустим, семь, восемь, одиннадцать и еще несколько каких-нибудь чисел, но ни разу не получилось двенадцать. В таком случае эмпирическое распределение двенадцать очков содержать не будет, а распределение максимальной энтропии — будет, поскольку гипотетически двенадцать выпасть могло, даже если в данной конкретной последовательности бросков оно не выпало. Это значит, что любой конкретный результат — будь то семь, восемь или одиннадцать — породит больше информации (снизит больше неопределенности, исключит больше альтернатив) применительно к распределению максимальной энтропии (включающему двенадцать), чем применительно к эмпирическому распределению (двенадцать не включающему).

По сравнению с измерением эмпирического распределения системы путем простого наблюдения за ней в течение какого-то времени измерять распределение максимальной энтропии обычно очень трудно. Подступиться к такому измерению можно двумя способами. Первый — воздействовать на систему всеми возможными способами и смотреть, что получится, примерно как ребенок нажимает на новой игрушке все кнопки подряд и смотрит, что она умеет. Второй — вывести распределение максимальной энтропии из исчерпывающего, полного знания о физическом механизме системы, ее «причинно-следственной структуре». Если знать о механизме все, иногда можно выяснить, на что он способен гипотетически, даже если на практике этого не демонстрирует¹³. Зная, что у игральной кости шесть граней, мы, даже не утруждая себя необходимостью что-то бросать, легко вычислим, что на двух костях при каждом броске будет выпадать сумма от двух до двенадцати. К сожалению, зачастую нам доступна лишь динамика системы — то, что система делает в действительности, а не то, на что она способна. Это относится и к мозгу. Я могу зафиксировать, что делает мозг (с разной степенью подробности), но получить исчерпывающее знание о его физической структуре у меня не получится, и воздействовать на его активность всеми возможными способами я тоже не могу. Поэтому наиболее отличительный постулат ТИИ — что Φ и есть сознание — оказывается наименее проверяемым.

При попытках измерить Φ возникают и другие трудности, под каким бы углом мы ни рассматривали информативность. Одна из этих трудностей заключается в том, что для измерения Φ необходимо подобрать подходящий способ разделения системы на части, чтобы оптимально сравнивать их с «целым». Есть системы, такие как расщепленный мозг, с которыми проделать подобное довольно просто (разделить надвое посередине), но обычно это нелегкая задача, поскольку число вариантов деления возрастает в геометрической прогрессии с увеличением размеров системы¹⁴.

Затем возникают еще более фундаментальные вопросы о том, что в принципе считать системой. Где начинается та пространственно-временная структурированность, для которой можно вычислить Φ ? С нейронов и миллисекунд или с атомов и фемтосекунд? Может ли обладать сознанием целая страна и может ли одна страна быть сознательнее другой?¹⁵ А вдруг уже пора рассматривать взаимодействие тектонических плит на протяжении геологических периодов как интегрирование информации в планетарных масштабах?

Важно понимать, что эти трудности, в том числе измерение имманентной информативности, а не внешней, существующей относительно наблюдателя, — это трудности только для нас, ученых, внешних наблюдателей, пытающихся вычислить Φ . У самой же системы — любой системы — Φ , согласно ТИИ, заведомо имеется. И система делает свое дело, интегрируя информацию точно так же, как брошенный вами камень описывает в воздухе дугу, не нуждаясь в том, чтобы предварительно высчитывать траекторию с учетом закона тяготения. Если теорию трудно проверить, это еще не значит, что она ошибочна. Ее всего лишь трудно проверить.

Отвлечемся пока от проблемы измерения Φ и зададимся вопросом: что даст нам ТИИ, если она все-таки верна? Как выясняется, если довести ее до логического конца, нам придется столкнуться с очень необычными последствиями.

Представьте, что я вскрыл ваш череп и подсунул в мозг пучок новых нейронов, каждый из которых подключается к имеющемуся у вас серому веществу неким определенным образом. Дальше вы занимаетесь привычными делами, живете обычной жизнью, и эти нейроны — собственно, они не делают ничего. Ни при каких обстоятельствах, ни в каких ситуациях, чем бы вы ни занимались, что бы вы ни видели, они не активируются. Ваш дополненный мозг по всем параметрам и во всех отношениях остается точно таким же, как был. Фокус в том, что ваши новые нейроны могут возбудиться, только если мозг окажется в каком-то состоянии, в котором он никогда в действительности не оказывается.

Допустим, например, что эти новые нейроны активируются лишь в том случае, если вы съедите арбуз сорта «денсукэ» — редкое лакомство, которое можно найти только на японском острове Хоккайдо. И хотя в обычной жизни вы этим арбузом так и не угоститесь и новые нейроны так никогда и не будут задействованы, согласно концепции ТИИ, весь ваш сознательный опыт с приобретением этих нейронов все же изменится, пусть и очень незначительно. Произойдет это потому, что теперь у вашего мозга прибавится потенциальных состояний, в которых он может находиться, — ведь потенциально новые нейроны активироваться все же могут, — а значит, должна измениться и Φ .

Не менее парадоксально выглядит противоположный сценарий. Представьте, что в глубинах вашей зрительной коры притаился, ничем себя не выдавая, пучок нейронов. Хотя эти нейроны связаны с другими — а значит, потенциально способны возбуждаться, если получают правильный входящий сигнал, — они не делают ничего. Затем в результате какого-нибудь хитрого экспериментального вмешательства у них отнимают способность «включаться», то есть они уже не просто бездействуют, они инактивированы намеренно. И хотя общая активность мозга в этом случае не изменится, согласно ТИИ, мы вновь получим изменение сознательного опыта, поскольку число потенциальных состояний, в которых мозг может побывать, уменьшится.

Интересно, что благодаря новой технологии под названием оптогенетика, позволяющей исследователям с невероятной точностью управлять активностью выбранных нейронов, подобный эксперимент вскоре можно будет провести. Используя методы генной инженерии, оптогенетика позволяет модифицировать

отдельные нейроны, сообщая им чувствительность к свету определенной длины волны. И когда затем на генетически модифицированный таким образом мозг животного воздействуют светом при помощи лазера или ламп LED, эти нейроны «включаются» или «выключаются»¹⁶. В принципе оптогенетику можно применять для инактивации бездействующих нейронов и оценивать, как это отразится на сознательном опыте (если отразится). Эксперимент непростой, и способом измерения Ф он служить не может. Но перспектива проверить хотя бы какую-то составляющую ТИИ очень заманчива, и мне посчастливилось участвовать в недавних дискуссиях — с Джулио Тонони и другими — о том, как воплотить эту перспективу в жизнь¹⁷.

Если посмотреть шире, мы увидим еще одно непривычное следствие из ТИИ: уверенно заявляя, что Ф и есть сознание, она подразумевает, что информация как таковая существует, то есть имеет определенный онтологический статус в нашей вселенной, как масса/энергия или электрический заряд. (Онтология — это учение о сущем, наука о том, «что существует».) В каком-то смысле это представление согласуется с концепцией «всё из бита», выдвинутой физиком Джоном Уилером, пожалуй, самым известным пропагандистом идеи, что все сущее в конечном итоге происходит из информации, то есть информация первична, а из нее уже следует все остальное¹⁸.

И здесь нас поджидает последняя трудность — панпсихизм. Коль скоро у системы имеется правильный механизм, правильная причинно-следственная структура, Ф окажется ненулевой и система будет обладать сознанием. Панпсихизм ТИИ — это ограниченная версия панпсихизма, не та, согласно которой сознание распределено по всей вселенной, как намазанный тонким слоем джем. С точки зрения ТИИ сознание обнаружится там, где отыщется интегрированная информация — Ф. Она может отыскаться где угодно, в разных местах, но не повсюду.

ТИИ оригинальна, амбициозна и бурлит идеями. Она по-прежнему единственная из нейробиологических теорий всерьез замахивается на трудную проблему сознания. Она очень непривычна, но непривычна не значит ошибочна. В современной физике почти все непривычно и при этом гораздо ближе к истине, чем в физике прошлого. Однако успех тех составляющих современной физики, которые признаны более близкими к истине, обусловлен прежде всего тем, что они экспериментально проверяемы. В этом проблема ТИИ. За свою дерзость ей приходится расплачиваться тем, что ее главный постулат — тождество Ф и уровня сознания — может не поддаваться проверке.

На мой взгляд, чтобы двигаться вперед, лучше всего будет, не отказываясь от основополагающей идеи ТИИ, что сознательный опыт одновременно информативен и интегративен, отбросить заносчивую мысль о том, что Ф является для сознания тем же, чем средняя кинетическая энергия молекул для температуры. Тогда предположения ТИИ, касающиеся структуры сознательного опыта, снова будут согласовываться с концепцией настоящей проблемы. Эта позиция откроет перед нами перспективу разработки альтернативных, применимых на практике версий Ф — измерений, которые в конечном итоге во многом будут похожи на измерение сложности, с которым мы познакомились в конце предшествующей главы.

Именно такой стратегии мы с моими коллегами Адамом Барреттом и Педро Медиано придерживаемся уже много лет. Мы разработали несколько версий Ф, ориентированных на информацию относительно наблюдателя, а не на имманентную. Они дают нам возможность измерять Ф на основании поведения системы, наблюдаемого на протяжении какого-то периода, не беря в расчет то, что она гипотетически могла бы делать, но не делает. Сейчас положение дел таково, что все разработанные нами версии Ф показывают себя по-разному, даже на очень простых

моделях систем¹⁹. Это значит, что нам еще многое предстоит сделать в разработке практически применимых версий Ф, которые, мы надеемся, получают эмпирическую привязку благодаря, а не вопреки укорененности в теоретических принципах. С нашей точки зрения, это значит рассматривать «интегративность» и «информативность» как главные свойства сознательного опыта, требующие объяснения, а не как аксиоматический постулат о том, что на самом деле является собой сознание. Иными словами, считать сознание скорее «жизнью», чем «температурой».

Путешествуя по уровням сознания, мы побывали в забвении анестезии и комы, миновали пределы вегетативного и минимального состояний сознания и мерцающие миры сна и сновидений и вышли к солнечному свету полного бодрствования и даже дальше — в причудливую гиперреальность психоделии. Все эти уровни связаны между собой идеей, что каждый сознательный опыт одновременно интегративен и информативен и располагается где-то посередине между порядком и беспорядком. На основе этой идеи были разработаны новые методы измерения, такие как ИСВ, сочетающие практическую применимость с возможностью наводить объяснительные мосты в духе настоящей проблемы между физическим и феноменальным. Благодаря ТИИ мы вплотную подошли к будоражающей воображение и одной из самых дискуссионных областей науки о сознании, где дерзость замирает перед барьером проверяемости и где, видимо, рассыпается наконец аналогия между сознанием и температурой. Но несмотря на весь мой скепсис по отношению к более громким заявлениям этой провокационной теории, мне по-прежнему невероятно интересно, как она будет развиваться дальше²⁰, — так же, как было много лет назад, когда мы ели джелато с Джулио Тонони.

Сейчас, оглядываясь в то время, я думаю, что более подходящего места, чем Лас-Вегас, для той дискуссии и придумать нельзя было. Реальна ли информация? Везде ли есть сознание? В Лас-Вегасе с трудом верится в реальность чего бы то ни было, кроме остроты самих ощущений. Даже сейчас, годы спустя, я без труда возвращаюсь мысленно в этот нескончаемый ранний «венецианский» вечер, к фальшивым гондолам, скользящим по своим раз и навсегда заданным траекториям. Я совершенно точно в сознании, но что именно я сознаю? В отеле «Венецианский» легко верится, что все кругом сплошной мираж и галлюцинация.

Как мы скоро увидим, в этой мысли неожиданно есть доля истины.



Часть II
Содержание

Глава 4

Восприятие, идущее изнутри

Я открываю глаза, и возникает мир. Я сижу на веранде деревянной развалюхи в горном кипарисовом лесу в нескольких милях к северу от калифорнийского Санта-Круза. Раннее утро, день только начинается. Мой взгляд уткнулся в высокие деревья,

еще окутанные холодным туманом, который каждую ночь накатывает с океана, обрушивая столбик термометра. Земли не видно, так что и веранда, и деревья словно плавают вместе со мной в этом белом молоке. На веранде несколько старых пластиковых стульев (на одном из них я и устроился), стол и поднос с кофе и хлебом. Слышно пение птиц, какие-то шорохи за спиной (это мои соседи возятся в доме) и невнятное бормотание вдалеке, не знаю откуда и чье. Так бывает не каждое утро, сегодняшнее выдалось приятным. И я в который раз пытаюсь внушить себе, что этот потрясающий мир — порождение моего мозга, конструкт, что-то вроде «контролируемой галлюцинации».

Когда мы в сознании, мы что-то сознаем — что-то одно или множество всякого разного. Это содержание сознания. Чтобы разобраться, откуда оно возникает и что я имею в виду, говоря о контролируемой галлюцинации, давайте попробуем поменять точку зрения. Представьте на минуту, что вы мозг.

Попробуйте действительно поставить себя на его место и вообразить, каково это, будучи заключенным в костяной темнице черепа, пытаться выяснить, что происходит там, снаружи, во внешнем мире. К вам не проникают ни свет, ни звуки, ничего — вокруг кромешная тьма и гробовая тишина. Для формирования впечатлений есть только нескончаемый поток бомбардирующих вас электрических сигналов, которые лишь косвенно связаны с происходящим снаружи, чем бы оно ни было. На этих сигналах от органов чувств нет карточек с подписями «это от чашки кофе», «это от дерева». У них даже модальность никак не обозначена — зрительный это сигнал, слуховой, осязательный или принадлежащий какой-нибудь менее известной модальности вроде термоцепции (ощущение тепла и холода) или проприоцепции (ощущение положения тела в пространстве)[\[12\]](#).

Как мозг преобразует эти заведомо неоднозначные сигналы от органов чувств в связный перцептивный мир, наполненный предметами, людьми и местами? Во второй части этой книги мы рассмотрим идею мозга как «предсказывающей машины», согласно которой все, что мы видим, слышим и чувствуем, есть не что иное, как «наиболее вероятное предположение» мозга о причинах и источниках поступающих к нему сенсорных сигналов. Доведа эту идею до логического конца, мы увидим, что содержание сознания — это что-то вроде сна наяву, контролируемой галлюцинации, которая одновременно превосходит реальный мир и уступает ему.

Вот общепринятое представление о восприятии. Назовем его «кажущееся».

Нас окружает независимая от нашего сознания реальность, полная предметов, людей, мест, обладающих подлинными свойствами, такими как цвет, форма, текстура и так далее. Наши органы чувств служат прозрачными окнами в этот мир — они наводятся на все эти объекты, улавливают их свойства и передают информацию мозгу, а тот считывает ее в ходе сложных нейронных процессов и формирует восприятие. Чашка кофе во внешнем мире вызывает восприятие чашки кофе, создающееся внутри мозга. Кто или что при этом совершает акт восприятия? Ну так понятно же — «личность», «внутреннее я», которое принимает накатывающие одна за другой волны сенсорных данных, руководствуется их перцептивным выводом в своем поведении и решает, что делать дальше. Во внешнем мире есть чашка кофе. Я воспринимаю ее и беру в руки. Я чувствую, я думаю, затем действую.

Вдохновляющая картина. Представления, складывавшиеся десятилетиями, а может, и столетиями, приучили нас к мысли, что мозг — это такой заключенный в черепной коробке компьютер, который, обрабатывая данные от органов чувств, рисует внутренний образ внешнего мира, чтобы им пользовалось наше «я». Картина настолько знакомая и привычная, что вообразить сколько-нибудь логичную и разумную альтернативу нелегко. Собственно, многие нейрочеловеки и психологи и

сейчас представляют себе восприятие именно так — как процесс выявления признаков, направленный «по восходящей».

Согласно этой картине, процесс происходит так: стимулы из внешнего мира — световые волны, звуковые волны, молекулы, передающие вкусы и запахи, и так далее — воздействуют на органы чувств и побуждают их посылать электрические импульсы «вверх» к мозгу или «внутри» него. Эти сенсорные сигналы проходят несколько различных ступеней обработки, показанных на рисунке ниже черными стрелками, и на каждом этапе анализируются все более сложные признаки. Давайте возьмем в качестве примера зрение. На первых ступенях обрабатываются такие признаки, как яркость или общие контуры, затем — подробности, более мелкие детали, такие как глаза или уши, колеса, зеркала заднего вида. На еще более поздних стадиях анализируется объект как целое или как категория — лица, машины, кофейные чашки.

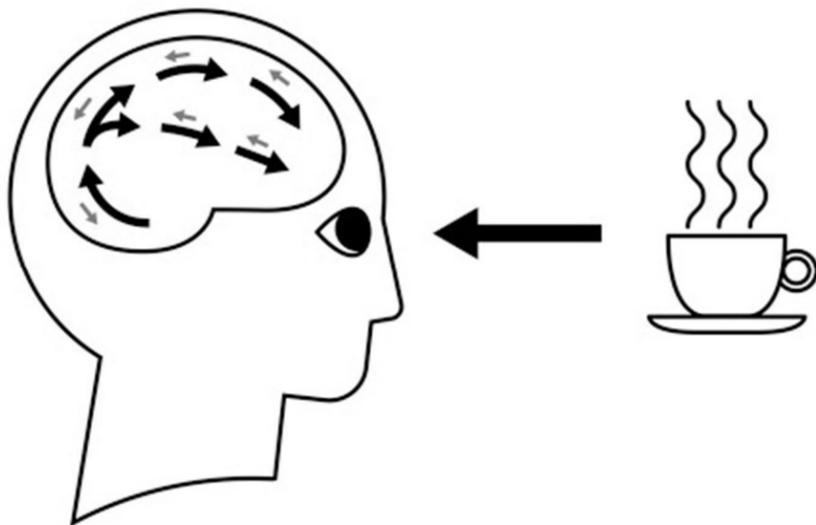


Рис. 3. Восприятие как восходящий процесс выявления признаков

Таким образом, внешний мир с его предметами, людьми и всем прочим резюмируется в виде серии признаков, извлеченных из льющегося в мозг потока сенсорных данных, словно рыба, которую рыбак ловит на удочку, и чем ниже по течению, тем этот улов крупнее и сложнее. Отмеченные на рисунке маленькими серыми стрелками сигналы, идущие в противоположном направлении — «сверху вниз» или «изнутри наружу», — служат только для того, чтобы уточнить или как-то еще сужать важнейший восходящий поток сенсорной информации.

Это представление о восходящем процессе восприятия отлично согласуется с тем, что нам известно об анатомии мозга, по крайней мере на первый взгляд. Каждая перцептивная модальность связана со своей областью коры — зрительной, слуховой и так далее. В каждой области обработка перцептивных данных организована иерархически. В зрительной системе нижние уровни, такие как первичная зрительная кора, располагаются ближе к источникам входящих сигналов, а более высокие уровни, такие как нижневисочная кора, находятся на несколько ступеней обработки дальше. Если смотреть с точки зрения связей и взаимодействия, сигналы от каждого уровня стекаются на следующий, вышестоящий, чтобы, как и следовало предполагать, нейроны на более высоких уровнях реагировали на признаки, возможно рассредоточенные во времени и пространстве¹.

На руку этим представлениям о восходящем процессе играют и исследования активности мозга. Проводившиеся на протяжении десятилетий эксперименты, в которых изучались зрительные системы кошек и обезьян, раз за разом показывали, что на ранних (низших) ступенях обработки зрительной информации нейроны реагируют на простые признаки, такие как общие очертания, а на более поздних (более высоких) — на сложные, такие как лицо. Схожие результаты применительно к человеческому мозгу демонстрируют относительно недавние эксперименты с использованием методов нейровизуализации, в частности фМРТ².

Так можно конструировать даже искусственные «системы восприятия», по крайней мере зачаточные. Предложенная специалистом по информационным технологиям Дэвидом Марром в 1982 г. классическая вычислительная теория зрения³ является одновременно стандартную отсылку к представлению о восприятии как о восходящем процессе и практическое пособие по разработке и конструированию искусственных зрительных систем. Более новые системы машинного зрения, применяющие искусственные нейросети, такие как сети «глубокого обучения», достигают сегодня впечатляющих высот в работе, в некоторых ситуациях сравнимых с тем, на что способен человек⁴. Эти системы тоже часто опираются на теории восходящего процесса восприятия.

При таком количестве свидетельств в пользу «кажущегося» представления о восприятии как о восходящем процессе еще бы оно не держалось на ногах так твердо!

Людвиг Витгенштейн: «Почему считается, что естественно полагать, будто Солнце вращается вокруг Земли, а не Земля вращается вокруг своей оси?»

Элизабет Энском: «Наверное, потому что все выглядит так, будто Солнце вращается вокруг Земли».

Людвиг Витгенштейн: «Хорошо, а как бы тогда все выглядело, если бы казалось, что Земля вращается вокруг оси?»

В этой восхитительной беседе⁵ со своей коллегой-философом (и биографом) Элизабет Энском легендарный немецкий мыслитель Витгенштейн на примере коперниканской революции иллюстрирует мысль, что кажущееся положение вещей совсем не обязательно совпадает с действительным. Хотя нам кажется, что Солнце вращается вокруг Земли, в действительности Земля вращается вокруг собственной оси, обеспечивая нам смену дня и ночи, а в центре Солнечной системы располагается Солнце, а вовсе не Земля. Так это же и так всем давно известно, скажете вы, — и будете правы. Но Витгенштейн мыслил глубже. Идея, к которой он подводил Энском, заключалась в том, что, даже когда мы знаем, как все обстоит в действительности, действительность по-прежнему будет казаться такой, как казалась. Солнце все так же встает на востоке и садится на западе.

То же самое и с восприятием. Я открываю глаза, и мне кажется, будто вокруг меня реальный мир. Сегодня я дома, в Брайтоне. Здесь нет кипарисов, как в Санта-Крузе, только обычный беспорядок на моем письменном столе, красное кресло в углу и вытянутые шеи дымовых труб за окном. Все эти объекты выглядят обладающими формой и цветом, а у тех, что поближе, вроде бы можно различить текстуру и запах. Таким представляется мир.

И хотя кажется, будто мои органы чувств служат прозрачными окнами в независимую от сознания реальность, а восприятие — это процесс «считывания и вывода» сенсорных данных, на самом деле, я убежден, происходит нечто совсем иное. Восприятие устремлено не снизу вверх или извне внутрь, оно движется преимущественно сверху вниз или изнутри наружу. Наш чувственный опыт строится на прогнозах мозга или «наиболее вероятных предположениях» о причинах и источниках сенсорных сигналов. Как и в случае с коперниканской революцией, это представление о восприятии как о нисходящем процессе, согласуясь со значительной частью имеющихся свидетельств, мало что меняет в видимом мире, однако на самом деле оно меняет все.

Идея эта отнюдь не нова. Первые проблески теории восприятия как нисходящего процесса появляются в Древней Греции, у Платона с его аллегорией пещеры. Узники в окопах, всю свою жизнь смотрящие на голую стену, видят лишь тени, которые на эту стену отбрасывают предметы, движущиеся за их спиной перед огнем, и они дают

этим теням имена, поскольку эти тени для них и есть реальность. Аллегория заключается в том, что наше сознательное восприятие, как и эти тени, представляет собой опосредованное отображение скрытых источников, непосредственно встретиться с которыми нам не дано.

Более тысячи лет спустя, но все равно очень давно, примерно за тысячу лет до нашего времени, арабский ученый Ибн аль-Хайсам писал, что восприятие здесь и сейчас зависит скорее от процессов «суждения и умозаключения»⁶, а не от непосредственного доступа к объективной действительности. Еще сотни лет спустя Иммануил Кант осознал, что хаос неограниченного потока данных от органов чувств останется бессмысленным, если его не будут структурировать уже имеющиеся представления и понятия, в число которых он включал такие априорные парадигмы, как пространство и время. Введенный Кантом термин «ноумен» относится к «вещам в себе» — Ding an sich — реальности, независимой от сознания, которая всегда будет недоступна человеческому восприятию, скрыта за сенсорной завесой.

В нейронауке эстафету принял немецкий физик и физиолог Герман фон Гельмгольц, внесший существенный вклад в научный прогресс. В конце XIX в. Гельмгольц выдвинул идею восприятия как «бессознательного умозаключения». Содержание восприятия, доказывал он, создается не сигналами от органов чувств как таковыми, — его нужно выводить как умозаключение, совмещая эти сигналы с ожиданиями или представлениями мозга об их источниках и причинах. Называя этот процесс бессознательным, Гельмгольц понимал, что мы не отдаем себе отчет в механике этих перцептивных умозаключений и учитываем только результат. Перцептивные суждения — «бессознательные умозаключения», по Гельмгольцу, — отслеживают и фиксируют свои источники во внешнем мире, активно и непрерывно обновляя наиболее вероятные перцептивные предположения по мере поступления новых сенсорных данных. По мысли Гельмгольца⁷, он подвел научное обоснование под кантовскую догадку, что восприятие не дает нам возможности знакомиться с предметами и явлениями окружающего мира напрямую, мы можем лишь строить умозаключения о том, что находится там, за сенсорной завесой.

На протяжении XX столетия оказавшаяся чрезвычайно влиятельной центральная идея Гельмгольца о «восприятии как умозаключении» принимала самые разные формы. В возникшем в 1950-е гг. направлении психологии «Новый взгляд» подчеркивалось влияние на восприятие социокультурных факторов. Например, как показало одно растиражированное впоследствии исследование, детям из бедных семей размер монет кажется больше, чем на самом деле, тогда как дети из состоятельных семей ошибаться в размере в большую сторону не склонны⁸. К сожалению, многие эксперименты такого рода при всей своей увлекательности проводились с точки зрения сегодняшних методологических стандартов некачественно, поэтому результатам их не всегда можно доверять.

В 1970-х психолог Ричард Грегори развил идеи Гельмгольца в несколько ином ключе, выстроив на них свою теорию восприятия как нейронную «проверку гипотез». Согласно этой теории, точно так же, как ученые проверяют и обновляют научные гипотезы на основании получаемых экспериментальных данных, мозг постоянно формулирует перцептивные гипотезы о том, как выглядит и устроен мир — на основании прошлого опыта и прочих форм накопления информации, — а потом проверяет эти гипотезы, получая данные от органов чувств. Перцептивное содержание, согласно теории Грегори, определяется наиболее подтвержденными гипотезами мозга⁹.

Оставшуюся половину века идея восприятия как умозаключения то завоевывала популярность, то отступала в тень, пока где-то в прошлом десятилетии не обрела новое дыхание. Родилась целая плеяда теорий, объединенная названиями «предиктивное кодирование» и «предиктивная обработка»: у каждой имеются свои

отличительные особенности¹⁰, но всех их роднит между собой предположение, что восприятие обуславливается теми или иными умозаключениями, делающимися в мозге.

Мою собственную интерпретацию бессмертной идеи Гельмгольца и ее современных воплощений точнее всего отражает представление о восприятии как о контролируемой галлюцинации — термин, который я впервые услышал много лет назад от британского психолога Криса Фрита^[13]. Основные постулаты теории контролируемой галлюцинации в моей трактовке я сейчас перечислю.

Во-первых, мозг постоянно строит предсказания насчет источников поступающих к нему сенсорных сигналов, и эти предсказания устремляются каскадом сверху вниз по иерархической лестнице его систем восприятия (серые стрелки на рис. 4). Если вы смотрите на чашку кофе, ваша зрительная кора будет формулировать прогнозы об источнике сенсорных сигналов, поступающих от этой чашки.

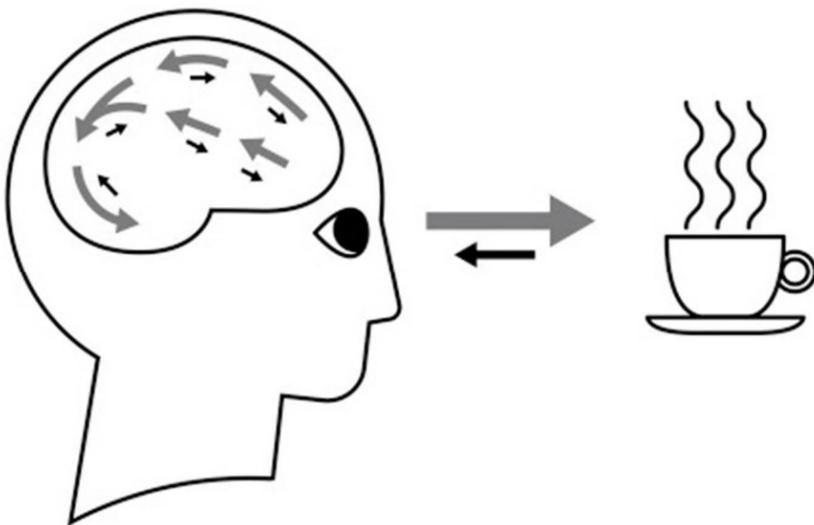


Рис. 4. Восприятие как нисходящий процесс умозаключения

Во-вторых, сенсорные сигналы, устремленные к мозгу снизу вверх или извне внутрь, полезным образом поддерживают привязку этих перцептивных прогнозов к их источникам (в данном случае к чашке кофе). Эти сигналы служат ошибками предсказания, обозначающими разницу между ожиданиями мозга и тем, что он получает на каждом уровне обработки данных. Корректируя нисходящие прогнозы так, чтобы они подавляли восходящие ошибки предсказания, наиболее вероятные перцептивные предположения мозга сохраняют связь со своими источниками во внешнем мире¹¹. С этой точки зрения восприятие представляет собой непрерывный процесс минимизации ошибок предсказания.

Третий и самый важный компонент теории контролируемой галлюцинации — это утверждение, что перцептивный опыт (в данном случае субъективный опыт «видения чашки кофе») определяется содержанием нисходящих предсказаний, а не восходящих сенсорных сигналов. Мы никогда не «испытываем» сами сенсорные сигналы, мы испытываем только их интерпретацию.

Смешайте эти три составляющие — получится коперниканский переворот в области представлений о восприятии. Нам кажется, что органы чувств раскрывают окружающий мир перед нашим сознательным разумом таким, какой он есть. С такими установками вполне закономерно считать восприятие процессом восходящего распознавания признаков — «считывания» окружающей нас действительности. На самом же деле мы воспринимаем нисходящую, направленную изнутри наружу нейронную фантазию, которую реальность просто обуздывает, а не смотрим на какую бы то ни было реальность сквозь прозрачное окно.

А как бы все выглядело — вернемся к Витгенштейну, — если бы восприятие представлялось нам наиболее вероятным нисходящим предположением? Да так же и выглядело бы. Как солнце по-прежнему восходит на востоке и заходит на западе, даже если бы восприятие казалось нам контролируемой галлюцинацией, чашка кофе

на столе — то есть совокупность чьего угодно перцептивного опыта — по-прежнему будет казаться такой, какой всегда была, и это не изменится.

Рассуждая о галлюцинациях, мы обычно представляем себе некое порождаемое чем-то внутри нас впечатление — зрительный или звуковой образ того, чего на самом деле нет, — как происходит при шизофрении или в психоделических приключениях, подобных тем, что пережил Альберт Хофман. В этих ассоциативных представлениях галлюцинации противопоставлены «нормальному» восприятию, которое предположительно отражает то, что существует в окружающем нас мире в действительности. В представлении о восприятии как о нисходящем процессе противопоставленные явления превращаются в обычные разновидности. И «нормальное» восприятие, и «ненормальные» галлюцинации включают в себя порождаемые чем-то внутри нас прогнозы относительно источников входящих сенсорных данных, и обе разновидности образуются одним и тем же комплектом основных механизмов в мозге. Разница лишь в том, что при «нормальном» восприятии воспринимаемое связано с источниками в реальном мире и контролируется ими, а в случае галлюцинаций восприятие в той или иной степени теряет связь с этими источниками. При галлюцинациях наши перцептивные прогнозы не обновляются должным образом с учетом ошибок прогнозирования.

Если восприятие — это контролируемая галлюцинация, то галлюцинацию по аналогии можно считать неконтролируемым восприятием. Они отличаются, но искать границу между ними — это все равно что искать границу между днем и ночью.

Давайте проверим теорию контролируемой галлюцинации на практике и выясним, что означает, допустим, перцептивный опыт цвета.

Наша зрительная система, сколь бы удивительной она ни была, реагирует только на крохотный отрезок полного электромагнитного спектра, расположенный между низкими частотами инфракрасного излучения и высокими частотами ультрафиолетового. Любой воспринимаемый нами цвет — собственно, любая часть совокупности каждого из наших зрительных миров — основывается на этом тонком срезе действительности. Уже одно это знание говорит нам о том, что перцептивный опыт не может выступать всеобъемлющей репрезентацией объективного внешнего мира. Он недотягивает до нее и одновременно превосходит ее.

Спросите нейропсихолога, и он, скорее всего, скажет, что вы воспринимаете тот или иной цвет, когда в сетчатке вашего глаза активируется определенная доля цветочувствительных колбочек. Этот ответ нельзя назвать неверным, однако он рисует далеко не полную картину. Между активностью цветочувствительных колбочек и восприятием цвета взаимно однозначного соответствия нет.

Воспринимаемый вами цвет определяется сложным взаимодействием света, отражающегося от поверхности, и общей освещенности той среды, в которой вы находитесь. Точнее, он определяется тем, как ваш мозг выводит умозаключения — наиболее вероятные предположения — о том, как складывается это взаимодействие.

Вынесите лист белой бумаги на улицу — он по-прежнему будет выглядеть белым, хотя свет, который он теперь отражает, имеет совсем другой спектральный состав в силу разницы между дневным солнечным светом (синеватым) и искусственным (желтоватым). Ваша зрительная система автоматически компенсирует эту разницу в общей освещенности — «делает поправку на источник света», как говорят исследователи зрения, — в результате чего ваше ощущение цвета распознает инвариантное свойство листа бумаги — то, каким образом бумага отражает свет. Мозг делает умозаключение об этом инвариантном свойстве, выдвигая наиболее вероятное предположение об источнике его постоянно меняющихся входящих

сенсорных данных. Белый — это феноменологическая составляющая данного умозаключения, проявление умозаключения мозга об этом инвариантном свойстве в нашем сознательном опыте.

Соответственно, цвет не является определенным свойством вещей в себе. Он, скорее, представляет собой найденное эволюцией полезное устройство, позволяющее мозгу распознавать и отслеживать предметы в условиях меняющейся освещенности. Когда я испытываю субъективный опыт наблюдения красного кресла в углу своей комнаты, это не значит, что кресло на самом деле красное, — как вообще кресло может обладать таким феноменологическим свойством, как краснота? Краснота не присуща креслу, как не присуще ему уродство, старомодность или авангардность. Просто у поверхности кресла имеется свойство — определенным образом отражать свет, — и мой мозг это свойство отслеживает с помощью своих перцептивных механизмов. Краснота — это субъективная, феноменологическая составляющая данного процесса.

Значит ли это, что краснота кресла переместилась из присутствия «там», в окружающем мире, в присутствие «здесь», внутри мозга? В некотором смысле ответом будет безоговорочное «нет». В мозге нет никакой красноты в том смысле, что там имеется некий воображаемый красный цвет, запечатлеваемый чем-то вроде миниатюрной видеокамеры, которая передает снятое в еще одну зрительную систему, в которой тоже имеется своя мини-камера, и так далее. Полагать, что для возникновения восприятия воспринимаемое свойство внешнего мира (краснота) должно как-то воспроизводиться в мозге, значит совершать ошибку «двойного преобразования», как ее называл философ Дэниел Деннет. В рамках этой ошибки внешняя «краснота» преобразуется сетчаткой в паттерны электрической активности, которые затем необходимо воспроизвести, то есть снова преобразовать, во внутреннюю «красноту». Как отмечает Деннет, такой ход мысли не объясняет ничего¹². Единственный смысл, в котором красноту можно обнаружить «в мозге», — тот, согласно которому именно там находятся механизмы, составляющие основу перцептивного опыта. И эти механизмы, разумеется, не красные.

Когда я смотрю на красное кресло, ощущение красного, которое я испытываю, зависит и от свойств самого кресла, и от свойств моего мозга. Оно соответствует содержанию набора перцептивных предположений о том, как определенный вид поверхности отражает свет. Красноты как таковой не существует — ни в мире, ни в мозге. Как сказал Поль Сезанн, «цвет — это точка, в которой мозг соприкасается со вселенной»¹³.

Если взглянуть шире, можно заявить, что этот принцип применим не только к области цветового восприятия. Он относится к любому восприятию. Иммерсивная мультисенсорная панорама вашей перцептивной сцены здесь и сейчас, в данный момент, — это выход на связь с внешним миром, это скорее письмо, чем чтение. Совокупность перцептивного опыта — это нейронная фантазия, которую удерживает в узде постоянный поток сменяющихся друг друга наиболее вероятных перцептивных предположений, контролируемых галлюцинаций.

Можно сказать даже, что все мы постоянно галлюцинируем. Просто, когда мы сходимся во мнении относительно этих галлюцинаций, мы называем их реальностью¹⁴.

Давайте посмотрим, как перцептивные ожидания формируют сознательный опыт, — возьмем три примера, которые вы можете проверить на себе.

Если в феврале 2015 г. вы пользовались соцсетями или читали газеты, то наверняка помните ажиотаж вокруг «феномена синего или белого платья». В одну февральскую среду по прибытии на работу на меня обрушилась лавина электронных писем и голосовых сообщений. Незадолго до этого я выступил соавтором небольшой книги о

зрительных иллюзиях¹⁵, а СМИ наперегонки искали объяснения взорвавшему интернет явлению. «Феномен» воплотился в случайно сделанном фотоснимке некоего платья, которое кому-то казалось сине-черным, а кому-то бело-золотым^[14]. Видевшие его в одном цветовом сочетании были настолько уверены в своей правоте, что попросту не представляли, как можно видеть что-то иное, вот интернет и забурлил доводами и возражениями.

Сперва я подумал, что это розыгрыш. Я, как и те первые четверо сотрудников лаборатории, которым я показал картинку, совершенно отчетливо видел черно-синее платье, поэтому, услышав от пятого «бело-золотое», опешил и в то же время обрадовался. Как и весь остальной мир, лаборатория разделилась примерно пополам: одна половина видела платье черно-синим, другая — бело-золотым.

Час спустя я уже выступал на ВВС, пытаюсь объяснить, в чем тут дело. Все постепенно приходило к единодушному мнению, что эффект возникает из-за поправки на источник света, то есть влияния общего освещения на восприятие цвета. Идея заключалась в том, что у разных людей этот процесс может происходить по-разному, и обычно эта разница никак себя не проявляет, поэтому о ней не знали, и только теперь ей представился случай сыграть решающую роль.

Участники обсуждений очень быстро заметили, что фото платья пересвечено и ему недостает контекста (большую часть снимка занимает само платье), — это вполне могло одурачить мозг, пытающийся опереться в своем умозаключении на контекст. Если вашей зрительной системе по той или иной причине привычен желтоватый искусственный свет — допустим, вы много времени проводите в помещении, — высока вероятность, что она сделает вывод о цвете платья как о сине-черном, полагая источник освещения желтоватым. И наоборот, если ваши лучшие друзья — солнце, воздух и вода, а с ними и голубоватый естественный свет, вы, возможно, увидите платье бело-золотым.

Все сразу же принялись экспериментировать — вглядываться в снимок в полутемной комнате, а потом выбегать на дневной свет; проводить параллели между преобладанием видящих платье бело-золотым и средним количеством солнечных дней в разных странах; выяснять, насколько среди видящих сине-черное больше пожилых, чем молодых. В мгновение ока возникла целая кустарная отрасль, занимающаяся отработкой этих и других гипотез¹⁶.

Наличие такой разницы в чувственном опыте, вызываемом одним и тем же изображением, и такая уверенность в своем опыте у тех, кто его испытывает, убедительно доказывает, что наше восприятие окружающего мира — это внутренний конструкт, формируемый индивидуальными особенностями нашей личной биологии и биографии. Как правило, мы считаем, что все видят окружающий мир примерно одинаково, и в большинстве случаев, наверное, так и есть. Но даже если это и в самом деле так, то не потому, что красные кресла действительно красные, а потому, что выявить тончайшие различия в том, как наш мозг отбирает наиболее вероятные предположения, способно только совсем нестандартное стечение обстоятельств, которое и возникло в ситуации с платьем.

В качестве второго примера мы возьмем популярную оптическую иллюзию «Шахматная доска Адельсона». Она демонстрирует, что прогнозы воздействуют на восприятие не только в нетривиальных ситуациях, таких как случай с платьем, а сплошь и рядом, на каждом шагу. Посмотрите на шахматную доску в левой части рис. 5 и сравните клетки А и В. Скорее всего, А покажется вам темнее В. Такой ее вижу я, такой ее видят все, кому я это изображение показывал. Ни намек ни на какие индивидуальные различия.

На самом же деле А и В имеют абсолютно один и тот же оттенок серого. Это доказывает шахматная доска в правой части рисунка, где А и В соединены двумя полосами, имеющими одинаковый оттенок серого по всей длине. И вглядевшись в получившийся прямоугольник, вы увидите, что оттенки не меняются, никаких переходов цвета нет. А и В действительно одинаково серые, хотя на шахматной доске в левой части они всё так же будут различаться. Знание о том, что они одного цвета, ничем не поможет смотрящему. Я разглядывал эти картинки тысячи раз, и все равно А и В слева ни в какую не хотят выглядеть одинаково [15].

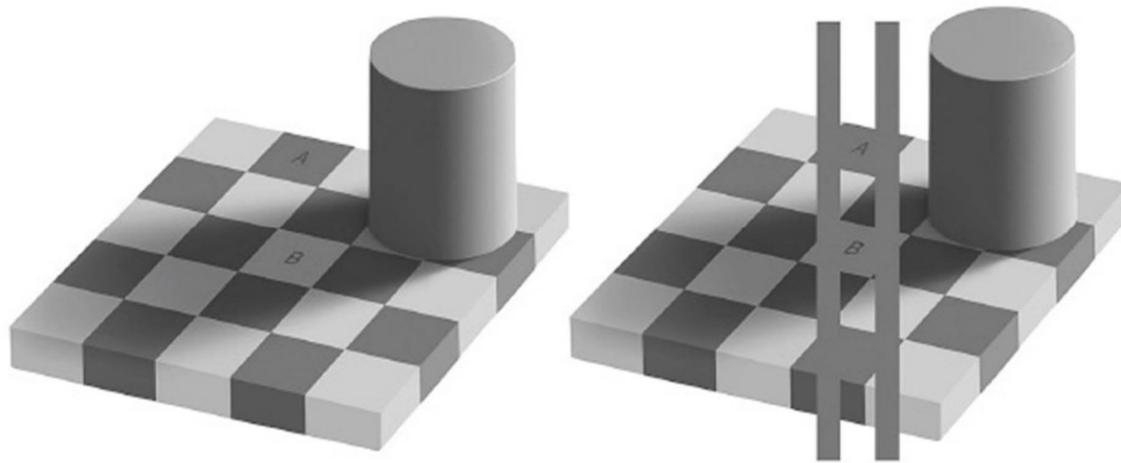


Рис. 5. Оптическая иллюзия Адельсона

Дело здесь в том, что восприятие серого определяется не теми световыми волнами, которые исходят от А и В (волны у них одни и те же), а наиболее вероятным предположением мозга о том, чем вызвано конкретное сочетание длин волны, и предположение это, как в случае с платьем, зависит от контекста. Клетка В находится в тени, клетка А — нет, а в нейронные сети зрительной системы мозга накрепко вшито, что в тени предметы кажутся темнее. И точно так же, как мозг корректирует свои перцептивные умозаключения с учетом общей освещенности, он корректирует умозаключения об оттенке клетки В с учетом своего предшествующего знания о тенях. Поэтому на шахматной доске в левой части клетка В кажется нам светлее, чем клетка А, на которую тень не падает. На шахматной доске в правой части контекст тени нарушается наложенными поверх клеток серыми полосами, и мы видим, что на самом деле А и В абсолютно одинаковы.

Происходит это совершенно автоматически. Вы не отдаете себе отчета (по крайней мере до сих пор не отдавали) в том, что, делая свои перцептивные прогнозы, мозг пользуется имеющимися у него заведомыми ожиданиями насчет теней. И это вовсе не недочет нашей зрительной системы. Результативная зрительная система — это не экспонометр, которым пользуются фотографы для измерения освещенности. Функция восприятия, по крайней мере в первом приближении, состоит в том, чтобы вычислять наиболее вероятные источники сенсорных сигналов, а не заставлять нас осознавать сами сенсорные сигналы, в чем бы это осознание ни заключалось.

Последний пример показывает, насколько быстро новые прогнозы влияют на сознательное восприятие. Взгляните на изображение ниже (рис. 6). Скорее всего, вы увидите только мешанину из черных и белых клякс. Потом, когда дочитаете предложение до конца, переверните несколько страниц и посмотрите на расположенную там фотографию (рис. 7).



Рис. 6. Что это?



Рис. 7. Это вот что. Исходный снимок для изображения на рис. 6^[16]

Посмотрели? Замечательно. Теперь снова взгляните на рис. 6 — на этот раз изображение покажется вам другим. Там, где прежде была неразбериха из пятен, теперь вырисовываются предметы, четкие очертания и какой-то сюжет. Перед нами черно-белая иллюзия, или «картинка в стиле Муни». Достаточно разглядеть сюжет каждой такой картинке один раз, и он будет виден там всегда. Для создания черно-белой иллюзии изображение переводится в оттенки серого, а затем осторожно сегментируется методом пороговой обработки, в результате которой все мелкие подробности пропадают, залитые радикальным черным или белоснежным. Если проделать все это надлежащим образом на подходящем изображении, будет очень трудно разобрать, что творится на картинке, пока вы не увидите оригинал, после чего черно-белая мешанина вдруг превратится в связный сюжет.

Примечателен этот пример тем, что сейчас, когда вы смотрите на черно-белую мешанину, ваши глаза получают всё те же сенсорные стимулы, что и в первый раз, сами стимулы за это время ничуть не изменились. Изменились только прогнозы мозга относительно источников этих сенсорных данных, и именно они меняют осознаваемую вами картину¹⁷.

Этот феномен характерен не только для зрения. Убедительным примером слухового восприятия такого рода служит «синусоидальная речь»¹⁸ — обработка произнесенной фразы, при которой отсекаются все высокие частоты, позволяющие

нам распознавать сказанное. Результат обычно похож на шумный свист, в котором нельзя разобрать ничего, то есть мы получаем слуховой эквивалент черно-белой иллюзии. Затем вы слушаете изначальную, необработанную, речь, а потом снова синусоидальную версию — и вдруг в ней все становится ясно. Точно так же, как с черно-белыми иллюзиями, четкий прогноз относительно источника сенсорных сигналов меняет — в данном случае обогащает — перцептивный опыт.

В совокупности эти примеры (признаю, намеренно простые) показывают, что восприятие — это творческий, генеративный акт, это проактивная, контекстозависимая интерпретация сенсорных сигналов и взаимодействие с ними. И как я уже говорил, принцип, согласно которому перцептивный опыт строится на рождающихся в мозге предсказаниях¹⁹, применяется повсеместно — не только к зрению и слуху, а ко всему нашему восприятию, постоянно.

Из этого следует, что мы никогда не воспринимаем мир таким, «какой он есть». Еще Кант, предлагая свою концепцию ноумена, заметил, что нам трудно представить, как выглядело бы такое восприятие. Мы уже убедились: даже столь простое явление, как цвет, существует только во взаимодействии мира и разума. И поэтому, как ни удивительно для нас, что перцептивные иллюзии — подобные уже рассмотренным нами — обнажают разницу между тем, что мы видим (слышим, осязаем), и тем, что есть на самом деле, мы должны постараться не оценивать перцептивный опыт исключительно с точки зрения «точности» его совпадения с реальностью. В таком понимании точное («достоверное») восприятие — это химера²⁰. Контролируемая галлюцинация мира, данного нам в ощущении, создавалась эволюцией, чтобы увеличить наши шансы на выживание, а не чтобы служить прозрачным окном во внешнюю действительность, не имеющим никакого понятийного смысла. В эти идеи мы углубимся в последующих главах, но прежде неплохо было бы разделаться с рядом контраргументов.

Первый из них заключается в том, что теория восприятия как контролируемой галлюцинации отрицает непреложные составляющие реального мира. «Если все, что мы ощущаем, это просто галлюцинация, — можете возмутиться вы, — идите прыгните под поезд и посмотрите, что получится».

Ничто из сказанного мной не нужно рассматривать как отрицание существования предметов и явлений окружающей действительности, будь то приближающийся поезд, кошка или чашка кофе. Слово «контролируемая» в нашем термине важно не менее, чем «галлюцинация». Изображая восприятие таким образом, мы не имеем в виду, что оно берется с потолка и может быть каким угодно, мы имеем в виду, что отображение окружающей действительности в перцептивном опыте — это конструкт нашего мозга.

В связи с этим полезно будет уяснить разницу между «первичными» и «вторичными» свойствами, как называл их философ эпохи Просвещения Джон Локк. К первичным качествам предмета он относил те, что существуют независимо от наблюдателя, в частности протяженность, плотность, подвижность²¹. Поезд на полном ходу этими первичными качествами обладает в избытке, поэтому лучше под него не прыгать, независимо от того, наблюдаете вы его или нет и каких взглядов на природу восприятия придерживаетесь. Вторичные качества — соответственно те, существование которых зависит от наблюдателя. Это свойства предметов, вызывающие ощущения — или «идеи» — в сознании, и существующими отдельно от объекта их назвать нельзя. Хорошим примером вторичного качества будет цвет, поскольку восприятие цвета зависит от взаимодействия определенного перцептивного аппарата с объектом.

С точки зрения контролируемой галлюцинации вызывать перцептивный опыт в ходе активного, конструктивного процесса могут как первичные, так и вторичные качества объектов. Однако ни в том ни в другом случае содержание этого перцептивного опыта не будет тождественно соответствующему качеству объекта.

Второй контраргумент касается нашей способности воспринимать новое. Может показаться, будто для всего, что нам когда-либо предстоит воспринять, необходимо сформированное заранее наиболее вероятное предположение и поэтому мы навеки заперты в перцептивном мире заведомых ожиданий. Представьте, что вы никогда не видели гориллу — ни в жизни, ни по телевизору, ни в кино, ни даже в книгах, — и вдруг она топает навстречу вам по улице. Я гарантирую, что гориллу вы в этом случае все равно увидите, и это будет новый и, вероятно, довольно пугающий перцептивный опыт. Но как он в принципе возможен в мире заведомых ожиданий?

Вкратце на этот вопрос можно ответить так: опыт «наблюдения гориллы» нельзя назвать целиком и полностью новым. Горилла — это животное с руками, ногами и шерстью, и вы — или ваши предки — непременно видели другие существа с такими же или частично такими же чертами. Если более общо, то горилла — это объект с отчетливыми (хотя и косматыми) контурами, он движется предсказуемым в разумных пределах образом, он отражает свет так же, как другие объекты схожего размера, цвета и текстуры. Новизна опыта «наблюдения гориллы» строится на перцептивных прогнозах, действующих на множестве разных уровней детализации и приобретаемых в течение разных периодов времени — от прогнозов относительно освещенности и очертаний до прогнозов относительно лиц и поз, — совокупность которых формирует новые общие наиболее вероятные перцептивные предположения, и поэтому гориллу вы видите впервые.

Чтобы ответить подробнее, нам нужно больше узнать о том, как мозг выполняет свою невероятно сложную нейронную акробатику, участвующую в перцептивных умозаклучениях, этим мы и займемся в следующей главе.

Глава 5

Вероятное неочевидное¹

Преподобный Томас Байес (1702–1761) — пресвитерианский священник, философ и статистик, проживший большую часть жизни на юге Англии, в Танбридж-Уэллсе, — так и не опубликовал теорему, которой суждено было обессмертить его имя. «Очерки к решению проблемы теории вероятностей» представил в Королевском научном обществе спустя два года после смерти Байеса его коллега — философ-проповедник Ричард Прайс, а значительную часть неподъемных вычислений выполнил позже французский математик Пьер-Симон Лаплас. Однако именно фамилия Байеса останется навеки связана с методом рассуждений, названным «вывод наилучшего объяснения», принципы которого играют главную роль в понимании того, как сознательное восприятие строится на наиболее вероятных предположениях, возникающих в мозге.

Байесовское рассуждение — это рассуждение на основе вероятностей. Если точнее, его суть в том, чтобы делать оптимальные выводы (умозаклучения, которые мы до сих пор называли наиболее вероятным предположением) в условиях неопределенности. «Умозаклучение» — это уже знакомый нам термин, означающий попросту выводы на основе свидетельств и логики. Байесовский вывод относится к категории абдуктивного рассуждения, которое нужно отличать от дедуктивного и индуктивного. При дедуктивном рассуждении мы руководствуемся только чистой логикой: если Джим старше Джейн, а Джейн старше Джо, значит, Джим старше Джо. Если посылки верны и логические правила применены правильно, дедуктивное

умозаключение гарантированно будет истинным. Индуктивное рассуждение — это экстраполяция из серии наблюдений: на протяжении всей обозримой истории солнце вставало на востоке, значит, оно всегда встает на востоке. Индуктивные умозаключения, в отличие от дедуктивных, могут быть ошибочными: первые три шарика, которые я вытащил из сумки, были зеленые, значит, все шарики в сумке зеленые. Этот вывод может оказаться как истинным, так и ложным.

Абдуктивное рассуждение — его разновидность, формализованная байесовским выводом, — состоит в том, чтобы найти лучшее объяснение для ряда наблюдений, когда эти наблюдения неполны, недостоверны и так или иначе сомнительны.

Абдуктивное рассуждение, как и индуктивное, тоже может привести к неверным выводам. Процесс поиска «лучшего объяснения» в абдукции можно представить как рассуждение не от причин к следствию, как в случае дедукции и индукции, а «наоборот», от наблюдаемых следствий к наиболее вероятным причинам.

Рассмотрим на примере. Выглянув поутру из окна своей спальни, вы видите, что газон мокрый². Получается, ночью шел дождь? Возможно, а может быть, вы просто забыли выключить разбрызгиватель. Задача — найти лучшее объяснение или гипотезу для наблюдаемого вами положения дел: если газон мокрый, какова вероятность, (1) что ночью шел дождь или (2) что вы не выключили полив? Иными словами, мы хотим вывести наиболее вероятную причину наблюдаемых данных.

Как это сделать, нам укажет байесовский метод. Он предлагает оптимальный способ обновления наших представлений по мере поступления свежих данных. Байесовское правило — это математический рецепт перехода от того, что мы уже знаем (априори), к тому, в чем мы должны увериться в дальнейшем (апостериори), на основании того, что мы узнаём сейчас (условная вероятность). Априорные, апостериорные и условные вероятности часто называют байесовскими «убеждениями», поскольку они представляют собой состояние знания, а не состояние мира. (Обратите внимание, что байесовское убеждение не означает, что убеждены должны быть именно мы лично. Сказать, что наша зрительная кора «убеждена», будто находящийся передо мной объект — это чашка кофе, мы можем с таким же успехом, как сообщить о своем убеждении, что Нил Армстронг высаживался на Луне.)

Априорная вероятность определяется состоянием дел до поступления новых данных. Допустим, вы живете в Лас-Вегасе, и тогда априорная вероятность ночного дождя очень невелика. Априорная вероятность, что вы забудете выключить полив, будет зависеть от того, как часто вы используете разбрызгиватель и насколько вы забывчивы. Она тоже невелика, но все же выше, чем априорная вероятность дождя.

Условная вероятность — это (если брать в широком смысле) противоположность апостериорной. Она формализует рассуждение «направленное вперед», от причин к следствию: если ночью будет дождь или останется включенным разбрызгиватель, какова вероятность того, что газон будет мокрым? Как и априорные вероятности, условные тоже могут варьироваться, но пока давайте исходить из того, что и дождь, и оставленный разбрызгиватель в равной степени способны увлажнить газон.

Байесовский метод представляет собой правило комбинирования априорных и условных вероятностей и выведения апостериорной вероятности для каждой гипотетической комбинации. Само правило простое³: апостериорная вероятность — это априорная вероятность, умноженная на условную и деленная на маргинальную вероятность (то есть «предшествующую данным» — в нашем случае это предшествующая вероятность мокрого газона и сейчас она нас не интересует, поскольку будет одинаковой для каждой гипотезы).

Увидев поутру за окном мокрый газон, Истинный Байесовец должен выбрать гипотезу с высочайшей апостериорной вероятностью, которая и будет наиболее вероятным объяснением имеющихся данных. Поскольку в нашем примере априорная

вероятность ночного дождя ниже, чем вероятность незапланированного орошения разбрызгивателем, апостериорная вероятность дождя тоже окажется ниже. Таким образом, Истинный Байесовец выберет гипотезу разбрызгивателя. Эта гипотеза и будет наиболее вероятным предположением Байесовца о причинах наблюдаемых данных, или «выводом наилучшего объяснения».

Если вам кажется, что это все и так понятно и ничего, кроме обычного житейского здравого смысла, для таких выводов не нужно, я соглашусь, но лишь потому, что наш пример большего и не требует. Однако есть множество других ситуаций, в которых байесовский вывод будет отличаться от того, что может предложить обычный здравый смысл. В частности, нам ничего не стоит ошибочно заподозрить у себя коварную болезнь на основании положительного медицинского анализа — из-за общей склонности завышать априорную вероятность развития редких болезней⁴. Даже если анализ обладает 99-процентной точностью, положительный результат может лишь немного повысить апостериорную вероятность наличия у вас соответствующего заболевания, если его распространенность среди населения довольно низка.

Давайте вернемся к сценарию с мокрым газоном и немного его разовьем. Рассмотрев как следует собственный газон, вы оглядываетесь на соседский — и видите, что он тоже мокрый. Это значимая новая информация. Теперь условная вероятность для каждой из гипотез будет разной: для подтверждения гипотезы разбрызгивателя мокрым должен быть только ваш газон, для подтверждения гипотезы дождя мокрыми должны быть оба газона. (Условная вероятность, напомню, направлена от предполагаемых причин к наблюдаемым данным.) Вы, как Истинный Байесовец, обновляете свои апостериорные вероятности, и теперь наилучшим объяснением для увиденного оказывается прошедший ночью дождь, поэтому вы меняете мнение.

Крупное преимущество байесовского вывода заключается в том, что при обновлении наиболее вероятных предположений он принимает в расчет надежность информации. Надежная (оцененная как надежная) информация имеет для байесовского убеждения больший вес, чем информация (оцененная как) ненадежная. Представьте, что окно в вашей спальне давно не мыто, а очки вы потеряли. Соседский газон вроде бы мокрый, но зрение у вас слабое, а окно грязное, поэтому новая информация очень ненадежна, и вы это знаете. В этом случае, хотя гипотеза дождя кажется при взгляде за соседский забор чуть более вероятной, в фаворитах может остаться изначальная гипотеза незапланированного полива.

Во многих ситуациях обновление наиболее вероятных байесовских предположений по мере получения свежих данных происходит снова и снова, и процесс вывода приобретает характер бесконечного цикла. В каждой итерации прежняя апостериорная вероятность становится текущей априорной, на ее основе трактуется новая порция данных и формируется новая апостериорная вероятность — новое наиболее вероятное предположение, затем цикл повторяется. Если ваш газон окажется мокрым и на следующее утро, то в этот раз вы в своем наиболее вероятном предположении о причинах будете отталкиваться от вчерашнего наиболее вероятного предположения, и так далее, с каждым новым днем.

Байесовский вывод с успехом используется в самых разных контекстах, от медицинской диагностики до поиска пропавших атомных подлодок, и ему постоянно находится все новое и новое применение⁵. Собственно, и сам научный метод можно рассматривать как байесовский процесс, в котором научные гипотезы обновляются с появлением новых свидетельств по итогам экспериментов. Подобное представление о науке отличается и от «сдвигов парадигмы» Томаса Куна, согласно которым целые научные построения переворачиваются, как только накапливается достаточный объем не совпадающих с ними свидетельств, и от «фальсификационизма» (опровергаемости) Карла Поппера, согласно которому гипотезы выдвигаются и

проверяются по очереди, одна за другой, словно воздушные шарик, которые отпускают в небо, а потом сбивают. В философии науки больше всего с байесовской концепцией схожи взгляды венгерского философа Имре Лакатоса, который в своих работах сосредоточивался прежде всего на том, что позволяет научно-исследовательским программам работать на практике, а не на том, в чем они в идеале должны состоять⁶.

Конечно, байесовское представление о науке подразумевает, что априорные убеждения ученых об обоснованности их теорий повлияют на то, в какой мере эти теории будут обновляться или опровергаться в свете новых данных. У меня, например, имеется сильное априорное убеждение, что мозг — это машина, выдающая байесовские прогнозы. Это сильное убеждение определяет не только мою трактовку экспериментальных данных, но и то, какие эксперименты я буду проводить, чтобы получить новые данные, отвечающие моим убеждениям. Мне иногда становится интересно, сколько понадобится данных, чтобы развенчать мое байесовское убеждение в том, что мозг, по сути, руководствуется байесовскими принципами⁷.

Давайте вернемся к нашему воображаемому мозгу, который, томясь в глухой темнице черепной коробки, пытается выяснить, что происходит снаружи, во внешнем мире. Теперь мы можем ухватиться за эту трудную задачу как за идеальную возможность вывести байесовское умозаключение. Выдвигая наиболее вероятные предположения об источниках своих зашумленных и неоднозначных сенсорных сигналов, мозг руководствуется принципами преподобного Томаса Байеса.

Перцептивные априорные вероятности могут кодироваться на разных уровнях абстракции и гибкости. Они варьируются от самых общих и относительно фиксированных — таких как «свет льется сверху» — до обусловленных конкретной ситуацией, таких как «приближающийся мохнатый объект — это горилла». Условные вероятности в мозге кодируют отображение потенциальных источников в сенсорных сигналах. Это составляющие перцептивного умозаключения по принципу «рассуждения, направленного вперед», которые, как и априорные вероятности, могут действовать в самых разных пространственно-временных масштабах. Мозг непрерывно комбинирует эти априорные и условные вероятности согласно байесовскому правилу, каждую долю секунды формируя новую байесовскую апостериорную вероятность — наиболее вероятное перцептивное предположение. И на каждом следующем витке постоянно меняющегося потока входящих сенсорных данных предшествующая апостериорная вероятность становится действующей априорной. Восприятие — это текучий процесс, а не застывший снимок.

Надежность сенсорной информации играет важную роль и здесь. Если вы не в зоопарке, то априорная вероятность, что замаячившая вдалеке темная косматая фигура окажется гориллой, очень низка. Поскольку этот «неопознанный объект» пока находится далеко, оцениваемая надежность входящих зрительных данных тоже будет низкой, поэтому вряд ли вашим наиболее вероятным перцептивным предположением сразу же будет горилла. Но по мере приближения зверюги зрительные сигналы будут становиться все надежнее и информативнее и мозг будет перебирать наиболее вероятные предположения одно за другим — большая черная собака, человек в костюме гориллы, собственно горилла, — пока вы воочию не увидите гориллу (будем надеяться, как раз вовремя, чтобы успеть убежать).

Проще всего представить байесовские убеждения — априорные, условные и апостериорные вероятности — как числа от 0 (обозначающего нулевую вероятность) до 1 (обозначающей стопроцентную вероятность). Но чтобы понять, как надежность сенсорных сигналов влияет на перцептивное умозаключение, и увидеть, как байесовское правило может на самом деле применяться в мозге, нам нужно еще

немного углубиться в теорию и начать рассуждать в терминах распределения вероятностей.

Диаграмма ниже демонстрирует пример распределения вероятностей для переменной x . Переменная в математике — это просто символ, который может принимать разные значения. Распределение вероятностей для x описывает вероятность того, что значения x будут располагаться в определенных пределах. Как видно на диаграмме, область значений можно изобразить в виде кривой. Вероятность того, что x находится в определенных пределах, задается площадью области под кривой, соответствующей этим пределам. В данном примере вероятность попадания x между двумя и четырьмя гораздо выше, чем вероятность его попадания между четырьмя и шестью. На графике распределения вероятностей общая площадь, лежащая под кривой, всегда равна единице. Это обусловлено тем, что при рассмотрении всех вероятных исходов какой-то непременно реализуется⁸.

Распределение вероятностей может принимать самую разную форму. Распространенная разновидность таких форм, к которой относится и данная кривая, — нормальное, или гауссовское, распределение. Оно полностью описывается средним значением или медианой (значением на вершине кривой — в нашем случае это три) и прецизионностью (насколько далеко друг от друга расположены значения — чем выше прецизионность, тем меньше растянута кривая). Эти величины — среднее и прецизионность — называются параметрами распределения^[17].

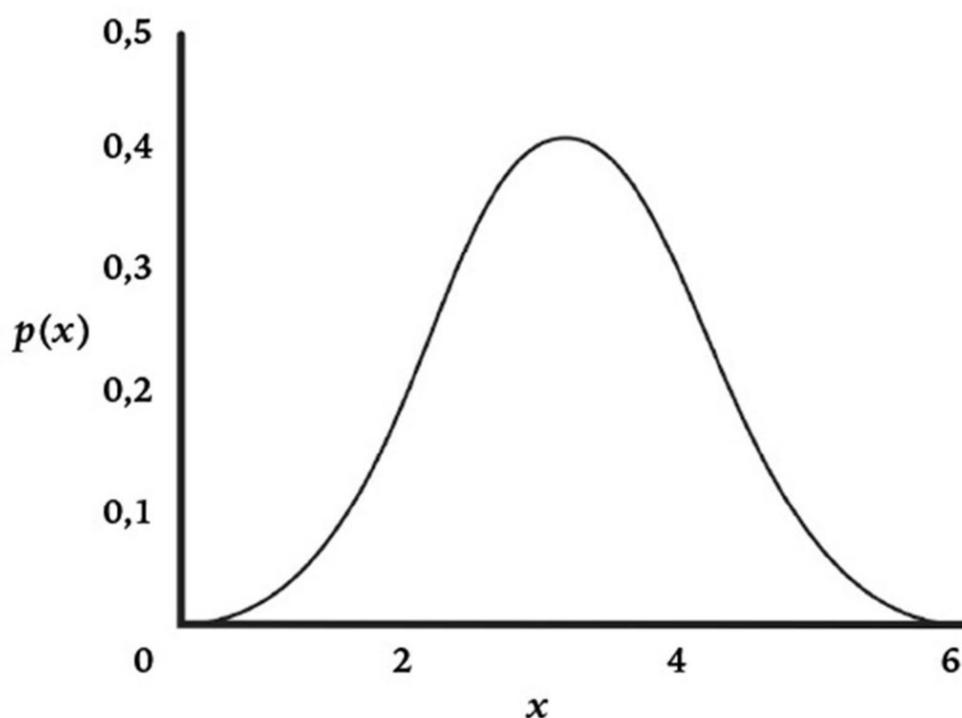


Рис. 8. Гауссовское распределение вероятностей

Нам это важно, поскольку байесовские убеждения можно с успехом представить в виде распределения Гаусса. В интуитивном смысле среднее означает содержание убеждения, а прецизионность — уверенность мозга по поводу этого убеждения. Сжатая заостренная кривая (высокая прецизионность) соответствует высокой уверенности в убеждении. Как мы еще увидим, именно в этой способности отражать уверенность, или надежность, заключена сила и действенность байесовского вывода. Вернемся к примеру с гориллой. Теперь мы можем представить соответствующие априорные, условные и апостериорные вероятности как распределения вероятностей, каждое из которых будет характеризоваться средним значением и прецизионностью. Для каждого распределения среднее будет означать вероятность того, что объект окажется гориллой, а прецизионность — уверенность мозга в оценке этой вероятности⁹.

Что происходит при появлении новых сенсорных данных? Процесс байесовского обновления данных лучше и проще будет изобразить графически. На следующей диаграмме пунктирная кривая представляет априорную вероятность встречи с гориллой. У этой кривой низкое среднее значение, указывающее на то, что гориллы

здесь считаются редкостью, и относительно высокая прецизионность, говорящая о том, что уверенность в этом априорном убеждении высока. Прерывистая кривая — это условная вероятность, соответствующая входящим сенсорным данным. Здесь среднее значение выше, но ниже прецизионность: если бы навстречу вам действительно двигалась горилла, именно такие сенсорные данные вы бы и получали, но вы в этом не особенно уверены. И наконец, сплошная кривая — это апостериорная вероятность, представляющая вероятность того, что перед вами горилла, на основании имеющихся сенсорных данных. Апостериорная вероятность, как всегда, выводится с применением байесовского правила. При использовании распределения Гаусса применение байесовского правила сводится к перемножению пунктирной и прерывистой кривой, при этом площадь области под конечной кривой — апостериорной — должна сохраняться равной единице.



Рис. 9. Байесовский вывод с распределением Гаусса для наиболее вероятного предположения насчет встречи с гориллой

Обратите внимание, что пик кривой апостериорной вероятности ближе к пику априорной, чем пик условной. Это потому, что комбинация двух распределений Гаусса зависит и от среднего значения, и от прецизионности. В данном случае, поскольку у условной вероятности прецизионность относительно низкая — сенсорные сигналы, указывающие на наличие гориллы, оцениваются как ненадежные, — наиболее вероятное апостериорное предположение недалеко уходит от априорного. Но уже при следующем взгляде сенсорные данные, касающиеся гориллы, могут оказаться более отчетливыми, поскольку она приближается, и так как новая априорная вероятность — это прежняя апостериорная, то новая апостериорная — новое наиболее вероятное предположение — сдвинется ближе к версии «горилла». И так далее, пока не станет ясно, что пора уносить ноги.

Теорема Байеса обеспечивает критерий оптимальности для перцептивных умозаключений. Она задает наиболее оптимистичные сценарии того, что должен делать мозг, вычисляя наиболее вероятные источники входящих сенсорных данных, будь то горилла, красное кресло или чашка кофе. Однако это далеко не весь процесс. Теорема Байеса не в состоянии показать, как именно с точки зрения нейронных механизмов мозг проворачивает этот фокус с выдачей наиболее вероятного предположения.

Чтобы узнать, как это происходит, нам придется вернуться к теории контролируемой галлюцинации и ее центральному постулату, что содержание сознания не просто формируется перцептивными прогнозами — оно и есть эти прогнозы.

В предшествующей главе мы ознакомились с концепцией восприятия как непрерывного процесса минимизации ошибок предсказания. Согласно данной концепции мозг постоянно создает прогнозы относительно сенсорных сигналов и сравнивает эти прогнозы с сенсорными сигналами, которые улавливают глаза и уши, а также нос, кожа и так далее. Из-за разницы между прогнозируемым и реальным возникают ошибки предсказания. Если поток перцептивных предсказаний устремлен преимущественно сверху вниз (изнутри наружу), то поток ошибок предсказания направлен снизу вверх (снаружи внутрь). На основании сигналов об ошибке предсказания мозг обновляет свои прогнозы и готовится к новому раунду входящей сенсорной информации. То, что мы воспринимаем, складывается из содержания всех нисходящих прогнозов, вместе взятых, после максимальной минимизации (или «разъяснения») сенсорных ошибок предсказания¹⁰.

Концепция контролируемой галлюцинации во многом схожа с другими «предиктивными» теориями восприятия и функционирования мозга, прежде всего с теорией «предиктивной обработки»¹¹. Однако между ними есть существенная разница — в том, на что они делают основной упор. Теория предиктивной обработки рассматривает механизмы, посредством которых мозг осуществляет восприятие (а также познание и действие). Теория контролируемой галлюцинации, в отличие от нее, рассматривает, как устройством механизмов мозга объясняются феноменологические свойства сознательного восприятия. Иными словами, теория предиктивной обработки рассказывает, как работает мозг, а концепция контролируемой галлюцинации идет дальше и обосновывает природу сознательного опыта. Важно, что фундаментом для обеих служит процесс минимизации ошибок предсказания.

И именно минимизация ошибок предсказания увязывает между собой концепцию контролируемой галлюцинации и байесовский вывод. Она берет байесовское утверждение о том, что должен делать мозг, и превращает в предположение о том, что он делает на самом деле. И оказывается, что, постоянно и повсеместно минимизируя ошибки предсказания, на самом деле мозг применяет байесовское правило. Точнее, он аппроксимирует байесовское правило. Этой связью и санкционирована идея, что перцептивное содержание представляет собой нисходящую контролируемую галлюцинацию, а не восходящее «считывание» сенсорных данных.

Давайте рассмотрим три основные составляющие минимизации ошибок предсказания в мозге — порождающие модели, перцептивные иерархии и «взвешивание прецизионности» сенсорных сигналов.

Порождающие модели определяют репертуар воспринимаемого. Чтобы воспринять гориллу, мой мозг должен вооружиться порождающей моделью, способной генерировать соответствующие сенсорные сигналы, то есть такие, которых следует ожидать, если я действительно имею дело с гориллой. Эти модели обеспечивают поток перцептивных прогнозов, которые при сравнении с поступающими сенсорными данными образуют ошибки предсказания, а они в свою очередь по мере того, как мозг пытается минимизировать эти ошибки, индуцируют обновление прогнозов¹².

Перцептивные прогнозы разыгрываются в самых разных пространственно-временных масштабах, поэтому мы воспринимаем структурированный мир, полный предметов, людей и пространств. Из высокоуровневого прогноза относительно наблюдения гориллы возникают прогнозы более низкого уровня относительно ее конечностей, глаз, ушей, шерсти, которые затем обрушиваются каскадом еще ниже, выливаясь в прогнозы относительно цвета, текстуры, очертаний и, наконец, в предугадываемые вариации яркости по всему полю зрения¹³. Перцептивная иерархия может затрагивать несколько сенсорных областей и даже распространяться за

пределы сенсорных данных. Если я вдруг услышу голос матери, зрительная кора включит предсказания, говорящие о том, что приближающаяся фигура — это моя мать. Если я знаю, что я в зоопарке, перцептивные области моего мозга будут больше подготовлены к встрече с гориллой, чем во время обычной прогулки по улице.

Здесь важно уточнить, что «прогнозы» при минимизации ошибок предсказания не обязательно касаются будущего. Прогноз в данном случае — это просто выход за рамки данных благодаря использованию модели. В статистике суть прогноза в том, чтобы отработать недостаток данных. Чем этот недостаток обусловлен — тем, что предсказания делаются относительно будущего (будущее вполне можно рассматривать как недостаток данных) или относительно текущего, но не полностью известного положения дел, — значения не имеет.

Последняя ключевая составляющая минимизации ошибок предсказания — это взвешивание прецизионности¹⁴. Мы уже видели, как относительная надежность сенсорных сигналов определяет степень обновления перцептивных умозаключений. Первый взгляд на маячащую вдалеке гориллу или на соседский газон через грязное окно принесет сенсорные сигналы низкой надежности, поэтому ваше наиболее вероятное байесовское предположение пока не слишком продвинется. Еще мы видели, как отражается надежность в прецизионности соответствующего распределения вероятностей. Как показывает график на рис. 9, сенсорные данные с низкой оцениваемой прецизионностью меньше влияют на обновление априорных убеждений.

Я говорю «оцениваемая прецизионность», а не просто «прецизионность», потому что прецизионность сенсорных сигналов не подается воспринимающему мозгу на блюдечке. Ее тоже нужно выводить как умозаключение. Мозгу приходится не только вычислять наиболее вероятные источники входящей сенсорной информации, но и устанавливать, насколько эта входящая сенсорная информация надежна. На практике это значит, что мозг постоянно корректирует степень влияния сенсорных сигналов на перцептивные умозаключения. Делает он это, временно меняя их оцениваемую прецизионность. Именно это и означает термин «взвешивание прецизионности». Снижение ее веса подразумевает, что сенсорные сигналы начинают меньше влиять на обновление наиболее вероятных предположений; увеличение веса, наоборот, увеличивает влияние сенсорных сигналов на перцептивные умозаключения. Таким образом, взвешивание прецизионности играет немаловажную роль в дирижировании этой изящнейшей переключкой прогнозов и ошибок предсказания, без которого не получится вывести наиболее вероятное перцептивное предположение.

Выглядит это все очень сложно, однако на самом деле роль взвешивания прецизионности в восприятии нам всем отлично знакома. Повышение оцениваемой прецизионности сенсорных сигналов — это не что иное, как «обращение внимания». Когда мы обращаем на что-то внимание — например, присматриваемся, пытаемся разглядеть, действительно ли там впереди горилла, — мозг усиливает взвешивание прецизионности для соответствующих сенсорных сигналов, что эквивалентно повышению их оцениваемой надежности или включению их «усиления». Теперь понятно, почему иногда мы вроде бы смотрим, но не видим. Если вы обращаете внимание на одни сенсорные данные — увеличиваете их оцениваемую прецизионность, — другие сенсорные данные будут меньше влиять на обновление наиболее вероятных перцептивных предположений.

Примечательно, что в некоторых ситуациях обойденные вниманием сенсорные данные полностью лишаются возможности повлиять на воспринимаемое. В 1999 г. психолог Дэниел Саймонс снял хорошо сейчас известную видеодемонстрацию¹⁵ этого феномена, который он назвал «слепота невнимания». Если вы не видели этот ролик, рекомендую посмотреть, прежде чем читать дальше^[18].

Вот что там происходит. Участники проводимого Саймонсом эксперимента смотрят короткий ролик, в котором сняты две команды по три игрока в каждой. Одна команда одета в черное, другая в белое. У каждой из команд имеется баскетбольный мяч, который они передают друг другу, хаотично перемещаясь и меняясь местами. Задача зрителя — подсчитать, сколько передач сделают «белые». Для этого ему приходится напрячь внимание, поскольку шесть игроков беспорядочно двигаются по всему пространству кадра, а мячей передается два.

Поразительно следующее: подсчитывая передачи, большинство людей совершенно не замечают человека в черном костюме гориллы, который проходит через кадр слева направо, по пути делая несколько характерных гориллиных движений. Когда им включают тот же ролик снова, на этот раз попросив найти гориллу, они видят ее сразу и часто отказываются верить, что это то же самое видео. Дело в том, что при сосредоточении внимания на игроках в белом сенсорные сигналы от игроков в черном и от гориллы получают низкую оцениваемую прецизионность, поэтому почти (или совсем) не влияют на обновление наиболее вероятных перцептивных предположений.

Что-то похожее произошло много лет назад и со мной, когда я ехал к моему любимому месту для сёрфинга в Сан-Диего. Я повернул налево под недавно установленный знак «поворот налево запрещен» — на короткий съезд к океану где-то у Дель-Мара. Поскольку никаких очевидных причин установки такого знака не наблюдалось, поскольку другие машины передо мной тоже туда поворачивали, поскольку за долгие годы я сворачивал на этот съезд, наверное, сотни раз и поскольку я страшно возмущаюсь, когда меня пытаются несправедливо оштрафовать, я доказывал в своих письменных показаниях, что знак был для меня в буквальном смысле невидим, даже если он «в принципе» находился в поле зрения. Я строил свою защиту на принципе слепоты невнимания. Да, новый знак имелся, но в силу взвешивания прецизионности в мозге я его воспринять не смог. Я дошел до Калифорнийского суда по делам о нарушении безопасности движения — пусть не до Верховного, но все же достаточно высокого, чтобы моя фамилия появилась в календаре слушаний. Я даже подготовил симпатичную презентацию в Power Point для судьи, но и это, увы, не помогло.

Слепотой невнимания (возможно, не подозревая, что она называется именно так) пользуются и фокусники. Салонные фокусы, в частности, требуют виртуозно отвлекать внимание зрителей, чтобы они не заметили, как иллюзионист закладывает за ухо даму червей, которая чуть погодя должна возникнуть там словно из ниоткуда¹⁶. Пользуются этой особенностью перцептивной физиологии и преуспевающие карманники. Я наблюдал однажды, как щипач-профи Аполло Роббинс без малейших усилий уводил у моих коллег часы, бумажники и сумочки, — трюк тем более примечательный, что многие из «потерпевших» великолепно разбираются в восприятии, знают все о слепоте невнимания и были предупреждены о намерениях Роббинса.

Нам, конечно, удобнее видеть свое взаимодействие с миром примерно так: сперва мы воспринимаем мир таким, какой он есть, затем решаем, что делать, — и делаем. Ощущаем, думаем, действуем. Возможно, так это и выглядит, но повторю, истинное положение вещей не всегда позволяет судить о себе по тому, каким оно представляется. Пора вводить в нашу картину действие.

Действие неотделимо от восприятия. В силу своей тесной связи восприятие и действие определяют и обуславливают друг друга. Каждое действие модифицирует восприятие, меняя входящие сенсорные данные, а любое восприятие оказывается таким, какое оно есть, чтобы направлять действие. Без действия восприятие просто

лишается смысла. Мы воспринимаем окружающий мир, чтобы эффективно в нем действовать, добиваться своих целей и в конечном итоге повышать свои шансы на выживание¹⁷. Мы воспринимаем мир не таким, какой он есть, а таким, каким он нам полезен.

Не исключено даже, что действие здесь первостепенно. Можно попробовать исходить не из того, что мозг выводит наиболее вероятные перцептивные умозаключения, чтобы затем направлять поведение, а из того, что его основополагающее занятие — порождение действий¹⁸ и постоянная их калибровка с учетом сенсорных сигналов, чтобы наилучшим образом достигать стоящих перед организмом целей. С этой точки зрения мозг предстает по самой своей сути динамичной, активной системой, непрерывно прощупывающей среду и изучающей последствия^[19].

В предиктивной обработке действие и восприятие — это две стороны одной медали. Обе опираются на минимизацию сенсорных ошибок предсказания. До сих пор я описывал этот процесс минимизации в категориях обновления перцептивных прогнозов, но это не единственный вариант. Сокращать ошибки предсказания можно и совершая действия, ведущие к изменению сенсорных данных, чтобы новые сенсорные данные соответствовали существующему прогнозу. Минимизация ошибок посредством действия называется активным выводом — этот термин предложил британский нейробиолог Карл Фристон¹⁹.

Проще всего представить себе активный вывод как некое самоисполняющееся перцептивное пророчество — процесс, при котором мозг посредством совершения действий ищет те сенсорные данные, которые позволят сбыться его перцептивным прогнозам. Действия эти могут быть самые простые — например, перевести взгляд. Сегодня утром я искал в вечной свалке на своем рабочем столе ключи от машины. И пока я шарил глазами по этому беспорядочному нагромождению, у меня не только ежесекундно обновлялись зрительные прогнозы (пустая кружка, пустая кружка, скрепки, пустая кружка...), мой сосредоточенный взгляд непрерывно исследовал открывающуюся мне картину, пока перцептивный прогноз относительно ключей не исполнился.

К тому или иному изменению данных приведет любое физическое действие — будь то перевод взгляда, переход в другую комнату, напряжение мышц живота. Даже такие крупномасштабные, «высокоуровневые» действия, как поиски работы или решение сочетаться браком, распадутся каскадом на последовательности физических действий, которые изменят входящую сенсорную информацию. Любое действие потенциально способно подавить сенсорные ошибки прогнозирования за счет активного вывода, поэтому любое действие непосредственно участвует в восприятии.

Активный вывод, как и все составляющие прогнозной обработки, зависит от порождающего моделирования. Точнее, активный вывод опирается на способность порождающих моделей прогнозировать сенсорные последствия действий. Это прогнозы вида «если я посмотрю туда, какие сенсорные данные я, скорее всего, получу?». Такие прогнозы называются условными, поскольку состоят в том, что случится при том или ином условии. Без условных прогнозов такого рода мозгу неоткуда было бы знать, какое действие из бесконечного множества возможных с наибольшей вероятностью снизит сенсорные ошибки прогнозирования. Действия, которые мой мозг спрогнозировал как наиболее вероятные для нахождения затерявшихся на столе ключей от машины, состояли в том, чтобы обшаривать взглядом поверхность стола, а не смотреть в окно или размахивать руками над головой.

Активный вывод способен не только исполнять уже имеющиеся перцептивные прогнозы, но и помогать улучшать их. На коротких временных отрезках действие

может собирать новые сенсорные данные, позволяя улучшить наиболее вероятное предположение или выбрать из нескольких соперничающих перцептивных гипотез одну. Мы видели такой пример в самом начале главы, когда взгляд на соседский газон мог поставить точку в соперничестве гипотез ночного дождя и оставленного разбрызгивателя. Другим примером было бы навести на столе порядок и убрать все кружки, чтобы проще стало найти ключи. В каждом случае выбор соответствующего действия опирается на наличие порождающей модели, способной спрогнозировать, как в результате изменятся сенсорные данные.

В конечном итоге на действиях основывается обучение, которое в данном случае означает улучшение порождающих моделей мозга за счет получения новой информации о причинах сенсорных сигналов и о каузальной структуре мира в целом. Заглядывая за соседский забор, чтобы вычислить причины намокания одного газона, я заодно узнаю что-то новое о причинах намокания газонов вообще. В наилучшем случае активный вывод может породить «добродетельный круг» положительной обратной связи: правильно выбранные действия будут выявлять полезную информацию об устройстве мира, которая будет затем встраиваться в усовершенствованные порождающие модели, они будут обеспечивать совершенствование перцептивных умозаключений и направлять новые действия, согласно прогнозу призванные принести еще больше полезной информации²⁰.

Самое контринтуитивное в активном выводе — это возможность рассматривать само действие как разновидность самоисполняющегося перцептивного пророчества. С этой точки зрения действие не просто участвует в восприятии — оно и есть восприятие. Когда я шарю взглядом в поисках ключей от машины или двигаю руками, убирая кружки, перцептивные прогнозы о положении и движениях моего тела способствуют своему исполнению.

В активном выводе действия выступают самоисполняющимися проприоцептивными пророчествами. Проприоцепция — это вид восприятия, состоящий в отслеживании положения и движений тела за счет регистрации сенсорных сигналов, поступающих от рецепторов, рассредоточенных по всем скелетным костям и мускулатуре. Мы не особенно задумываемся о ней, потому что в каком-то смысле она всегда с нами, но уже одно то, что мы можем с закрытыми глазами коснуться собственного носа — попробуйте! — показывает, насколько существенную роль она играет во всех наших действиях. С точки зрения активного вывода дотронуться до носа значит позволить группе проприоцептивных прогнозов о движениях и положении руки самоисполниться, то есть восторжествовать над сенсорными данными о том, что пальцы в данный момент НЕ касаются носа. И здесь снова вступает в игру взвешивание прецизионности. Чтобы проприоцептивные прогнозы обеспечивали свое исполнение, необходимо уменьшить вес ошибок предсказания, указывающих мозгу, где и как в данный момент располагается тело. Это можно считать противоположностью внимания — чем-то вроде «безразличия» к телу, благодаря которому оно, собственно, и двигается²¹.

Представляя себе действие подобным образом, мы отчетливее видим, что действие и восприятие — это две стороны одной медали. Мы рассматриваем их не как ввод (восприятие) и вывод (действие) по отношению к некоему «центральному» разуму, а как разновидности совершающегося в мозге предсказания. В основе обеих находится общий процесс байесовского вывода наиболее вероятных предположений, тщательно срежиссированный танец перцептивных прогнозов и сенсорных ошибок предсказания, разница лишь в том, кто в этом танце ведет, а кто ведомый.

Давайте в последний раз наведемся к нашему воображаемому мозгу, томящемуся в своей костяной темнице. Теперь мы знаем, что он отнюдь не отрезан от внешнего

мира. Он купается в потоках сенсорных сигналов от окружающей среды и организма и постоянно дирижирует действиями — самоисполняющимися проприоцептивными пророчествами, — которые проактивно формируют этот сенсорный поток. Входящий сенсорный шквал сталкивается с каскадом нисходящих прогнозов — сигналы ошибок предсказания устремляются вверх, стимулируя все более совершенные прогнозы и вызывая новые действия. Этот неуклонный процесс порождает аппроксимацию байесовского вывода, когда мозг раз за разом выбирает сменяющие друг друга наиболее вероятные предположения о причинах и источниках окружающей его сенсорной среды, в результате рождая на свет живой и яркий перцептивный мир — контролируруемую галлюцинацию.

Теперь, понимая контролируруемую галлюцинацию именно так, мы имеем все основания считать, что нисходящие прогнозы не просто определяют наше восприятие. Они его составляют. Наш перцептивный мир, полный цветов, форм и звуков, — это не что иное, как наиболее вероятные предположения нашего мозга о скрытых причинах и источниках его бесцветных, бесформенных и беззвучных входящих сенсорных данных.

И как мы увидим далее, таким образом объясняется не только восприятие нами кошек, кофейных чашек и горилл, а совершенно любая составляющая нашего перцептивного опыта.

Глава 6

Вклад зрителя

Наш путь к глубинным структурам перцептивного опыта начнется с путешествия в Вену начала XX в. Будь вы в те годы завсегдатаем элегантных венских кофеен, салонов и опиумных притонов, вам наверняка довелось бы столкнуться с некоторыми весьма занятыми личностями, а именно представителями Венского философского кружка, в который входили Курт Гедель, Рудольф Карнап и время от времени Людвиг Витгенштейн. Кроме них к этому кружку принадлежали основоположники модернизма в живописи Густав Климт, Оскар Кокошка и Эгон Шиле, а также историк искусства Алоиз Ригль. И разумеется, Зигмунд Фрейд.

В кипучей интеллектуальной атмосфере тогдашней Вены наука и искусство смешивались легко и непринужденно. Науку не ставили превыше искусства в том привычном теперь смысле, согласно которому искусство и пробуждаемый им у человека отклик ничего не стоят без научного объяснения. Вместе с тем и искусство не пыталось удалиться в недостижимые для науки пределы. Художники и ученые и их критики были союзниками, объединяющимися в попытке понять человеческий опыт во всем его богатстве и разнообразии. Недаром нейробиолог Эрик Кандель назвал эту эпоху «веком самопознания»¹.

Одна из самых влиятельных идей, рожденных в ту пору, — предложенная Риглем и позже популяризованная другим венцем, одним из крупнейших деятелей истории искусства XX в. Эрнстом Гомбрихом, концепция «вклада зрителя»². Эта идея подчеркивает роль наблюдателя — в данном случае зрителя, который творчески «завершает» произведение искусства. Вклад зрителя — это та часть перцептивного опыта, которую вносит адресат и которая в самом произведении (или окружающем мире) не содержится³.

Концепция вклада зрителя практически напрашивается на то, чтобы увязать ее с прогнозными теориями восприятия, такими как теория контролируемой галлюцинации. Кандель писал: «Открытие участия восходящих процессов в восприятии зрителя убедило Гомбриха в том, что "непредвзятый взгляд" в принципе невозможен, то есть зрительное восприятие основано на интерпретации зрительной

информации. Гомбрих утверждал, что мы не воспринимаем то, что не в состоянии классифицировать»4.

Мне кажется, особенно заметен вклад зрителя при взаимодействии с такими художниками, как Клод Моне, Поль Сезанн и Камиль Писсарро. Когда я стою перед каким-нибудь из шедевров импрессионизма, например «Инеем»5 Писсарро (создан в 1873 г. и сейчас представлен в парижском музее Орсе), меня словно затягивает в совершенно другой мир. Почему подобные картины обретают над нами такую власть? В том числе и потому, что они оставляют нашей зрительной системе простор для совершения интерпретационной работы. «Остатки краски, счищенные с палитры и наляпанные на грязный холст», как отозвался о полотне Писсарро критик Луи Леруа6, прекрасно работают на создание перцептивного впечатления заиндевелшего поля.

Импрессионистские пейзажи стремятся изъять художника из акта живописи и воскресить гомбриховский «непредвзятый взгляд»7, сообщая произведению те вариации яркости, которые послужат для перцептивного умозаключения исходным материалом, а не конечным продуктом. Для этого художнику необходимо выработать тонкое и глубокое понимание того, как проявляются субъективные, феноменологические составляющие зрительного восприятия, и уметь это понимание применять. Каждую работу художника можно считать упражнением в «обратной разработке» человеческой зрительной системы, начиная с входящих сенсорных данных и заканчивая связным субъективным опытом. Картины становятся экспериментами в области прогнозного восприятия и природы сознательного опыта, порождаемого этими процессами8.

Такие произведения, как картина Писсарро, — это не просто эхо или предтеча науки о восприятии, а вклад зрителя — это не просто историко-живописная разновидность минимизации ошибок прогнозирования. Гомбрих со товарищи преподнес нам глубокое понимание феноменологической, эмпирической природы восприятия — понимание, которое очень легко упустить, закопавшись в винтиках и шестеренках априорных вероятностей и ошибок прогнозирования.

«Говоря, что пятна и мазки на импрессионистском полотне "вдруг оживают", мы имеем в виду, что нас побуждают самим спроецировать на эти разноцветные точки пейзаж»9. В этом объяснении Гомбрих ухватывает кое-что очень важное для сознательного восприятия, кое-что выходящее за рамки искусства как такового и относящееся к сути чувственного опыта в принципе10. Наше восприятие окружающего мира как «существующего на самом деле» — это не пассивное выявление объективной действительности, а живая, текущая проекция — установление мозгом контакта с миром.

В научной лаборатории попытки разобраться в том, как субъективный опыт опирается на перцептивные ожидания, начинаются с простых экспериментов. Одна из экспериментальных гипотез состоит в том, что ожидаемое мы должны воспринимать быстрее и легче, чем неожиданное11. Эту гипотезу несколько лет назад взялся проверить мой тогдашний постдокторант, а сейчас старший преподаватель Амстердамского университета Яир Пинто, по нашему обыкновению сосредоточив свое исследование на зрительном восприятии.

Яир использовал метод под названием «непрерывное подавление стимула вспышкой»12, при котором левому и правому глазу предъявляются разные изображения. Одному глазу предъявляется четкая картинка (в данном случае либо дом, либо лицо), а другому — быстро меняющийся рисунок из частично перекрывающихся друг друга прямоугольников. Когда мозг пытается объединить эти два образа в одной сцене, у него ничего не получается — человек видит только

меняющиеся прямоугольники. Осознанное восприятие картинки подавляется «непрерывно вспыхивающими» фигурами. В нашем эксперименте, отраженном на рис. 10, использовалась вариация этого метода, в которой сначала контрастность прямоугольников задавалась высокой и затем постепенно снижалась, а контрастность картинки сначала задавалась низкой и постепенно повышалась¹³. Это значит, что максимум через несколько секунд картинка — лицо или дом — начинала восприниматься осознанно.

Чтобы выяснить, как перцептивные ожидания влияют на сознательное восприятие, перед каждой экспериментальной попыткой мы давали участнику подсказку — либо слово «дом», либо слово «лицо». Важно отметить, что эти ожидания не обязательно соответствовали истине. Когда участникам выдавалось слово «лицо», лицо присутствовало на изображении лишь в 70% попыток, а в остальных 30% на картинке изображался дом. То же самое происходило и с подсказкой «дом». Замеряя, через какое время изображение проявится, преодолев подавление вспышкой, мы определяли, насколько быстро распознается картинка — лицо или дом — в том случае, когда испытуемый ее ожидает, и в том, когда не ожидает.

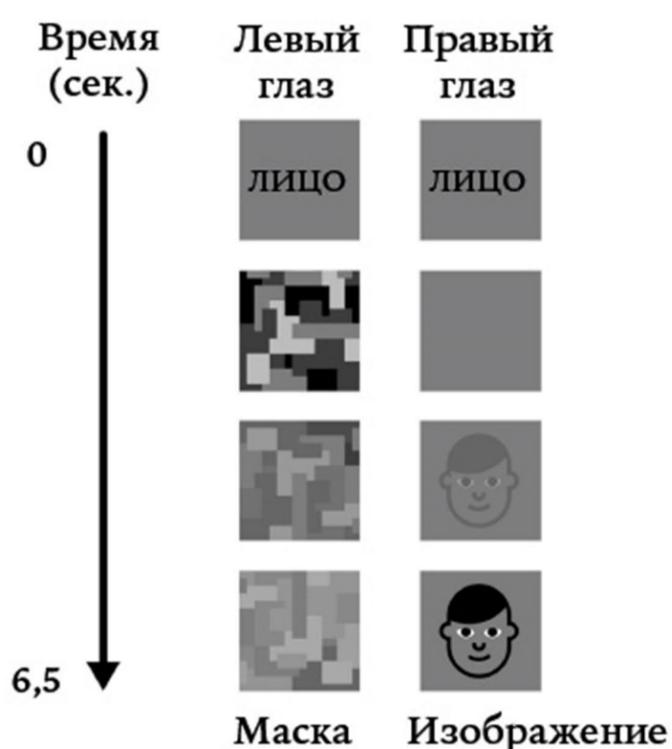


Рис. 10. Непрерывное подавление стимула вспышкой с изменением контрастности^[20]

Как мы и предполагали, быстрее и точнее испытуемые распознавали дом, когда именно его они и ожидали увидеть. То же самое с лицами. Разница в скорости распознавания была невелика — около одной десятой секунды, — но о ней можно было говорить с уверенностью. В нашем эксперименте справедливые перцептивные ожидания действительно вели к более быстрому и точному осознанному восприятию.

Наше исследование лишь одно из растущего числа призванных выяснить, что происходит, когда в игру вступают перцептивные ожидания. Авторы другого изящного эксперимента — Мика Хейлброн, Флорис де Ланге и их коллеги из Института Дондерса в Неймегене — воспользовались так называемым эффектом превосходства слова. Отдельные буквы, в частности U, проще идентифицировать в составе слова, скажем HOUSE (дом), а не когда они входят в ничего не значащий набор букв, например AEUVR. Группа де Ланге показывала испытуемым множество примеров слов и бессмысленных наборов букв, всегда на фоне визуального шума. В полном соответствии с эффектом превосходства слова отдельные буквы оказалось проще распознавать в слове, чем в бессвязном наборе.

Отличительной особенностью этого исследования стал хитроумный способ анализа активности мозга испытуемых, зафиксированной с помощью фМРТ. При анализе данных с использованием мощной технологии «чтение мозга»¹⁴ было обнаружено, что специфическая нейронная активность, связанная с буквой («сигнатура»), в

зрительной коре выделялась «резче», то есть ее легче было отличить от нейронных репрезентаций других букв, когда она находилась в слове, а не в бессмысленном наборе. Это значит, что перцептивные ожидания, создаваемые контекстом слова, меняли активность мозга на ранних стадиях зрительной обработки в сторону усиления восприятия — в полном соответствии с тем, что предполагает концепция контролируемой галлюцинации.

Увы, какими бы показательными ни были эти эксперименты, лабораторным условиям по-прежнему далеко до того богатства и разнообразия, которое отличает сознательный чувственный опыт, получаемый «в природе». Но чтобы выбраться из кабинетов на широкие просторы, нужно изменить сам образ мысли.

Одним относительно недавним летним днем я впервые в жизни положил под язык крохотную дозу ЛСД и вытянулся на траве — дожидаться, что будет дальше. День был теплый, дул легкий ветерок, на бело-голубом небе вихрились бледные завитки облаков. Примерно через полчаса, точно так же, как у Альберта Хофмана много лет назад, мир зашевелился и начал преображаться. Холмы, небо и облака принялись пульсировать, становясь ярче, обволакивая меня, сплетаясь и переплетаясь со мной, почти как живые. Я, как положено ученому, пытался вести записи, но, заглянув в них на следующий день, понял, что попытки эти довольно скоро заглохли. Из того, что мне все-таки запомнилось, — облака принимали хотя и подвижные, но отчетливые формы, которые, по крайней мере отчасти, как будто мне подчинялись. Как только облако начинало напоминать лошадь, или кошку, или человека, я без особого труда мог это сходство усилить и подчеркнуть, иногда с умопомрачительным результатом. В какой-то момент по небу прошествовала целая процессия Силл Блэк[21].

Галлюцинации, порождаемые ЛСД, неплохо переубеждают любого сомневающегося в том, что чувственный опыт создается именно мозгом. Еще несколько дней после этого эксперимента мне казалось, что я вижу свои перцептивные впечатления «насквозь» и хотя бы отчасти воспринимаю их так, как им и подобает, то есть как конструкторы. Я по-прежнему ощущал вместе с перцептивным посланием отголоски средства его передачи.

Разумеется, разглядеть в облаках лицо можно и без допинга — по крайней мере увидеть намеки и наброски, тени и контуры, проецируемые на небо. Способность складывать из разрозненных предметов общий образ называется парейдолией (от греческих слов «рядом» и «изображение»). В силу значимости распознавания лиц для человека — и для некоторых других животных — у нашего мозга имеются сильные предустановленные ожидания, связанные с лицами. Именно поэтому мы склонны так часто видеть лица в неодушевленных предметах, будь то облака, кусок гренки или даже старая раковина, как на фотографии ниже¹⁵. И поскольку для нас это обычное явление, мы, как правило, не считаем парейдолию галлюцинацией. Когда шизофреник слышит голос, велящий ему изувечить себя или объявляющий его новым воплощением Христа, и, кроме него, этот голос никто не слышит, — это другое дело, это явная галлюцинация. И увиденное мной под ЛСД шествие Силл Блэк по небу тоже было галлюцинацией.



Рис. 11. Раковина, напоминающая нам лицо

Как мы теперь знаем, было бы ошибочно считать любое из этих явлений — даже самое на первый взгляд диковинное — в корне отличающимся от обычной выдачи наиболее вероятных перцептивных предположений. Любой наш чувственный опыт, независимо от того, обозначаем ли мы его как галлюцинаторный, всегда и всюду основан на проекции перцептивных ожиданий на нашу сенсорную среду. То, что мы называем галлюцинацией, возникает, когда наши перцептивные априорные убеждения оказываются настолько сильны, что одерживают верх над сенсорными данными и мозг начинает терять связь с источниками этих данных в окружающем мире.

Вдохновленная этой общностью нормального восприятия и галлюцинаторного, наша лаборатория принялась искать новые способы исследования, как из наиболее вероятных перцептивных предположений возникает перцептивный опыт, и наши эксперименты завели нас туда, куда нам и не снилось.

Если, выйдя из моего кабинета, подняться на один этаж, то, поплутав в закоулках бывшего химфака, вы окажетесь в одной из наших импровизированных лабораторий — ее назначение подсказывает приклеенный на дверь клочок бумаги с надписью «Лаб. VR/AR». Здесь, обратившись к услугам стремительно развивающихся технологий виртуальной и дополненной реальности (VR/AR), мы исследуем восприятие человеком мира и себя самого невозможными прежде способами. Несколько лет назад мы решили сконструировать «галлюцинаторную машину»¹⁶ и посмотреть, получится ли у нас породить галлюцинаторноподобный опыт в экспериментально контролируемой обстановке, симулируя сверхактивные перцептивные априорные предположения. Проект возглавил Кейсукэ Судзуки, старший научный сотрудник нашей лаборатории и штатный специалист по виртуальной реальности.

Для начала мы сняли камерой, дающей обзор на 360°, панорамное видео реальной среды. В качестве панорамы была выбрана главная площадь университетского кампуса во вторник, в обеденное время, когда студенты и преподаватели бродят по еженедельно раскидывающемуся там продуктовому рынку. Затем мы обработали видео с помощью программы, созданной Кейсукэ на основе гугловских алгоритмов Deep Dream («глубокого сна»), создавая симулируемую галлюцинацию.

Алгоритм «глубокого сна» предполагает использование «наоборот» искусственной нейронной сети, обученной распознаванию объектов на изображениях. Подобные сети состоят из множества слоев симулированных нейронов, связи между которыми устроены так, что они отчасти напоминают восходящие пути биологической

зрительной системы¹⁷. Поскольку эти сети содержат только восходящие связи, их легко обучать с использованием стандартных методов машинного обучения. Нашу сеть обучили идентифицировать более тысячи разных видов объектов на изображениях, включая множество разных пород собак. Она отлично справляется, с успехом различая даже разновидности хаски, которые для меня все похожи как две капли воды.

Обычно эти нейросети используют так: предъявляют изображение и спрашивают, что оно, «по мнению» сети, содержит. Алгоритм «глубокого сна» идет от противного: нейросети сообщают, что изображение содержит некий объект, и она соответствующим образом обновляет картинку¹⁸. Иначе говоря, алгоритм проецирует перцептивный прогноз на изображение, в избытке внося в нее зрительский вклад. Для нашей галлюцинаторной машины мы покадрово обработали таким образом все панорамное видео и добавили еще несколько примочек, чтобы обеспечить непрерывность, сглаживание и так далее. Это видео в формате «глубокого сна» затем воспроизводилось в шлеме VR, обеспечивая испытуемому, оглядывающемуся вокруг, полное погружение в виртуальную реальность. Так появилась на свет галлюцинаторная машина.

Опробуя ее впервые, я не подозревал, что действо окажется настолько захватывающим¹⁹. Хотя его нельзя сравнить с настоящим наркотическим трипом или (насколько мне известно) психической галлюцинацией, мир тем не менее изменился в процессе существенно. На этот раз обошлось без Силлы Блэк, зато было много собак и частей собачьего тела, которые органично вписывались в окружающий меня пейзаж, — совсем не так, как вышло бы, если бы мы просто вмонтировали изображения собак в имеющееся видео (см. черно-белый стоп-кадр на рис. 12). Сила галлюцинаторной машины — в ее способности симулировать вывод из нисходящих наиболее вероятных предположений о присутствии рядом собак и тем самым резюмировать в сильно утрированном виде процесс, в ходе которого мы воспринимаем и интерпретируем зрительные сцены в реальном мире.



Рис. 12. Стоп-кадр из видео для «галлюцинаторной машины»

Слегка варьируя программу галлюцинаторной машины, можно варьировать и симулируемый галлюцинаторный опыт. Если, например, сосредоточить активность в одном из средних уровней сети, а не на уровне исходящих действий, мы получим галлюцинации частей объекта, а не целого. В этом случае разворачивающаяся перед вами сцена будет кишеть глазами, ушами и лапами, то есть все ваше зрительное пространство заполнит мешанина из частей собачьего тела. Если сосредоточить активность на еще более низких уровнях, мы получим галлюцинации, которые лучше

всего обозначаются как «геометрические» — в них необычайно ярко, живо и выпукло проступают низкоуровневые черты зрительной среды, такие как контуры, линии, текстура и рисунок.

Галлюцинаторная машина — это упражнение в «вычислительной феноменологии»: в перекидывании объяснительных мостов от механизмов к свойствам перцептивного опыта с помощью вычислительных моделей. Ее непосредственная ценность заключается в том, чтобы сопоставить вычислительную архитектуру прогнозного восприятия с феноменологией галлюцинации. Так мы сможем сделать шаг к пониманию того, почему определенные разновидности галлюцинаций именно таковы, каковы они есть²⁰. Но за этим применением скрывается более глубокое и для меня более интересное предположение, что, пролив свет на галлюцинации, мы сможем лучше разобраться и в нормальном, повседневном перцептивном опыте. Галлюцинаторная машина четко — прямо, живо и мгновенно — показывает, что галлюцинации, как мы их называем, — это разновидность неконтролируемого восприятия. А нормальное восприятие, совершающееся здесь и сейчас, — это действительно разновидность контролируемой галлюцинации.

Может показаться, что концепция контролируемой галлюцинации сводится к таким объяснениям, как «я вижу стол, потому что это наиболее вероятное предположение моего мозга о причинах текущих входящих сенсорных данных». (Вместо стола можно подставить лицо, кошку, собаку, красное кресло, брата жены, авокадо или Силлу Блэк.) Я думаю, мы можем пойти гораздо дальше и найти объяснение тому, что мне нравится называть «глубинной структурой» восприятия, — тому, как содержание сознания возникает в нашем чувственном опыте, во времени, в пространстве и в разных модальностях.

Возьмем такое банальное вроде бы наблюдение, что наш зрительный мир в основном состоит из объектов и пространства между ними. Глядя на чашку кофе на своем столе, я в определенном смысле воспринимаю и ее обратную сторону, хотя в данный момент она от меня скрыта. Чашка видится как занимающая некоторый объем пространства, в отличие от чашки на фотографии или рисунке. Это феноменология «объектности». Объектность представляет собой свойство того, каким содержание сознательного зрительного восприятия представляется в общем, а не в качестве отдельного сознательного чувственного опыта.

Объектность хотя и распространенная черта зрительного опыта, но не универсальная. Если в солнечный день вы посмотрите на безоблачный, равномерно голубой небосвод, у вас не возникнет впечатления, что небо — это отдельный объект. И цветные круги, которые заплывут у вас перед глазами, когда вы посмотрите на солнце, а потом отведете взгляд, тоже будут восприниматься не как объект, а как временное нарушение зрения. То же самое относится и к другим модальностям: люди, страдающие тиннитусом, не воспринимают донимающий их шум или звон в ушах как относящийся к реальной окружающей действительности.

Значимость объектности для человеческого восприятия давно осознали художники. Рене Магритт в своей уже набившей оскомину картине «Вероломство образов» (рис. 13) исследует разницу между предметом и изображением предмета. Пабло Пикассо на протяжении значительной части своего кубистского периода изучает, насколько на наше восприятие объектности влияет взгляд от первого лица. Его произведения дробят и перемешивают предметы самым разным образом, представляя их сразу с нескольких точек зрения. Можно расценивать эти картины и другие им подобные как исследование принципов объектности в свете вклада зрителя²¹. Работы Пикассо, в частности, вовлекают адресата в процесс творческого создания перцептивных объектов из хаоса вероятностей. Как выразился философ Морис Мерло-Понти, при

помощи своей картины художник исследует средства, позволяющие объекту стать для нас видимым²².



Рис. 13. Рене Магритт «Вероломство образов» (1929)

© ADAGP, Paris and DACS, London 2021.

В когнитивной науке феноменологию объектности всесторонне и тщательно изучает «теория сенсомоторных совпадений»²³. Согласно этой теории, переживаемый нами чувственный опыт зависит от того, насколько «мастерски» действия меняют входящую сенсорную информацию. Когда мы что-то воспринимаем, содержание этого восприятия не передается сенсорными сигналами, а возникает из подспудного имплицитного знания мозга о том, как стыкуются действия и ощущения. В рамках этой концепции зрение и остальные перцептивные модальности — это действия, которые производит организм, а не пассивное закачивание данных в централизованный «разум».

В главе 4 мы описывали ощущение красного с точки зрения создаваемых в мозге прогнозов о том, как поверхности отражают свет. Давайте распространим это объяснение на объектность. Если я держу перед собой помидор, я воспринимаю его как имеющий обратную сторону, в отличие от тех случаев, когда я смотрю на изображение помидора (или на изображение трубки, как на картине Магритта), на безоблачное синее небо или когда у меня перед глазами вертятся цветные круги после взгляда на солнце. Согласно теории сенсомоторных совпадений, даже не видя в данный момент обратную сторону помидора, я все равно перцептивно ее осознаю благодаря «вшитому» в мой мозг имплицитному знанию о том, как изменятся входящие сенсорные сигналы, если я поверну помидор другой стороной.

Необходимая прошивка имеет форму порождающей модели. Как мы знаем из предшествующей главы, порождающие модели могут предсказывать сенсорные последствия действий. Эти прогнозы «условны» или «контрфактуальны» в том смысле, что они описывают происходящее или гипотетически происходящее с сенсорными сигналами при том или ином действии. Несколько лет назад я написал научную статью, в которой высказал предположение, что феноменология объектности зависит от богатства этих условных или контрфактуальных прогнозов²⁴. Порождающие модели, кодирующие множество различных прогнозов такого рода (например, что кожица помидора одинаково красная по всей поверхности), обеспечивают сильную феноменологию объектности. Те же порождающие модели, которые кодируют лишь несколько подобных прогнозов или ни одного (как в случае с гладким и плоским синим небом или цветными кругами после взгляда на солнце), дают слабую объектность или не дают никакой.

Еще один случай, для которого типично отсутствие объектности, — «графемно-цветовая синестезия»²⁵. Синестезия — это что-то вроде «смещения чувств». Обладатели графемно-цветовой ее разновидности видят буквы в цвете: например, буква А может вызывать ощущение яркого красного независимо от того, каким цветом в действительности напечатана или раскрашена буква. И хотя эти ощущения цвета возникают устойчиво и непроизвольно (то есть буква каждый раз «окрашивается» в один и тот же цвет и для этого не требуется никаких сознательных усилий), синестеты обычно не путают свои синестетические цвета с настоящими, «существующими в окружающем мире». Я объясняю это тем, что за синестетическими цветами не скрывается, в отличие от «настоящих», богатый репертуар сенсомоторных прогнозов. Синестетический «красный» не особенно варьирует в зависимости от ваших движений или от смены освещения, поэтому феноменология объектности здесь отсутствует.



Рис. 14. Виртуальные объекты, которые должны выглядеть незнакомыми и непривычными

Некоторые из этих идей мы начали проверять в своей лаборатории виртуальной реальности. В ходе недавнего эксперимента²⁶ мы создали ряд намеренно непривычных и незнакомых виртуальных объектов, состоявших из бесформенных сгустков с выступами, и демонстрировали их участникам посредством виртуального шлема (см. рис 14). Мы модифицировали метод подавления стимула вспышкой, использовавшийся в нашем предшествующем эксперименте с домом/лицом, сделав так, чтобы каждый объект был поначалу невидим, но в конце концов прорывался в сознание. Если в эксперименте с домом/лицом мы манипулировали ожиданием по поводу каждого изображения, то в эксперименте с виртуальной реальностью у нас появилась возможность манипулировать достоверностью сенсомоторных прогнозов, меняя реакцию объектов на действие. Испытуемые вращали виртуальные объекты с помощью джойстика, и мы могли задать характер этого вращения — либо соответствующий тому, как объект вел бы себя в реальном мире, либо не соответствующий, и тогда объект вращался как попало, в непредсказуемых направлениях. Мы предполагали, что виртуальные объекты, ведущие себя нормально, прорвутся в сознание быстрее, чем нарушающие сенсомоторные прогнозы, — и именно такие результаты и получили. Да, эксперимент, возможно, далек от совершенства, поскольку скорость выхода в осознанное восприятие может служить лишь приблизительным аналогом феноменологии объектности. И все же он демонстрирует, что достоверность сенсомоторных прогнозов способна вполне определенным измеримым образом на это восприятие влиять.

В число многих интуитивных, но ошибочных представлений о восприятии входит и такое: изменения в нашем восприятии непосредственно соответствуют изменениям в реальном мире. Однако изменения, как и объектность, — это еще одно проявление глубинной структуры перцептивного опыта. Изменения в восприятии вызываются не просто изменением сенсорных данных. Мы воспринимаем изменения в силу тех же принципов наиболее вероятного предположения, которым обязаны своим существованием другие составляющие восприятия.

Как показал уже не один эксперимент, физические изменения, происходящие в окружающем мире, не являются ни необходимыми, ни достаточными для того, чтобы изменилось восприятие. Ярким примером служит «змеистое» изображение (рис. 15)^[27], на котором ничего не двигается, однако возникает перцептивное впечатление движения, особенно если позволить взгляду поблуждать по всей поверхности. Происходит это потому, что мелкие детали рисунка, видимые периферийным зрением — краем глаза, что называется, — убеждают зрительную кору делать вывод о движении, хотя движения никакого нет.

Противоположная ситуация — физическое изменение без перцептивного — возникает при «слепоте к изменениям». Такое может происходить, когда какие-то составляющие среды меняются слишком медленно или когда меняется все сразу, но значение имеют лишь отдельные черты. В одном впечатляющем видеопримере^[22] цвет меняет вся нижняя половина изображения, но, поскольку все это растянуто во времени примерно секунд на сорок, большинство зрителей не замечают превращения совсем, даже если смотрят прямо на ту часть кадра, которая из красной становится фиолетовой. (Однако это только в том случае, если им не сообщают заранее, что цвет будет меняться. Если же знать, куда смотреть и чего ждать, перемену легко заметить.) Этот пример в чем-то схож с примером слепоты невнимания, который я приводил в предыдущей главе (когда люди не замечают гориллу, возникающую в кадре в разгар игры в баскетбол). Разница между примерами в том, что при слепоте к изменениям человек не замечает самой «перемены».

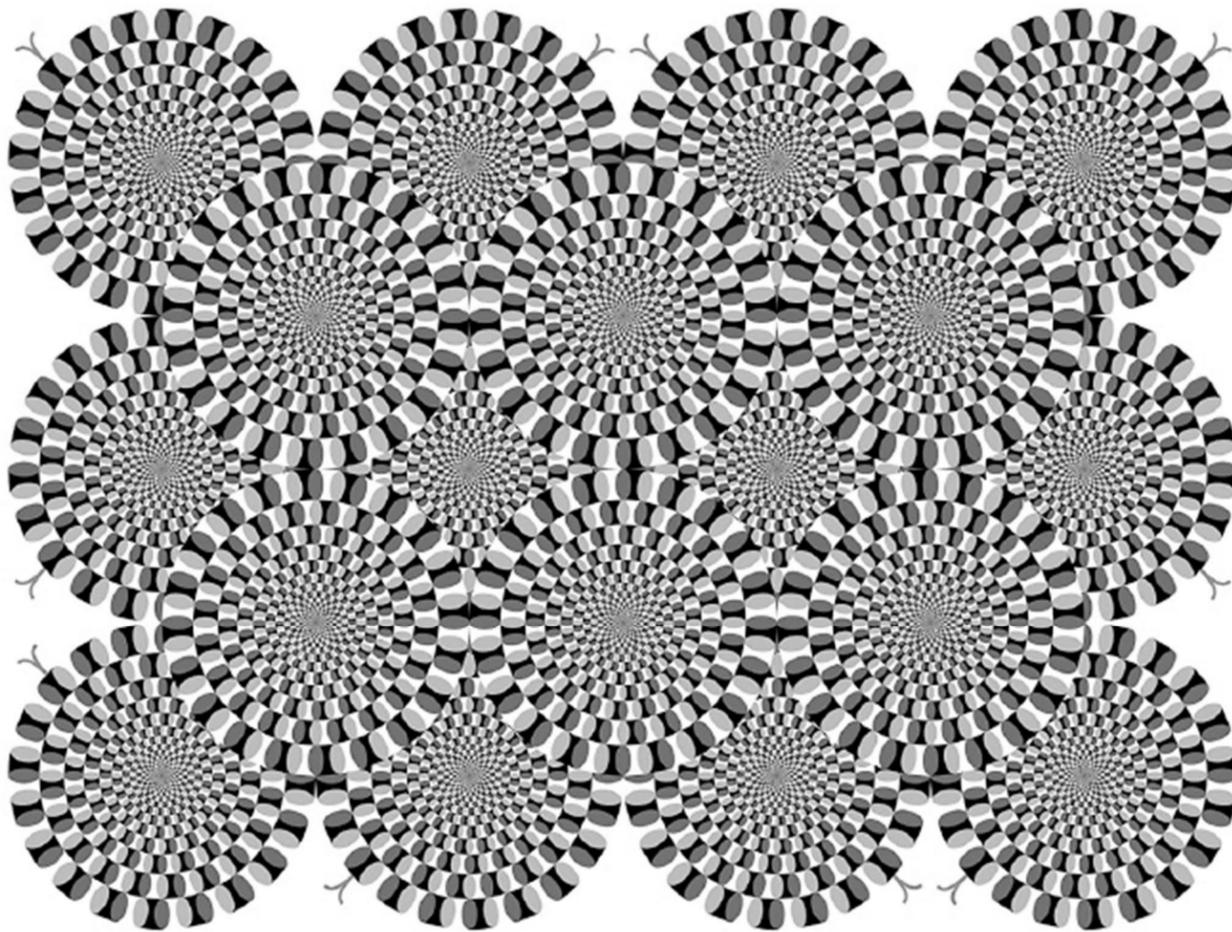


Рис. 15. Иллюзия «Вращающиеся змейки»^[23]

Автор: Акиёси Китаока

Некоторые усматривают в слепоте к изменениям философскую дилемму: после смены цвета вы по-прежнему воспринимаете нижнюю часть кадра красной (хотя она уже фиолетовая), или теперь вы воспринимаете ее фиолетовой, но тогда какой же вы воспринимали ее раньше, учитывая, что саму смену цвета вы не ощутили? Чтобы разрешить эту дилемму, нужно отвергнуть исходную ее посылку и признать, что восприятие изменений и изменение восприятия — это не одно и то же. Восприятие изменений — это еще одно перцептивное умозаключение, еще одна разновидность контролируемой галлюцинации.

И если восприятие изменений — это перцептивное умозаключение, то и восприятие времени тоже.

Время — одна из самых сложных и запутанных тем и в философии, и в физике, и в нейронауках. Если физики пытаются разобраться, что оно собой представляет и почему течет (если действительно течет), то перед нейрочеными задача стоит не менее мудреная. Все наши перцептивные переживания происходят «во времени» и «со временем». Даже наше ощущение настоящего момента всегда словно размазывается между относительно фиксированным прошлым и частично открытым будущим. Для нас время тоже течет — впрочем, иногда ползет, а иногда летит или мчится.

Мы переживаем секунды, часы, месяцы и годы, но никакого «датчика времени» в нашем мозге нет. Если для зрения у нас имеются фоторецепторы в сетчатке, а для слуха — «волосковые клетки» в ушах, то для восприятия времени никакой специальной сенсорной системы не существует. Более того, если не принимать в расчет циркадные ритмы (побочным эффектом которых выступает пресловутый джетлаг — ухудшение самочувствия при смене часовых поясов), нет никаких свидетельств, что у нас в голове имеются некие «нейронные часы», отмеряющие воспринимаемое нами время. Но если бы они существовали, то являли бы собой идеальный пример «двойного преобразования», как называл его Дэниел Деннет, то есть воспроизведения внутри мозга тех или иных свойств окружающего мира для предполагаемого внутреннего наблюдателя. В действительности же восприятие времени — это такая же контролируемая галлюцинация, как и восприятие изменений и все остальное наше восприятие²⁸.

Вот только чем контролируемая? Если нет специального сенсорного канала, который мог бы обеспечивать эквивалент сенсорных ошибок прогнозирования? Простое и элегантное предположение высказал мой коллега — когнитивист Уоррик Роузбум, который пришел к нам в Центр в 2015 г. и сейчас возглавляет собственную научно-исследовательскую группу, занимающуюся восприятием времени. Его гипотеза состоит в том, что умозаключения о времени мы основываем не на тиканье внутренних часов, а на темпах изменений перцептивного содержания других модальностей, — и он придумал хитрый способ эту гипотезу проверить.

Под руководством Уоррика наша группа записала библиотеку видеоматериалов разной продолжительности в разных контекстах — оживленная городская улица, пустой офис, несколько коров, пасущихся на лужайке неподалеку от университета. Затем мы демонстрировали эти видео участникам эксперимента и просили оценить, сколько длился каждый ролик. Все испытуемые продемонстрировали в своей оценке когнитивное искажение характеристики²⁹: склонность преуменьшать длительность долгих роликов и преувеличивать длительность коротких. Дали о себе знать и искажения, связанные с контекстом сцены: более насыщенные сюжеты казались испытуемым длиннее спокойных, хотя на самом деле те и другие имели одинаковую продолжительность.

Затем Уоррик показал те же видео искусственной нейросети, имитировавшей работу человеческой зрительной системы. Собственно, это была та же сеть, которую мы использовали для галлюцинаторной машины. Оценка длительности для каждого видео вычислялась на основе, грубо говоря, накапливаемого темпа изменения активности самой сети. Для этой оценки не требовалось никаких «внутренних часов». Примечательно, что оценки нейросети практически совпали с человеческими, демонстрируя те же предубеждения по поводу длительности и контекста³⁰. Это значит, что восприятие времени может, по крайней мере в принципе, проистекать из «наиболее вероятного предположения» о темпе изменения сенсорных сигналов, совершенно не нуждаясь ни в каких внутренних хронометрах.

Некоторое время назад мы сделали следующий шаг в этом исследовании, начав искать свидетельства данного процесса в самом мозге. В проекте, который возглавила постдокторантка Максин Шерман, мы фиксировали с помощью фМРТ активность мозга людей, смотревших тот же самый набор видеороликов и оценивавших их длительность³¹. Нам нужно было узнать, можно ли на основании активности зрительной коры спрогнозировать воспринимаемую длительность каждого видео, как нам удалось сделать в предшествующем исследовании под руководством Уоррика с помощью компьютерной модели зрительной системы. Как выяснила Максин, можно. Активность зрительной системы мозга, в отличие от активности других его областей, позволяла очень точно спрогнозировать субъективную длительность. Это весомое свидетельство в пользу того, что восприятие длительности действительно возникает из наиболее вероятных перцептивных предположений, а не из тиканья нейронных часов.

Другие эксперименты, которые могли бы выявить наличие у нас «внутреннего хронометра», ничего подобного не выявили. Мой безусловный фаворит в этой категории (остальные рядом не стояли) — эксперимент, участники которого прыгали со строительных кранов³². Нейробиолог Дэвид Иглмен задался целью проверить распространенное интуитивное представление о том, что субъективное время замедляется в острые драматические моменты, например перед автокатастрофой. Он рассудил, что причиной этого субъективного замедления может быть ускорение хода внутренних часов, то есть за отрезок времени они успевают натикать больше, и поэтому отрезок этот в нашем восприятии удлиняется. Это, в свою очередь, ведет к «ускорению» темпа восприятия, поскольку ускорение часов должно означать улучшение способности воспринимать короткие отрезки происходящего.

Чтобы проверить эту гипотезу, Иглмен и его группа разработали особые цифровые часы, которые показывали ряд цифр, мелькающих так быстро, что прочитать их в нормальных условиях было невозможно. Затем он уговорил нескольких смельчаков-добровольцев на вызывающие выброс адреналина жуткие прыжки в бездну, во время которых они должны были смотреть на эти часы с мелькающими на дисплее цифрами. Если внутренние часы и вправду ускоряются, то, по логике Иглмена, добровольцы-испытуемые во время свободного падения должны были увидеть вместо нечитаемого вихря на дисплее обычные цифры. Этого не произошло, и наличие у нас внутренних часов этот эксперимент не подтвердил. Конечно, отсутствие доказательства — это не доказательство отсутствия, но каков эксперимент!

На переднем крае наших исследований глубинной структуры перцептивного опыта находится проект, посвященный изучению восприятия «реальности» как таковой. В этом проекте, которым руководит талантливый докторант Альберто Мариола, используется новая разработка, получившая название «замещающая реальность»³³. Среду виртуальной реальности, даже с полным погружением, всегда можно отличить от реальной. Участники эксперимента с нашей галлюцинаторной машиной всегда, как бы ни затягивала их эта психоделия, знали, что она не настоящая. Замещающая реальность пытается этот недостаток преодолеть. Задача заключалась в том, чтобы создать систему, в которой человек воспринимал бы окружающую его среду как реальную и верил в ее реальность, хотя реальной она не является.

Идея проста. Как и в эксперименте с галлюцинаторной машиной, мы записали панорамное видео, но на этот раз снимали ту самую лабораторию виртуальной/дополненной реальности, в которой проводим свои исследования. Участника, подходящего в лабораторию, усаживают на табурет в центре комнаты и надевают шлем виртуальной реальности с прикрепленной спереди камерой. Затем ему предлагается оглядеться вокруг и осмотреть комнату через камеру на шлеме. В

какой-то момент мы переключаем подачу изображения, и камера начинает показывать не происходящее на самом деле в данную секунду, а видеозаготовку. Надо сказать, что большинство людей в такой ситуации продолжают воспринимать видимое в камере как реальность, хотя это уже не она.

Эта разработка дает нам возможность проверять гипотезы, касающиеся условий, при которых человек воспринимает окружающую среду как реальную и, что, возможно, еще важнее, при которых этот обычно универсальный аспект сознательного опыта начинает распадаться³⁴. Подобное может происходить и происходит не только в случае, например, возникновении послеобразов (как при взгляде на солнце) — глобальная утрата воспринимаемой реальности мира и самого себя характерна и для таких инвалидизирующих психиатрических заболеваний, как деперсонализация и дереализация.

Исследование того, что придает перцептивному опыту «реальность», возвращает нас к витгенштейновской идее коперниканской революции: даже если мы понимаем, как все обстоит на самом деле (Земля вращается вокруг Солнца, восприятие — это контролируемая галлюцинация), выглядеть все по-прежнему будет так, как выглядело всегда. Когда я смотрю на красное кресло в углу комнаты, его краснота (и субъективная «кресельность») по-прежнему кажется реально существующим, достоверным свойством действительности, независимой от разума, а не замысловатым построением мозга, выводящего наиболее вероятные предположения.

Давным-давно, еще в XVIII в., Дэвид Юм сделал аналогичное наблюдение по поводу каузальности — еще одной универсальной черты нашего восприятия мира. Юм доказывал, что физической каузальности как объективного свойства окружающей действительности не существует, мы лишь «проецируем» каузальность на собственный мир на основании повторяющегося восприятия того, что происходит почти непрерывно, в плотной временной последовательности. Мы не наблюдаем и не можем наблюдать «каузальность» в окружающем мире непосредственно. Да, в мире что-то происходит, но то, что мы воспринимаем как каузальность, — это перцептивное умозаключение, в том же самом смысле, в котором все наше восприятие — это проекции структурированных ожиданий нашего мозга на нашу сенсорную среду, то есть упражнения во вкладе зрителя. Как писал Юм, разум имеет огромную склонность распространяться во внешний мир, и мы «золотим и пятнаем» естественные объекты «красками, заимствованными у внутренних ощущений»³⁵. И не только красками: и формы, и запахи, и «кресельность», и изменения, и продолжительность, и каузальность — все фигуры и фоны наших перцептивных миров — всё это юмовские проекции, составляющие контролируемой галлюцинации.

Почему мы воспринимаем наши перцептивные конструкторы как объективную реальность? С точки зрения теории контролируемой галлюцинации задача восприятия — направлять действия и поведение, чтобы повысить шансы данной особи на выживание. Мы воспринимаем мир не таким, какой он есть, а таким, какой он нам полезен. Поэтому вполне логично, что феноменологические свойства, такие как краснота, «кресельность», «силлаблэковость» и каузальность, кажутся объективными, достоверными, присущими реально существующей внешней среде. Воспринимая происходящее в мире как реально существующее, мы будем реагировать на это быстрее и эффективнее. Это имманентное нашему перцептивному опыту представление об окружающем мире как о действительно существующем «там», за пределами нашего тела, выступает, на мой взгляд, необходимым свойством порождающей модели, способной предугадывать входящие сенсорные данные, чтобы успешно направлять поведение.

Иначе говоря, хотя перцептивные свойства зависят от нисходящих порождающих моделей, мы не воспринимаем модели как модели. Скорее, мы воспринимаем

посредством наших порождающих моделей³⁶, и в процессе этого восприятия из обычного механизма возникает структурированный мир.

В начале книги я обещал, что с помощью настоящей проблемы мы будем постепенно подкапываться под трудную проблему — почему и как какой бы то ни было физический механизм может породить сознательный опыт, соотноситься с ним или быть ему тождественным. Получается ли у нас?

Получается. Начав с принципа, что мозг должен делать умозаключения о скрытых источниках своей входящей сенсорной информации, мы пришли к новому пониманию, почему и как наша внутренняя вселенная оказывается заполненной всем тем — от чашек кофе до оттенков цвета и каузальности, — что кажется нам свойствами внешней объективной действительности, где это впечатление само по себе составляет свойство перцептивного вывода. И именно это свойство «казаться реальным» подогревает дуалистические интуитивные представления о том, как взаимосвязаны сознательный опыт и физический мир, — представления, которые, в свою очередь, подводят нас к идее трудной проблемы. Именно потому, что наше восприятие носит феноменологический характер «реальности», так необычайно трудно признать, что на самом деле перцептивный опыт не обязательно должен (если вообще должен) непосредственно соотноситься с тем, что существует независимо от сознания. Кресло существует независимо от сознания, а кресельность — нет.

Как только мы это уясним, нам станет проще признать трудную проблему не такой уж трудной — или даже совсем не проблемой. Верно и обратное: если трактовать содержание нашего перцептивного опыта как реально существующее во внешнем мире, трудная проблема сознания покажется особенно трудной. А ведь именно к такой автоматической трактовке подталкивает нас феноменология нормального сознательного восприятия.

Как было и с изучением жизни столетие назад, необходимость найти «особый соус» для сознания снижается прямо пропорционально нашей способности различать разные составляющие сознательного опыта и объяснять их с точки зрения механизмов, находящихся в их основе. Развеивать трудную проблему постепенно — это не то же самое, что решить ее разом или безоговорочно опровергнуть, но это оптимальный способ продвижения вперед, и он гораздо лучше, чем превозносить сознание как волшебную тайну или отмахиваться от него как от метафизически иллюзорной не-проблемы. И когда мы учитываем, что перцептивными построениями выступают не только картины нашего восприятия мира, наше исследование набирает обороты.

Самое время спросить: кто или что осуществляет всю эту перцепцию?



Глава 7

Делирий

Летом 2014 г. в отделении экстренной хирургии оксфордской больницы им. Джона Рэдклиффа моя мама впала в вегетативное состояние. У нее развилась энцефалопатия — заболевание головного мозга. Причину так окончательно и не установили; маму клали в больницу с раком кишечника, и неврологические проблемы стали полной неожиданностью. Я срочно прилетел с конференции в Брисбене, опасаясь самого худшего. Мама выкарабкалась, хотя и медленно, однако этот ее нарастающий с каждым часом распад сознания я не забуду никогда. Сама она, к счастью, почти ничего из этого не помнит.

Прошло четыре года. На дворе 2018-й, разгар небывалой жары и очередного летнего чемпионата мира по футболу. На этот раз никакого вегетативного состояния. Теперь у мамы, которой уже 83, возник так называемый больничной делирий — другая разновидность распада ощущения себя и окружающего мира. За две недели до этого ее срочно привезли в клинику Рэдклиффа с рецидивом выраженных болей в кишечнике. Спустя два дня после поступления в больницу, пока врачи пытались понять, удастся ли обойтись без операционного вмешательства, у нее начались сильные галлюцинации и бред, и я приехал из Брайтона, чтобы быть рядом.

Всплывшее как будто из средневековых трактатов слово «делирий» происходит от латинского delirare — «быть вне себя, безумствовать». В словаре делирий обозначается как «острое состояние расстроенного сознания, характеризующееся беспокойством, наличием иллюзий и бессвязностью речи». В отличие от деменции, которая представляет собой хроническое дегенеративное состояние, делирий рано или поздно отступает, хотя длиться может неделями. По мне, так от этого слова веет сумасшедшими домами викторианской эпохи, и поэтому так странно видеть его в официальных диагнозах в английской больнице XXI в. Однако, если подумать, ничего удивительного здесь нет — просто очередное напоминание о том, что наша психиатрия еще только в самом начале пути.

Мамино состояние это словарное определение описывает вполне точно. Когда я прихожу в палату, она сидит, сгорбившись, в кресле, без улыбки, растрепанная, помятая, с пустым безразличным взглядом. Рассказывает мне о людях, которые ползают по стенам, говорит, что не помнит, где она и почему она здесь. Связь с реальностью и понимание того, кто она такая, слабеют.

В пятницу становится совсем плохо. Мама никому не верит, убежденная, что над ней проводят грандиозный жестокий эксперимент, а мы — в том числе и я, поскольку в ее паранойе я часто оказываюсь зачинщиком творящихся безобразий, — намеренно вызываем у нее галлюцинации с помощью лекарств, которые с какими-то неведомыми, но явно нехорошими целями заставляем ее принимать. Обаятельная и добрая в обычной жизни, сегодня она возмущенно рывкает на медсестер, требует, чтобы ее выписали, несколько раз пытается сбежать и велит врачам убрать ее сына — сумасшедшего ученого — с глаз долой. Это не моя мама. Эта женщина выглядит точь-в-точь как она, но это не мама.

В число факторов, провоцирующих больничной делирий, входит эпилепсия, инфекции, обширное оперативное вмешательство (или состояние, требующее такого вмешательства), жар, обезвоживание, недоедание и недосыпание, побочные эффекты медицинских препаратов и, что важно, незнакомая и непривычная обстановка. У мамы все это разом, полный набор. И именно из-за незнакомой обстановки этот вид делирия называется «больничным».

Немного найдется на свете мест, настолько способствующих отрыву от реального мира, как отделение неотложной хирургии. Постоянный писк приборов, мигающие огоньки и лампочки, почти полное отсутствие признаков существования внешнего мира — разве что краешек ухватишь в окне, если повезет, — все сжимается до пределов койки, кресла, может, коридора. Вокруг товарищи по несчастью в

состоянии разной степени стресса и расстройства и нескончаемый парад разных, но неотличимых друг от друга медсестер, ординаторов и консультантов. Один день повторяет другой. Делирий, хотя и являет собой неотложное состояние, часто таковым не признается, и справиться с ним не пытаются. Пациенты поступают в больницу с физическим заболеванием, так что лечат именно его, а не возникающие иной раз попутно расстройства психики и мозга.

Больничный делирий развивается едва ли не у трети пожилых пациентов, поступающих в отделение неотложной помощи, а среди тех, кого оперируют, эта доля еще выше¹. И хотя со временем делирий обычно проходит, он может оставлять довольно ощутимые отголоски, в том числе ухудшение когнитивных способностей, повышение вероятности смерти в последующие месяцы и риска новых эпизодов делирия и деменции². Я съездил к маме домой, в тот самый сельский дом, где вырос, набрал знакомых и привычных ей предметов в надежде, что они помогут сориентироваться заново, — что-то вроде обломков кораблекрушения, за которые можно уцепиться, барахтаясь в волнах. Фотография в рамке, очки, кофта, старый мягкий лев из моего далекого детства.

Бредовые иллюзии редко бывают случайными. В мамином делирии есть своя логика, хотя и кривая. Мама верит — нет, знает, — что стала жертвой бесчеловечного эксперимента, в котором против нее сговорились все, включая меня. Да, я действительно провожу эксперименты с участием людей, и, когда я прихожу к маме в больницу, роль у меня получается двойственная: я одновременно и сын, и врач, я вроде бы успокаиваю ее, но при этом постоянно читаю выписки и справки и перешептываюсь с консультантами и ординаторами насчет паранеопластической энцефалопатии и еще каких-то непрогнозируемых страшных вещей. Мозг всегда пытается прийти к какому-то выводу, выстроить наиболее вероятное предположение.

Среди изменений в мамином поведении есть почти незаметные. Например, она произносит каждое слово во фразе отдельно, а не связным потоком. «Я не могу найти очки, не знаю, где они». Такое состояние длится еще много дней после того, как отступает первоначальный делирий. Оно тоже то усиливается, то ослабевает. Сегодня вечером у нас откат, очередное ухудшение, рушащее мои надежды, что скоро ее выпишут.

Моя собственная жизнь тоже начинает казаться слегка ненастоящей. Я у мамы единственный близкий родственник, так что быть с ней рядом, кроме меня, некому. Утро и вечер я провожу в больнице, а днем, если повезет, разгребаю завалы на работе, гуляю и плаваю в Темзе. В такие дни я отправляюсь на Порт-Медоу — огромный заливной луг, где есть гуси, лебеди, пасутся коровы и дикие лошади. Невероятное место даже для Оксфорда, в котором наслаиваются друг на друга самые разные миры. Жара длится уже не первую неделю, поэтому обычно хлюпающая под ногами земля сейчас напоминает африканскую саванну. Сегодня утром за мной погналась лошадь, и, когда я шел по мосту через железную дорогу обратно в город, сердце в груди колотилось как бешеное.

Сегодня 14-й день, как мама в больнице, для меня 12-й. Острая спутанность сознания прошла, но она все еще не совсем в себе, ее состояние неустойчиво. Однако теперь она изумляется моим рассказам о том, как она думала, будто я провожу над ней эксперименты, и просила убрать меня отсюда. Я держу ее за руку, убеждаю, что все будет хорошо, и надеюсь, что она окончательно придет в себя.

Но что значит «в себя»? Это что-то, из чего можно выходить и возвращаться?

Наше «я» тоже совсем не то, чем кажется.

Оправдание ожиданий

Нам кажется, что «я» — наше «я» — и есть «деятель», осуществляющий восприятие. Но это не так. «Я» — это такой же продукт восприятия, очередная контролируемая галлюцинация, хотя и совершенно особая. Все множество и разнообразие составляющих нашего «я» — начиная личностной идентичностью (я ученый или я сын) и заканчивая ощущением обладания телом, просто существованием в виде тела, — представляют собой байесовские наиболее вероятные предположения, выработанные эволюцией для выживания.

Давайте начнем исследование своего «я» с короткого путешествия в будущее. Перенесемся лет на сто вперед, когда уже будут изобретены телепорты, создающие точные копии любого человека. Они работают как аппараты в «Звездном пути»: сканируют человека в мельчайших подробностях, вплоть до устройства каждой молекулы, и на основе этих данных конструируют вторую версию человека где-нибудь далеко-далеко, например на Марсе.

Поначалу люди относятся к этой технологии с опаской, но быстро привыкают и пользуются ею как эффективным видом транспорта. Привыкают даже к такому условию, как немедленное распыление оригинала после создания копии, — эту процедуру пришлось ввести как неотъемлемую часть работы телепорта, чтобы мир не заполнили одинаковые люди. Для самого путешественника — пусть это будет женщина, назовем ее Ева, — практической проблемы нет никакой. Заверенная оператором, что все пройдет благополучно, Ева просто ощущает, что только что была в месте X (Лондон), а в следующий миг уже находится в месте Y (Марс).

Но однажды случается сбой. Распыляющий модуль в Лондоне не срабатывает, и Ева — та Ева, которая в Лондоне, — ощущает себя так, будто ничего не произошло и она по-прежнему на телепортационной станции. Небольшая заминка. Придется перезагрузить аппарат и попробовать снова, а может, отложить до завтра. Но потом в помещение входит техник с пистолетом в руке. И бормочет что-то в духе: «Не волнуйтесь, вы благополучно телепортировались на Марс, все как положено, просто в правилах сказано, что нам все равно нужно... и потом, смотрите, вы же подписали согласие...» Он медленно поднимает пистолет, и у Евы впервые закрадывается подозрение, что им насчет этой самой телепортации что-то недоговаривали.

Задача мысленного эксперимента «парадокс телепортации» — обозначить некоторые когнитивные искажения, входящие у большинства из нас в понятие «быть собой»¹.

Парадокс телепорта поднимает две философские проблемы. Первая — проблема сознания в целом: можем ли мы быть уверены, что копия будет обладать сознательным опытом, или это будет идеально функционирующий, но лишенный внутреннего мира эквивалент. Эта проблема у меня особого интереса не вызывает. Если копия создается достаточно подробной — тождественной оригиналу вплоть до устройства молекул! — нет никаких оснований сомневаться, что она будет сознательной, причем настолько же, насколько сознателен оригинал. Если копия тождественна не полностью, то мы возвращаемся к полемике о разновидностях философских зомби, и нам совершенно незачем проговаривать все это снова.

Гораздо интереснее проблема личностной идентичности или самоидентичности. Считать ли Еву на Марсе (назовем ее Ева-2) тем же человеком, что и Ева-1 (та, что в Лондоне)? Велик соблазн ответить утвердительно: Ева-2 будет чувствовать все то же самое, что чувствовала бы Ева-1, если бы ее физически переместили в один миг из Лондона на Марс. Для самоидентичности важна, судя по всему, не физическая, а психологическая непрерывность и целостность^[24]. Но тогда, если Еву-1 не распылили, какая из двух Ев настоящая?

На мой взгляд, правильный, хотя и странный ответ — обе. Они обе настоящие.

Мы интуитивно отделяем восприятие себя от восприятия мира и относимся к ним по-разному. В том, что касается ощущения «бытия собой», кажется еще труднее сопротивляться интуитивной склонности считать, будто оно отражает истинное положение вещей — в данном случае наше действительное «я», — а не набор перцепций. Убеждение в существовании действительного «я» интуитивно влечет за собой среди прочего представление о том, что такое «я» может быть только одно, а не два, не три, не множество.

Идея неделимости, неизменности, трансцендентности, своеобразия нашего «я» вписана в картезианский идеал нематериальной души и до сих пор глубоко нам, особенно в западном обществе, созвучна. В то же время ее то и дело разглядывают под скептическим микроскопом философы и духовные деятели, а в последнее время еще и психонавты, экспериментирующие с психоделическими веществами, медики и нейрочеловеки.

Кант в своей «Критике чистого разума» доказывал, что считать личность «простой субстанцией» неправильно, а Юм рассматривал личность как «пучок восприятий»². Не так давно немецкий философ Томас Метцингер написал совершенно гениальную книгу «Быть никем» (Being No One), которая представляет собой мощную деконструкцию единого «я»³. Буддисты издавна утверждают, что никакого вечного и неизменного «я» не существует, и пытаются посредством медитации абсолютно отрешиться от всего личностного. Участников церемоний с распитием отвара аяуаска, изначально проводившихся в Южной Америке, а теперь набирающих популярность и за ее пределами, лишает ощущения себя забористая смесь обрядовых действий и диметилтриптамина.

В области нейронаук Оливер Сакс и другие задокументировали множество вариантов распада личности в результате повреждения или расстройства мозга, тогда как операция по разделению мозга, о которой мы говорили в главе 3, говорит о вероятности того, что одно «я» может разделиться надвое. Самое любопытное здесь — сиамские близнецы-краниопаги, которые не просто срастаются черепами, но и имеют ряд общих структур мозга. Что значит быть отдельной личностью в мире, где один близнец-краниопаг, оказывается, ощущает, как другой пьет апельсиновый сок?⁴ «Быть собой» не так просто, как может показаться на первый взгляд.

Тем временем Ева-1 на телепортационной станции сумела ускользнуть от вознамерившегося уничтожить ее техника и пытается освоиться с этим новым для нее положением дел, а Ева-2 пребывает в блаженном неведении о разворачивающейся на Земле драме.

Если в момент копирования обе Евы были объективно и субъективно тождественны, то теперь их личности начинают расходиться. Как у гомозиготных близнецов, каждый из которых живет собственной жизнью, со временем расхождение нарастает и усложняется. Если сейчас поставить Еву-1 рядом с Евой-2, у них возникнет пусть незначительная, но разница во входящих сенсорных данных, которая приведет к едва заметным различиям в поведении, и не успеем мы оглянуться, как Евы будут испытывать не одно и то же, запоминать не одно и то же и станут разными людьми.

Эти сложности с личностной идентичностью у каждого из нас проявляются по-разному. У моей мамы самоидентичность резко изменилась во время делирия, и, хотя сейчас мама восстановилась, она кажется, по крайней мере мне, одновременно и такой, и не такой, как раньше. В точности как в случае с двумя Евами. В принципе Ева-1 и Ева-2 могут быть связаны примерно как вы нынешний и вы 10 лет назад или вы 10 лет спустя.

Когда мы говорим о том, кто вы такой или кто я такой — та личность, которую объективно и субъективно являет собой Анил Сет, — все оказывается не так просто, как на первый взгляд. Начнем с того, что ощущение личностной идентичности — то самое внутреннее «я» — это лишь одна составляющая «бытия собой» в нашем сознании.

Давайте посмотрим, что включает в себя человеческое «я».

Ощущения, непосредственно связанные с телом, — это физическое, или телесное, «я». В их число входят ощущения идентификации с определенным объектом, который является нашим организмом, — мы испытываем определенное чувство обладания телом, неприменимое к другим объектам окружающего мира. Кроме того, к телесному «я» относятся эмоции и настроения, а также состояния возбуждения и настороженности. Под этим слоем обнаруживаются более глубокие и более смутные, расплывчатые ощущения просто бытия живым организмом — бытия телесным существом, не имеющие четко очерченных пространственных границ или конкретного содержания. К этому фундаменту нашей «самости» мы вернемся чуть позже, а пока достаточно будет обозначить его как «ощущение себя живым».

Движемся дальше, отходя от телесного. Здесь у нас будет ощущение восприятия мира с определенной точки зрения, взгляд от первого лица — субъективная отправная точка перцептивного опыта, которая обычно мыслится где-то внутри головы, сразу за лбом на оси между глазами. Наилучшей иллюстрацией этого перспективистского «я» служит автопортрет австрийского физика Эрнста Маха, известный также под названием «С точки зрения левого глаза».

Центральную роль для нашего «я» играют также ощущения волеизъявления, намерения и выступления причиной происходящего — агентность. Это волевое «я». Рассуждая о «свободе воли», обычно подразумевают именно эти составляющие нашей личности. Для многих людей понятие «свобода воли» включает ту сторону «бытия собой», которую они наотрез отказываются признавать прерогативой науки.

Все эти варианты «бытия собой» могут предшествовать любому понятию личностной идентичности — связанной с именем, биографией, будущим. Как мы уже убедились на примере парадокса телепортации, для существования личностной идентичности необходима персонализированная предыстория, нанизывание автобиографических воспоминаний, хранящееся в памяти прошлое и проецируемое будущее.



Рис. 16. Эрнст Мах «Автопортрет» (1886)

Возникающее ощущение личностной идентичности можно назвать нарративным «я». Оно несет с собой способность испытывать сложные эмоции, такие как сожаление, в противовес простому разочарованию. (Человек, в отличие от животных, способен даже на «предвосхищающее сожаление» — уверенность в том, что действие, которое он сейчас выполнит, ни к чему хорошему не приведет; что он тем не менее это действие выполнит и это скажется и на нем самом, и на окружающих.) Здесь видно, как взаимодействуют и ветвятся разные уровни «я» — появление личностной идентичности меняет растущий спектр доступных эмоциональных состояний и в то же время отчасти этим спектром определяется.

Социальное «я» — это то, как я воспринимаю окружающих, которые воспринимают меня. Это часть меня, обусловленная моей включенностью, встроенностью в сеть социальных взаимоотношений. Социальное «я» вырабатывается постепенно, начиная с детства, и затем развивается на протяжении всей жизни, хотя при таких состояниях, как аутизм, этот процесс может идти не так, как у других. Развитие социального «я» влечет за собой собственный спектр эмоциональных возможностей — новых способов испытывать как неприятные чувства, скажем вину или стыд, так и приятные: гордость, любовь и ощущение сопричастности и востребованности.

Для каждого из нас в нормальных обстоятельствах эти разнообразные составляющие личности увязаны воедино, представляют одно целое и входят в объединяющее их ощущение «бытия собой». Цельность этого ощущения кажется настолько естественной — такой же естественной, как перцептивное объединение цвета и формы, когда мы смотрим на красное кресло, — что ее очень легко счесть само собой разумеющейся.

Но это будет ошибкой. Точно так же, как восприятие красного не указывает на существующую во внешнем мире «красноту», ощущение единства личности не означает существования «действительного "я"». Собственно, ощущение бытия единой, цельной личностью разрушается в два счета⁵. Чувство личностной идентичности, основанное на нарративном «я», может размыться или полностью исчезнуть при деменции или тяжелых случаях амнезии и искажаться при делирии, больничном или иного рода. Волевое «я» может давать сбой при шизофрении и при синдроме чужой руки, когда у человека ослабевает ощущение связи со своими собственными действиями, или при таком расстройстве, как акинетический мутизм, когда человек совершенно перестает взаимодействовать с окружающим миром. Перспективистское «я» страдает при чувстве «выхода из собственного тела» и других диссоциативных расстройствах; нарушения обладания телом варьируются от синдрома фантомной конечности (не проходящие, зачастую болевые ощущения от ампутированной руки или ноги) до соматопарафрении (чувства, что ваша собственная конечность принадлежит не вам, а кому-то другому). При ксеномелии — крайней форме соматопарафрении — человек испытывает острое желание отрезать себе руку или ногу и в некоторых, пусть нечастых случаях эту радикальную меру применяет.

Личность — это не неизменная сущность, украдкой поглядывающая на внешний мир через окна глаз и управляющая телом, как пилот самолетом⁶. Ощущение бытия мной или бытия вами — это тоже восприятие, точнее говоря, совокупность восприятий, туго увязанный «пучок» кодируемых нейронами предсказаний, призванных обеспечить выживание организма. И это, на мой взгляд, все, чем нам нужно быть, чтобы быть собой.

Рассмотрим ощущение отождествления с определенным объектом физического мира, который представляет собой ваше тело. Изменчивость и хрупкость этого ощущения

видны не только в таких состояниях, как соматопарафрения или синдром фантомной конечности, но выявляются и в ходе простых лабораторных экспериментов. Самый известный пример — «иллюзия резиновой руки», впервые описанная более 20 лет назад и сейчас выступающая краеугольным камнем исследований в области теории телесности⁷.

Эксперимент с иллюзией резиновой руки можно без труда провести в домашних условиях — для этого вам понадобится доброволец-испытуемый, картонка, чтобы соорудить перегородку, пара кисточек и резиновая рука. Как все расставить, показано на рис. 17. Испытуемый кладет собственную настоящую руку за картонную перегородку, вне поля своего зрения. На то место, которое эта рука должна была бы занимать в обычной ситуации, кладется резиновая искусственная рука. Затем экспериментатор берет кисточки и начинает синхронно водить ими по настоящей руке испытуемого и по искусственной. При одновременном прикосновении кисточек у испытуемого возникает странное ощущение, будто резиновая рука действительно принадлежит ему, хотя он и знает, что это не так. Если же кисточки касаются рук не одновременно, иллюзии не возникает и испытуемый не включает резиновую руку в ощущение собственного тела.

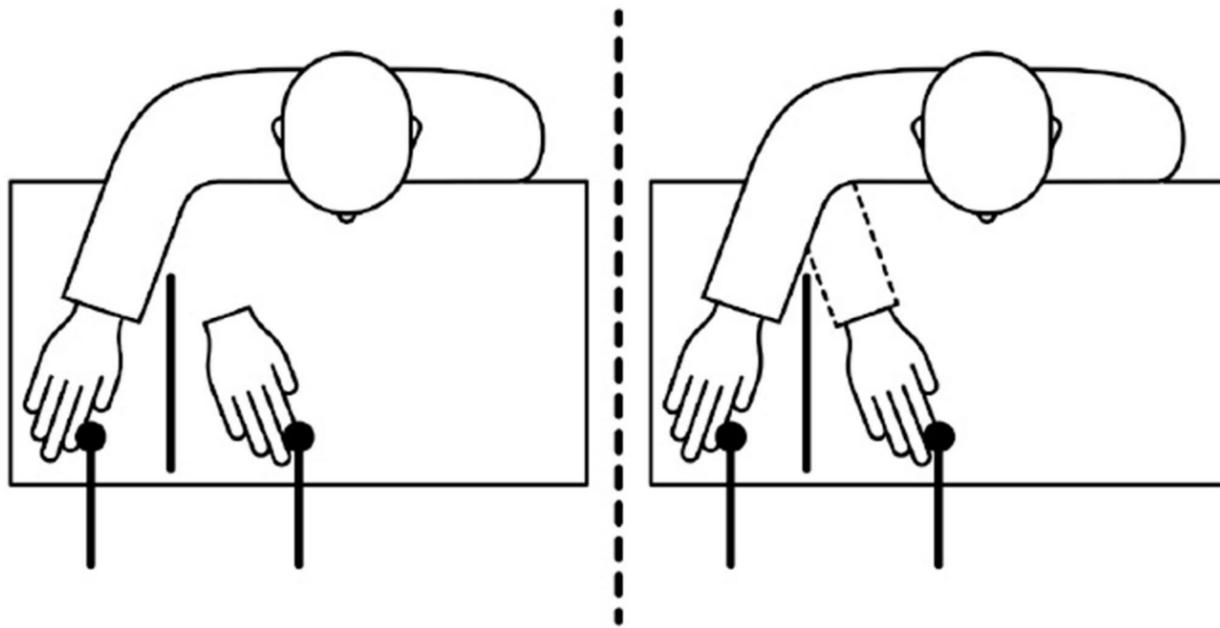


Рис. 17. Иллюзия резиновой руки. Когда резиновую и настоящую руку гладят кисточкой одновременно (слева), ощущение обладания телом размывается и резиновая рука начинает восприниматься как часть тела

У кого-то все происходит точь-в-точь как описано выше, и ощущение, что безусловно искусственная рука некоторым образом становится (хотя и не полностью) частью тела, возникает хотя и странно, но отчетливо. Однако в действительности у кого-то иллюзия может выражаться сильнее, у кого-то слабее. Один из способов это выяснить — внезапно замахнуть на резиновую руку молотком или ножом: если иллюзия действительно сработала, вы получите немедленное живейшее тому подтверждение.

Иллюзия резиновой руки отлично согласуется с идеей, что ощущение обладания телом — это особая разновидность контролируемой галлюцинации. Иллюзия обусловлена тем, что при синхронном поглаживании кисточкой сочетание наблюдения за касанием резиновой руки и одновременное ощущение (без наблюдения) касания, испытываемое настоящей рукой, обеспечивает мозгу достаточно сенсорных данных, чтобы вывести наиболее вероятное перцептивное предположение: резиновая рука некоторым образом принадлежит телу. Происходит это только при синхронном (но не асинхронном) поглаживании в силу априорного ожидания, что поступающие одновременно сенсорные сигналы должны, скорее всего, иметь общий источник — и в данном случае им оказывается резиновая рука. Иначе могут ощущаться не только части тела. Иногда смещается ощущение всего тела — и взгляда «от первого лица».

В 2007 г. в престижном научном журнале Science почти одновременно вышли две статьи. В обеих описывалось, как с применением новых методов использования технологии виртуальной реальности вызывать чувство «выхода из собственного тела». В основе этих экспериментов лежала иллюзия резиновой руки, но теперь она распространялась на все тело. В одном из исследований, проведенных лозаннской научной группой под руководством Олафа Бланке, испытуемые надевали шлем виртуальной реальности и наблюдали виртуальное отображение собственной спины с расстояния в два метра (см. рис. 18). Затем они с этой же точки наблюдали, как виртуальную спину гладят кисточкой — либо синхронно, либо асинхронно касанию кисточкой их собственной спины. При синхронном поглаживании большинство участников, по их словам, ощущали виртуальное тело в какой-то мере своим, и, когда их просили подойти туда, где они сейчас, по собственным ощущениям, располагаются, они делали движение, устремленное к виртуальному телу⁸.

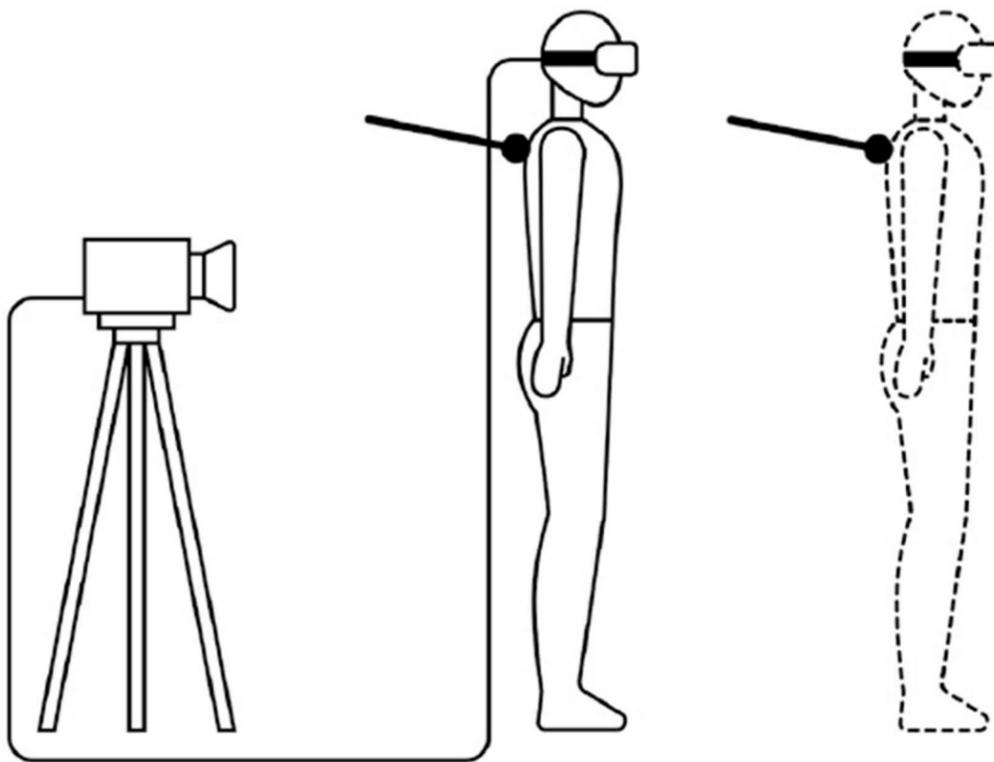


Рис. 18. Создание иллюзии «всего тела»

Точно так же, как иллюзия резиновой руки указывает на гибкость и изменчивость ежесекундного чувства обладания телом, такие эксперименты, называемые «иллюзиями всего тела», позволяют предположить, что субъективным чувством обладания телом и местонахождением взгляда от первого лица тоже можно без труда манипулировать. Эти эксперименты служат ошеломляющим свидетельством того, что ощущение «чем является мое тело» можно отделить, по крайней мере в некоторой степени, от ощущения «где я».

Идея возможности покидания физического тела «смотрящим от первого лица» — в виде «выхода из собственного тела», или внетелесного переживания (ВТП), — глубоко укоренена в нашей истории и культуре⁹. Рассказы о внетелесных или схожих с ними переживаниях, когда человек оказывается на грани смерти, на операционном столе, при наступлении или окончании эпилептического приступа, подпитывают веру в существование нематериальной сущности, сосредоточивающей в себе наше «я». Еще бы: раз мы можем посмотреть на себя со стороны, значит, основа нашего сознания отделима от мозга?

Однако, если расценивать взгляд от первого лица как одну из разновидностей перцептивного умозаключения, необходимость хвататься за эти дуалистические подпорки отпадает. В пользу этой концепции говорят не только результаты экспериментов Олафа Бланке и других с виртуальной реальностью, но и исследования с использованием стимуляции мозга, история которых уходит к серии прорывных экспериментов, проведенных в 1940-х гг. канадским неврологом Уайлдером Пенфилдом.

Одна из пациенток Пенфилда, фигурировавшая в записях под инициалами G. A., в ходе электрической стимуляции правой верхней височной извилины (часть височной доли мозга) воскликнула вдруг: «У меня странное чувство, будто я не здесь... Я словно наполовину тут, а наполовину нет»¹⁰. Собственно, и сам Бланке заинтересовался ВТП, когда о схожих ощущениях («Я вижу самого себя в кровати, я смотрю сверху, но вижу только свои ноги»¹¹) сообщил его пациент при стимуляции аналогичной области мозга — угловой извилины, находящейся на стыке височной и теменной долей.

Все подобные случаи объединяет необычная активность в тех областях мозга, которые имеют дело с вестибулярными входящими данными (вестибулярная система отвечает за ощущение равновесия) и при этом участвуют в интеграции данных от всех органов чувств¹². Судя по всему, при нарушении нормальной деятельности этих систем мозг может выдать необычное «наиболее вероятное предположение» о нахождении своего взгляда от первого лица, даже если остальные составляющие нашего «я» останутся незатронутыми.

Связь с нарушениями этих процессов прослеживается и у схожих с ВТП ощущений, которыми иногда сопровождаются эпилептические приступы. Эти ощущения обычно подразделяют на аутоскопические галлюцинации, при которых человек видит окружающую его обстановку под другим углом, и геатоскопические (также называемые двойниковыми), при которых человек видит под другим углом самого себя. Обширный массив документальных свидетельств подобных ощущений, накапливающихся уже не одну сотню лет, тоже подтверждает пластичность¹³ взгляда от первого лица^[25].

Не стоит отмахиваться от сообщений о сверхъестественных, казалось бы, или абсурдных ощущениях вроде ВТП как от выдумки. Люди, скорее всего, действительно испытывают именно то, о чем сообщают. Внетелесные переживания случаются у людей не одно тысячелетие, но это не значит, что физическое тело и вправду покидает некая нематериальная сущность или неизменная вечная душа. Эти рассказы говорят лишь о том, что взгляд от первого лица складывается куда более сложным, непостоянным и нестабильным образом, чем нам способен подсказать непосредственный субъективный доступ.

Возможность менять в виртуальном мире взгляд от первого лица находит самое ошеломляющее применение, во многих случаях строящееся на иллюзии с интригующим названием «обмен телами», описанной в исследовании 2008 г. под руководством шведского ученого Хенрика Эрссона¹⁴. В эксперименте с обменом телами два человека надевают шлемы виртуальной реальности с камерой. Переключая подачу потока данных с камеры своего шлема на шлем соседа, каждый испытуемый может увидеть себя глазами другого. Но во всей своей полноте эффект «обмена телами» ощущается, только когда участникижимают руки: одновременно наблюдая и ощущая это рукопожатие, они получают мультисенсорную стимуляцию, при которой в сочетании с нисходящими ожиданиями каждый из них ощущает, будто каким-то чудом переместился в чужое тело и пожимает руку самому себе. Это переживание позволяет, пусть виртуально, побывать в чужой шкуре.

Я пробовал виртуальный обмен телами на небольшой встрече в калифорнийском городе Охай зимой 2018 г. Мы были там с Даанишем Масудом, посредником ООН в мирных урегулированиях, параллельно трудящимся в области исследований виртуальной реальности. К тому времени Масуд уже несколько лет тесно сотрудничал с лабораторией BeAnotherLab — детищем барселонского нейрочеловека Мела Слейтера. Лаборатория занималась адаптацией технологии обмена телами для новейших устройств «эмпатического поколения»¹⁵. Авторы проекта исходили из

того, что, побывав «в чужой шкуре», то есть увидев мир чужими виртуальными глазами, человек естественным образом начнет проявлять эмпатию и в других ситуациях.

Дааниш привез свою группу в Охай, чтобы продемонстрировать созданную ими систему под названием «Машина перевоплощения» (The Machine to Be Another). Это устройство добавляет к базовому принципу обмена телами немного изящной хореографии, благодаря которой эффект оказывается еще ярче. Надев шлем, каждый из пары участников сперва смотрит на свои колени — и видит при этом колени партнера. Затем, следуя подробным инструкциям, участники совершают ряд координированных движений, и, если при этом им удастся максимально синхронизироваться, возникает впечатление, что они способны управлять своими новыми телами, и чувство перевоплощения усиливается. Некоторое время спустя участникам показывают зеркало, и каждый принимает отражение другого за свое. В последнем акте разделяющий двух участников занавес убирается, и они смотрят на себя глазами другого, а потом подходят друг к другу и обнимаются.

Когда настала моя очередь попробовать, моей партнершей по обмену оказалась довольно ухоженная дама в возрасте за 70. Я даже не ожидал, что этот эксперимент настолько меня захватит. Помню, как я опустил взгляд, согнул свою (ее) руку и с изумлением уставился на свою (ее) обувь — мерцающие блестками кроссовки. Особенно сильное впечатление на меня произвели зеркало и финальные объятия — не знаю даже, что повлияло больше, ощущение нахождения в чужом теле или возможность видеть себя чужими глазами. И только позже, за ужином, я задумался, каково же было моей партнерше вдруг начать смотреть на мир глазами англо-индийского нейробиолога в скучнейших ботинках.

Мне кажется поразительным, что этими привычными и принимаемыми как данность составляющими «я» — субъективным обладанием телом и взглядом от первого лица — так легко удается манипулировать, будь то с помощью искусственной руки и кисточки или новейших технологий виртуальной и дополненной реальности. Однако у этих манипуляций есть предел. Как я уже говорил выше, обычно иллюзия резиновой руки выражается в том, что рука начинает ощущаться как собственная, хотя человек прекрасно осознает, что это не так. Но это «обычное» ощущение у разных людей выражается по-разному, а кто-то и вовсе ничего похожего не чувствует. То же самое относится к иллюзиям всего тела и иллюзиям обмена телами.

В этом смысле экспериментальные манипуляции принадлежностью тела сильно отличаются от классических зрительных иллюзий, таких как шахматная доска Адельсона, которую мы видели в главе 4. В случае с шахматной доской мы убеждены перцептивно, что квадраты отличаются оттенком серого, и сильно удивляемся или даже изумляемся, когда оказывается, что оттенок у них один. Такого удивления, характерного для зрительных иллюзий, при иллюзиях обладания телом не возникает почти никогда. Для меня самой убедительной телесной иллюзией из всех испробованных был обмен телами в Охае, но и там я ни на миг не поверил, что стал кем-то другим или куда-то переместился.

Субъективную слабость иллюзий обладания телом подчеркивает недавнее исследование, посвященное роли гипнотической внушаемости в иллюзии резиновой руки¹⁶. В нем принял участие и я. Исследование, проводимое психологами Питером Лашем и Золтаном Дьенешем, строилось на предположении, что сама модель эксперимента, обеспечивающая иллюзию, вызывает у испытуемого сильное подспудное ожидание по поводу того, что ему предстоит ощутить, и у некоторых уже одно это ожидание меняет ощущение обладания телом. В подтверждение этой гипотезы мы обнаружили, что индивидуальные различия в силе иллюзии

коррелируют с тем, насколько человек внушаем, если измерять по стандартной шкале гипнабельности. Отличающиеся высокой гипнотической восприимчивостью сообщали о сильном ощущении принадлежности искусственной руки их телу (при синхронном поглаживании кисточками), тогда как продемонстрировавшие низкие показатели гипнабельности почти не ощутили иллюзию.

С одной стороны, эти результаты отлично согласуются с представлениями об обладании телом как о контролируемой галлюцинации, поскольку гипнотическое внушение можно рассматривать как сильное нисходящее ожидание, пусть даже такое, о котором сам участник не подозревает¹⁷. С другой стороны, это серьезная проблема для экспериментальных исследований в этой области, поскольку возникает вероятность, что иллюзия резиновой руки большей частью или целиком вызвана эффектом внушения. И если исследования иллюзий воплощения не будут принимать в расчет разницу во внушаемости (а они в основном не принимают), им трудно будет сказать что-то конкретное по поводу механизмов создания иллюзии. Это относится и к резиновой руке, и к внетелесным переживаниям, и к иллюзиям обмена телами, и любым другим ситуациям, в которых человека явно или подспудно подводят к ожиданию того или иного ощущения, связанного с телом.

Эти субъективно безобидные иллюзии обладания телом резко контрастируют с сильно измененными переживаниями, наблюдаемыми при таких клинических состояниях, как соматопарафрения, ксеномелия и синдром фантомной конечности, или при ярких внетелесных переживаниях, связанных с припадками или вызываемых прямой стимуляцией мозга. Настолько резкие искажения гораздо больше схожи с классическими зрительными иллюзиями, прежде всего большей убедительностью для тех, кто их испытывает. И поэтому они намного более явно свидетельствуют о том, что ощущение обладания телом и взгляд от первого лица — это действительно только конструкторы, выстраиваемые мозгом.

Давайте от вопросов личностной идентичности перейдем к появлению «нарративного» и «социального» «я». Как мы видели на примере парадокса телепортации, именно на этих уровнях наша сущность воспринимается как непрерывно перетекающая из одного мига в другой, длящаяся изо дня в день, из недели в неделю, из месяца в месяц и до некоторой степени на протяжении всей жизни. Это те уровни «я», на которых имеет смысл увязывать его с именем, воспоминаниями и планами на будущее. На этих уровнях мы осознаем, что «я» у нас имеется, то есть приобретаем подлинное самосознание¹⁸.

Эти более высокие планы личности полностью отделимы от физического, телесного «я». Телесное «я» способны ощущать многие животные, а также маленькие дети, не имея при этом никакого сопровождающего его чувства личностной идентичности. И хотя взрослые люди обычно ощущают все эти формы личности как единое интегрированное целое, когда нарративная и социальная составляющие слабеют или разрушаются, результат может быть катастрофическим.

Клайв Уэринг — британский музыковед, известный как редактор произведений композитора эпохи Возрождения Орландо ди Лассо, хормейстер и человек, полностью изменивший музыкальный формат «Радио-3» корпорации ВВС в начале 1980-х гг. В марте 1985 г., находясь на пике карьеры, он перенес опасное воспалительное заболевание мозга (герпетический энцефалит), сильно повредившее его гиппокамп в обоих полушариях и вызвавшее один из самых тяжелых случаев амнезии в истории медицины^[26].

Уэрингу стало невероятно трудно извлекать уже имеющиеся воспоминания (это расстройство называется ретроградной амнезией), но в первую очередь у него возникли затруднения с накоплением новых (антероградная амнезия). Он словно

существует исключительно в текущем моменте — в отрезке времени длиной от семи до 30 секунд. Сейчас ему за 80, и, скорее всего, он по-прежнему воспринимает жизнь как непрерывную череду мини-пробуждений, как будто примерно каждые 20 секунд он приходит в себя после комы или общего наркоза. Его нарративное «я» уничтожено.

Вид памяти, который утратил Клайв, — это память эпизодическая, автобиографическая, то есть память о локализованных во времени и пространстве событиях, включающая, самое важное, события с его собственным участием. Читать дневники Уэринга больно¹⁹. Там раз за разом повторяются упоминания о «первом» пробуждении, при этом предшествующие записи, сделанные иногда всего несколькими минутами раньше, вычеркиваются, а иногда и замазываются в ярости.

8:31 Вот теперь я целиком и полностью очнулся.

9:06 Теперь я абсолютно очнулся и всецело это ощущаю.

9:34 Теперь я, вне всякого сомнения, по-настоящему очнулся.

Эти дневники и беседы с Клайвом, которые его жена Дебора приводит в своей книге «Вечное сегодня» (Forever Today), говорят об ударе по его ощущению личностной идентичности, вызванном поражением мозга. Его неспособность выстраивать единый нарратив на протяжении времени означает, что «бытие собой», более 30 лет представлявшее для Клайва непрерывный поток, теперь обнулилось, превратившись в мимолетное ощущение, лишённое стабильного «я» — того стержня, на который можно наматывать нить восприятия мира и себя. Утрата прошлого и будущего настолько выбила Уэринга, увязшего в зыбучих песках амнезии, из колеи, что он начал сомневаться, жив ли он и был ли жив. Дебора Уэринг пишет: «Клайву постоянно казалось, будто он только что очнулся от потери сознания, поскольку никаких свидетельств предшествовавшего бодрствования у него не было. "Я ведь ничего не слышал, ничего не видел, ничего не касался, мне ничем не пахло, — говорил он. — Это все равно что умереть"»²⁰.

При этом другие составляющие ощущения себя у Клайва совершенно не пострадали. У него не нарушено чувство обладания телом, местонахождение взгляда от первого лица и даже совершение волевых действий. Его любовь к жене ничуть не ослабла, хотя иногда он не может вспомнить, как он с ней познакомился, — они поженились всего за год до его болезни. И когда Клайв играет на фортепиано, или поет, или дирижирует, музыка льется так же свободно и живо, как прежде, позволяя ему снова почувствовать себя цельным.

Клайва эти моменты любви и музыки преображают и спасают. Оливер Сакс в статье в «Нью-Йоркере» описал его случай так: «У него больше нет внутреннего нарратива, он не проживает жизнь в том смысле, в котором проживаем ее мы. Но любой, кто видит его за фортепиано или с Деборой, моментально убеждается, что в такие минуты он снова становится собой и ощущает всю полноту жизни»²¹.

Несмотря на эти светлые моменты, случай Клайва безусловно трагичен. Разрушение нарративного «я» — это не просто ухудшение памяти, оно ведет к неспособности воспринимать себя как непрерывно существующего во времени, а эта неспособность, в свою очередь, размывает основополагающее чувство личностной идентичности, которое большинство из нас вполне естественно считает само собой разумеющимся. Памятью наше «я» не исчерпывается, но, как показывает история Уэринга и как многие из нас видят на примере своих друзей и родных, которых засасывает болото деменции или болезни Альцгеймера, без непрерывности и устойчивости самовосприятия человеку обойтись нелегко.

Сила любви Клайва и Деборы, способная восстанавливать ощущение личностной идентичности Клайва, плавно подводит нас к разговору о социальном «я».

Человек, как и многие другие животные, существо общественное. Способность понимать состояние чужого сознания — критически важное умение для общественных живых существ в любых контекстах и в любых сообществах. Часто считается, что эта способность — ее еще называют «теорией разума» — у человека развивается довольно медленно, однако она играет ключевую роль почти для всех нас на протяжении всей нашей жизни.

Временами мы ощущаем ее довольно остро — например, когда пытаемся угадать, что думает о нас партнер, друг, коллега. Но даже когда мы не раздумываем осознанно над своими социальными взаимоотношениями, наша способность воспринимать чужие намерения, убеждения и желания всегда работает в фоновом режиме, направляя наше поведение и формируя эмоции.

О социальном восприятии и теории разума написано немало, и эти работы охватывают и психологию, и социологию, и (с недавнего времени) еще совсем молодую область социальной нейронауки²². В большинстве случаев названные вопросы исследуются с точки зрения их важности для руководства социальными взаимоотношениями. Я же хочу развернуть объектив в другую сторону и посмотреть, насколько наше ощущение «бытия собой» зависит от того, как я воспринимаю восприятие себя другими.

Социальное восприятие — восприятие состояния психики окружающих — это не просто эксплицитное рассуждение или «раздумья» о том, что думают или не думают другие. В большинстве своем наше социальное восприятие носит автоматический и непосредственный характер. Мы формируем восприятие чужих убеждений, эмоций и намерений так же естественно и легко, как формируем восприятие кошек, кофейных чашек, кресел и даже собственного тела. Когда я, наливая себе бокал вина, вижу, что подруга подставляет мне свой пустой бокал, мне не нужно вычислять ее намерение логически, я просто воспринимаю, что ей тоже хочется еще вина и что, вообще-то, нужно было бы сначала предложить налить ей. Я воспринимаю эти состояния психики с такой же легкостью, хотя и не обязательно так же правильно и точно, как сам винный бокал.

Как это происходит? Ответ, на мой взгляд, снова нужно искать в концепции мозга как прогностической машины и в восприятии как процессе вывода источников сенсорных сигналов.

И несоциальное, и социальное восприятие подразумевает построение мозгом наиболее вероятных предположений относительно источников входящей сенсорной информации. Но если ошибок в восприятии того, что происходит в уме другого человека, у нас, как известно, хватает, то ошибиться в восприятии винного бокала и перепутать его, допустим, с машиной мы вряд ли сможем (если у нас не галлюцинации). Одна из причин заведомой неоднозначности социального восприятия состоит в том, что источники входящих данных спрятаны гораздо глубже. Световые волны, вызывающие восприятие винного бокала, проистекают более или менее непосредственно от самого бокала, тогда как сенсорные сигналы, связанные с состоянием психики другого человека, неизбежно проходят ряд промежуточных этапов — мимику, жесты, речевые акты, — и на каждом из них повышается вероятность перцептивной промашки.

Социальное восприятие точно так же, как зрительное, зависит от контекста и ожиданий, и мы пытаемся минимизировать «ошибки социального прогнозирования», меняя сенсорные данные — это межличностная разновидность активного вывода, — а также обновляя прогнозы. Активный вывод в социальном восприятии выражается в поведении, которое изменит состояние психики другого человека, приведя его в

соответствие с нашим прогнозом или желанием²³. Так, например, мы улыбаемся не только для того, чтобы выразить собственную радость, но и чтобы изменить чувства нашего собеседника, а говоря что-то, пытаемся внедрить те или иные мысли в его сознание.

Теперь посмотрим, как эти представления о социальном восприятии можно увязать с социальным «я». Для способности делать выводы о психическом состоянии окружающих, как и для перцептивного умозаключения, требуется порождающая модель. Порождающие модели, как мы знаем, могут порождать сенсорные сигналы, соответствующие той или иной перцептивной гипотезе. Применительно к социальному восприятию это означает предположение о психическом состоянии другого человека, что подразумевает высокую степень взаимности. Моя наиболее вероятная модель вашего психического состояния будет включать модель того, как вы моделируете мое психическое состояние. Иначе говоря, я пойму, что у вас на уме, только если попытаюсь понять, как вы воспринимаете содержание моего сознания. Именно в этом смысле мы воспринимаем себя через призму сознания других. В этом и заключается социальное «я», и эта социальная «матрешка» прогностического восприятия составляет важную часть общего ощущения бытия человеком²⁴.

Такая трактовка социального «я» подразумевает, кроме прочего, интригующее предположение, что самосознание — высший уровень личности, включающий в себя и нарративное и социальное «я», — непременно требует социального контекста. Если вы существуете в мире, в котором других сознаний нет (точнее, нет других подходящих сознаний), вашему мозгу не нужно будет предсказывать психическое состояние окружающих, а значит, не нужно будет делать вывод о том, что его собственные ощущения и действия принадлежат какой-то личности. Так что рассуждение поэта XVII в. Джона Донна о том, что «человек не остров», можно воспринимать буквально²⁵.

Вы сегодня тот же, кем были вчера? Переформулирую получше: вы ощущаете себя сегодня таким же, каким ощущали вчера? Скорее всего — если, конечно, с вами за ночь не произошло чего-то катастрофического, — ваш ответ будет утвердительным. А по сравнению с прошлой неделей, прошлым месяцем, прошлым годом, с собой десятилетней давности, с собой четырехлетним? А когда вам будет 94, как думаете, вы будете все тем же? Вы будете казаться себе таким же?

Поразительное, но часто упускаемое из виду свойство сознательного «я» состоит в том, что обычно мы воспринимаем себя как существующих во времени непрерывно, как единое целое. Это свойство можно назвать субъективной стабильностью личности. Оно относится не только к непрерывности автобиографической памяти, но и к более глубокому ощущению себя как сохраняющегося из мига в миг, будь то на уровне биологического тела или личностной идентичности²⁶.

По сравнению с перцептивным восприятием внешнего мира восприятие, связанное с нами самими, невероятно стабильно. Наше восприятие мира постоянно меняется, объекты и сцены появляются и исчезают в непрерывном потоке событий.

Восприятие, связанное с собой, судя по всему, не так изменчиво. И хотя мы знаем, что меняемся со временем — фотодоказательств этого у большинства имеется в избытке, — нам эти перемены кажутся не особенно существенными. Если мы не страдаем никаким психическим или неврологическим заболеванием, бытие собой ощущается как незыблемый остров посреди постоянно меняющегося мира. Об этом говорил еще в XIX в. один из основоположников психологии Уильям Джеймс: «В отличие от восприятия объекта, который может восприниматься с разных точек зрения или даже перестать восприниматься, ощущение, что мы по-прежнему находимся в привычном старом теле, пребудет с нами всегда»²⁷.

Конечно, вы можете решить, что здесь и рассматривать-то нечего. В конце концов, тело — и прочие объекты восприятия, связанного с нами самими, — действительно меняется меньше, чем все воспринимаемое нами во внешнем мире. Сколько бы я ни переходил из одной комнаты в другую, мое тело, мои действия и мой взгляд от первого лица будут неизменно меня сопровождать. При таком раскладе неудивительно, что наше «я» воспринимается как меняющееся меньше, чем мир вокруг. Но я думаю, дело не только в этом.

Как мы видели в главе 6, восприятие изменений — это тоже перцептивное умозаключение. Даже когда наше восприятие меняется, мы совсем не обязательно воспринимаем его как меняющееся. Эту разницу иллюстрирует феномен «слепоты к изменениям», при котором медленные перемены (в окружающем мире) не вызывают соответствующего ощущения. Тот же принцип применим и к восприятию себя. Мы постоянно становимся другими. Наше восприятие себя постоянно меняется: сейчас вы уже не тот человек, каким начинали читать эту главу, но это не значит, что мы эти перемены воспринимаем.

Субъективная слепота к изменениям нашего «я» не обходится без последствий. Прежде всего, она подпитывает ложное интуитивное представление о том, что «я» — это незыблемая монументальная сущность, а не «пучок перцепций». Но эволюция выработала у нас именно такое ощущение самих себя не поэтому. Я думаю, что субъективная стабильность «я» — это нечто большее, чем просто слепота к изменениям, продиктованная медленным темпом перемен в нашем теле и мозге. Мы живем с преувеличенной, крайней формой слепоты к изменениям в себе, и, чтобы понять почему, нам нужно разобраться, почему мы в принципе воспринимаем себя. Мы делаем это не для того, чтоб познать, а для того, чтобы себя контролировать.

Глава 9

Теория животного-машины

Мы видим вещи не такими, какие они есть, а такими, какие мы есть¹.

Анаис Нин

Суть самовосприятия не в выяснении, что происходит снаружи, в окружающем мире, или внутри нашего тела. Самовосприятие — это физиологический контроль и регуляция, призванные обеспечить нам выживание. Чтобы разобраться, почему это так и что это значит для всего нашего сознательного опыта, давайте для начала вспомним один старый спор о том, как связаны жизнь и разум.

На вершине Великой цепи бытия² — средневековой христианской иерархии всего сущего и живого — располагается Бог. Прямо под ним — ангелы, затем люди (сгруппированные в соответствии с социальной принадлежностью), затем другие животные, растения и, наконец, минералы. Все на своих местах, у каждого свои способности и силы, определенные его местом в Цепи.

Внутри Цепи мы, люди, неловко балансируем между миром горним, обителью Бога и ангелов, и миром дольным, физическим, населенным животными, растениями и минералами. Мы обладаем бессмертной душой, у нас есть разум, любовь и воображение, но при этом нас одолевают физические страсти — боль, голод, плотское желание, — поскольку мы привязаны к своему физическому телу.

Столетиями Великая цепь бытия (или *Scala Naturae*, «лестница природы») служила, особенно в Европе, устойчивым шаблоном, по которому человек определял свое место в мире, а также свою ценность по сравнению с другими людьми (короли в Цепи располагаются выше крестьян). Затем, в XVII в., Рене Декарт упразднил ряд ступеней этой лестницы, разделив мир на две модели существования —

res cogitans («вещь мыслящую», разум) и res extensa («вещь протяженную», материю).

Это резкое упрощение пестрой картины мира, оставлявшее ей всего две краски, повлекло за собой множество новых проблем. В том числе и метафизическую проблему того, как эти две категории должны взаимодействовать, и с тех пор этот вопрос так или иначе, к добру или к худу (чаще к худу), задавал рамки исследований сознания. Кроме того, нарушился и тот филигранно выстроенный порядок, от которого зависела политическая и религиозная власть. Если у животных имеются элементы res cogitans, любые признаки разума, что помешает им точно так же, как людям, притязать на духовное? И любые попытки копаться в душе рассудочными инструментами, как вроде бы предлагал Декарт, несомненно, раздражали могущественную католическую церковь³.

Декарт всегда осторожничал с церковью, вплоть до того, что в третьей и пятой главах своих «Размышлений о первой философии» пытался доказать существование всемилостивого Бога⁴. Что же до животных помимо человека, часто утверждается, будто Декарт полностью отказывал им в наличии сознания. Но, хотя наверняка сказать мы не можем, не факт, что он действительно так считал^[27]. Основной постулат Декарта, касающийся животных, состоял в том, что у них нет души, а значит, и всего разумного, духовного и сознательного, что к ней прилагается. Историк Уоллес Шагг резюмировал его взгляды на этот вопрос так:

Тело как человека, так и зверя <...> не более чем машина, которая дышит, переваривает, воспринимает и движется за счет определенного устройства ее частей. Но только у человека движения тела управляются разумом так, что он готов к любым случайностям; только человек выказывает признаки наличия у него разума, владея подлинной речью. Без разума, управляющего движениями тела или восприятием чувств, животных следует считать бездумными, бесчувственными машинами, которые движутся как часовой механизм⁵.

С этой точки зрения свойства живых существ, связанные с плотью и кровью, то есть их природа как живых организмов, не имеют равным счетом никакого отношения к наличию разума, сознания или души (чем бы она ни была). Животных (помимо человека) лучше всего считать bêtes-machines — то есть животными-машинами. В картезианской картине мира разум и жизнь так же отделены друг от друга, как res cogitans и res extensa.

Подчеркивая особенное положение человека, Декарт задабривал своих потенциальных преследователей, и ему это удалось, однако философия в результате вступила на скользкий путь. Если животные — это машины, а человек тоже животное (как-никак он явно создан из той же плоти, крови, жил и костей), значит, способности разума и сознания тоже должны объясняться в механистических и физиологических терминах?

Французский философ Жюльен Офре де Ламетри, писавший в середине XVIII в., именно так все и представлял. Он распространил довод Декарта о животных-машинах на человека, доказывая, что человек тоже машина — l'homme machine⁶, отрицая тем самым особое нематериальное положение души и ставя под вопрос существование Бога. Ламетри был не из тех, кто станет смягчать доводы в угоду религиозным властям, поэтому жизнь его, в отличие от Декарта, очень быстро осложнилась. В 1748 г. он был вынужден бежать из приютивших его Нидерландов в Берлин, ко двору прусского короля Фридриха, где три года спустя философ скончался, пережив паштета.

Если Декарт считал разум и жизнь независимыми друг от друга, то Ламетри усматривал между ними тесную связь, полагая, что разум можно расценивать как

свойство жизни. Полемика о том, существует ли расхождение между жизнью и разумом в основополагающих механизмах и принципах, не утихает по сей день⁷.

Мои симпатии в этой полемике на стороне Ламетри, но я предпочитаю говорить не о «разуме» в общем, а именно о сознании. И тут мы подходим к сути моей теории животного-машины о сознании и личности. Наше осознанное восприятие окружающего мира и себя в нем складывается при участии, посредством и благодаря нашему живому организму. Наше животное устройство не просто согласуется с осознанным восприятием себя и мира. Моя гипотеза состоит в том, что понять свойства и происхождение этих осознанных переживаний мы можем не иначе как в свете нашей природы живых существ.

Под многоуровневыми пластами личности, включающими воспоминания о прошлом и планы на будущее, перед эксплицитным чувством личностной идентичности, ниже уровня «я» и даже прежде появления взгляда от первого лица и ощущения обладания телом располагаются более глубокие слои, которые нам еще только предстоит открыть. Они тесно связаны не столько с телом как с объектом окружающего мира, сколько с его внутренним устройством, и варьируют от эмоций и настроений (того, что психологи называют «аффективными» переживаниями) до базального, бесформенного непреходящего ощущения «бытия» телесным, живым организмом.

Мы начнем изыскания в этих глубинах с эмоций и настроений. Эти формы содержания сознания играют центральную роль в чувственном опыте телесного «я», и, как любое восприятие, их можно рассматривать как байесовское наиболее вероятное предположение об источниках сенсорных сигналов. Отличие аффективных переживаний в том, что соответствующие источники обнаруживаются внутри тела, а не во внешнем мире.

Размышляя о восприятии, мы обычно думаем о нем как о разных способах ощущения окружающего мира, в частности о привычных модальностях зрения, слуха, вкуса, осязания и обоняния. Эти ориентированные во внешний мир разновидности ощущения и восприятия называются экстероцепцией. Восприятие того, что происходит внутри тела, именуется интероцепцией — это «ощущение внутреннего физиологического состояния тела»^{[28]8}. Интероцептивные сенсорные сигналы обычно поступают от внутренних органов тела в центральную нервную систему, передавая сведения о состоянии этих органов, а также о функционировании организма в целом⁹. Интероцептивные сигналы сообщают о сердцебиении, об артериальном и венозном давлении, о различных низкоуровневых показателях химии крови, натяжении стенок желудка, дыхании и так далее. Эти сигналы, пройдя сквозь сложную сеть нервов и глубокие отделы мозга в стволе и таламусе, принимаются теми частями коры, которые отвечают за обработку интероцептивных данных, в частности островковой корой^{[29]10}. Ключевое свойство интероцептивных сигналов состоит в том, что они так или иначе отражают качество физиологической регуляции организма. То есть показывают, хорошо ли мозг обеспечивает выживание тела.

С эмоциями и настроениями интероцептивные сигналы стали увязывать довольно давно. Еще в 1884 г. Уильям Джеймс и Карл Ланге независимо друг от друга доказывали, что эмоции — это не «вечные и священные психические сущности», какими их рисовали философы древности, и не механизмы, возникшие в мозге в процессе эволюции, как утверждал Дарвин. Эмоции, утверждали они, — это восприятие изменений в состоянии организма. Мы плачем не потому, что нам грустно, — наоборот, нам грустно, потому что мы воспринимаем состояние организма как плачущее. Эмоцию страха, согласно этой концепции, составляет (интероцептивное) восприятие целого диапазона телесных откликов, выдаваемых организмом, который распознает опасность в окружающей среде. С точки зрения

Джеймса, восприятие изменений в организме по мере их возникновения и есть эмоция: «Мы огорчаемся, потому что плачем, сердимся, потому что наносим удар, боимся, потому что дрожим, а не наоборот, — плачем, наносим удар и дрожим, потому что огорчаемся, сердимся или боимся»¹¹.

В свое время теория Джеймса встретила суровый отпор, отчасти потому, что она переворачивала общепринятое интуитивное, соответствующее тому, как все выглядит, представление о том, что эмоции вызывают телесный отклик, а не наоборот. Как же иначе, это ведь испуг (например, если на вас неожиданно выйдет медведь гризли) заставляет сердце бешено стучать, уровень адреналина в крови — подскочить, а ноги — мчаться. Но мы уже научены скептически относиться к внешнему впечатлению и не руководствоваться им в попытке понять, как все обстоит на самом деле, поэтому отвергать гипотезу Джеймса только на этом основании было бы недальновидно.

Гораздо более существенное возражение состоит в том, что разница между телесными состояниями может быть не настолько велика, чтобы поддерживать весь спектр эмоций, который мы испытываем. И хотя по поводу отдельных составляющих этого возражения еще возникают разногласия¹², в 1960-х гг. появился вполне весомый ответ в виде «оценочных теорий» эмоций¹³. Согласно этим теориям эмоции — не просто считывание изменений в телесном состоянии. Они зависят от более высокоуровневой когнитивной оценки контекста, в котором происходят физиологические изменения.

Оценочные теории снимают проблему эмоционального диапазона, поскольку эта концепция не требует отдельного телесного состояния для каждой конкретной эмоции. Две родственные эмоции — апатия и безразличие, например, — могут основываться на одном и том же телесном состоянии, а различие между ними будет объясняться разной когнитивной его интерпретацией. Разумеется, с таким же успехом может оказаться — и я подозреваю, именно так и оказывается, — что у каждой эмоции действительно имеется своя «фирменная» телесная характеристика, просто нюансы этих отличительных особенностей очень трудно распознать.

Мое любимое экспериментальное исследование оценочной теории было выполнено в рамках остроумного проекта Дональда Даттона и Артура Арона в 1974 г.¹⁴ В этом исследовании женщина-интервьюер подходила к прохожим мужского пола на одном из двух мостов через реку Капилано в Норт-Ванкувере. Первый — шаткий подвесной мост длиной почти в 140 метров, с низким сетчатым ограждением, качающийся высоко над быстрой и мелкой горной речкой. Второй — более короткий и надежный, срубленный из толстых кедровых бревен и расположенный выше по течению, всего в трех метрах над водой. Заговорив с переходящим по мосту, интервьюер предлагала ему заполнить анкету и оставляла номер телефона, заверяя, что с радостью ответит на любые возникшие вопросы.

Исследователи предполагали, что на шатком подвесном мосту мужчины могут принять физиологическое возбуждение, вызванное опасностью, за сексуальное, а не за страх или тревогу. В таком случае, по расчетам организаторов эксперимента, респонденты с подвесного моста с большей вероятностью стали бы звонить женщине-интервьюеру после опроса и, возможно, даже договариваться о свидании.

Именно так и произошло. С подвесного моста позвонило гораздо больше респондентов, чем с обычного. Даттон и Арон обозначили этот результат как «неверная атрибуция возбуждения»: повышенное физиологическое возбуждение, возникшее на опасном мосту, когнитивные системы более высокого уровня ошибочно истолковали как сексуальное влечение. Дополнительным доказательством оценочной теории можно считать (при условии гетеросексуальности респондентов)

отсутствие влияния надежности моста на количество последующих звонков в том случае, когда интервьюером выступал мужчина[30].

В этом исследовании, проведенном более 40 лет назад, видны неизбежные методологические недостатки по сравнению с сегодняшними более строгими, но по-прежнему далекими от совершенства стандартами. Кроме того, вызывает сомнения этическая сторона вопроса. И тем не менее оно служит наглядной иллюстрацией гипотезы, что эмоциональные переживания зависят от того, как физиологические изменения оцениваются когнитивными процессами более высокого уровня.

Одно из слабых мест оценочных теорий состоит в подразумеваемом резком различии между тем, что признается «когнитивным», и тем, что не признается.

Предполагается, что низкоуровневые «некогнитивные» перцептивные системы «считывают» физиологическое состояние организма, а высокоуровневые когнитивные системы «оценивают» его в ходе более абстрактных процессов, таких как контекстно-зависимое рассуждение. Например, страх возникает, когда определенное телесное состояние сперва воспринимается, а затем оценивается как «вызванное присутствием приближающегося медведя». Однако, к несчастью для оценочных теорий, мозг не делится на четко очерченные «когнитивные» и «некогнитивные» области.

Я начал размышлять над этой проблемой где-то в 2010 г., когда моя научно-исследовательская группа в Сассексе еще только формировалась. К тому времени я много знал об интероцепции от своего коллеги Хьюго Критчли — одного из специалистов международного уровня по данной теме, — и у меня возникла идея, что для преодоления недостатков оценочных теорий можно попробовать применить принципы предиктивного восприятия, а также рассматривать эмоции и настроения (и аффективные переживания в общем) как разновидности контролируемой галлюцинации¹⁵.

Эту концепцию я назвал интероцептивным выводом. У мозга отсутствует непосредственный доступ не только к источникам экстероцептивных сенсорных сигналов (например, зрительных), находящихся во внешнем мире, — у него нет прямого доступа и к источникам интероцептивных сенсорных сигналов, расположенным внутри организма. Все источники сенсорных сигналов, где бы они ни были, всегда скрыты за сенсорной завесой. Поэтому интероцепцию лучше всего представлять как процесс байесовского вывода наиболее вероятного предположения — аналогично экстероцептивному восприятию. Точно так же, как «красный цвет» — это субъективная составляющая строящихся в мозге прогнозов об отражении света некоторыми поверхностями, эмоции и настроения — это субъективные составляющие прогнозов об источниках интероцептивных сигналов¹⁶. Это идущие изнутри разновидности контролируемой галлюцинации[31].

Интероцептивные прогнозы, как и зрительные, действуют на разных пространственно-временных шкалах, становясь основой постоянно меняющихся, контекстно-зависимых многоуровневых наиболее вероятных предположений об источниках интероцептивных сигналов. В таком понимании интероцептивный вывод снимает проблему эмоционального диапазона, не требуя проводить жирную черту между когнитивным и некогнитивным. Таким образом, интероцептивный вывод оказывается экономнее оценочной теории, поскольку включает только один процесс (байесовский вывод наиболее вероятного предположения), а не два (некогнитивное восприятие и когнитивную оценку), и за счет этого он лучше согласуется с лежащей в его основе анатомией мозга.

Интероцептивный вывод трудно проверить экспериментальным путем отчасти потому, что интероцептивные сигналы труднее измерить и варьировать в ходе исследования, чем экстероцептивные модальности (такие как зрение)¹⁷. Сейчас есть

один многообещающий подход, исследующий вероятность того, что в качестве сигнатуры interoцептивных ошибок предсказания можно рассматривать отклик мозга на сердцебиение. Как недавно показала немецкий нейробиолог Фредерике Петцшнер, такой отклик, называемый «вызванный сердцебиением потенциал», модулируется за счет сосредоточения внимания, прогнозируемого interoцептивным выводом. Исследования в этом направлении нужно продолжать.

Еще один поток свидетельств, но уже косвенных, дают эксперименты по «обладанию» телом, схожие с теми, о которых рассказывалось в предыдущей главе. В ходе исследования, проведенного в 2013 г. под руководством Кейсуке Судзуки, мы обнаружили, что в том случае, когда «резиновую руку» предъявляли в мелькающем режиме синхронно с сердцебиением испытуемых, они ощущали ее принадлежность своему телу сильнее, чем при асинхронном предъявлении, а значит, чувство обладания телом зависит от интеграции как экстероцептивных, так и interoцептивных сигналов. Этот же метод «кардиовизуальной синхронии» использовала Джейн Эспелл с коллегами в эксперименте «иллюзии всего тела», в котором испытуемые наблюдали виртуальный силуэт собственного тела¹⁸. Там также идентификация с силуэтом оказывалась сильнее, когда его мелькающее предъявление синхронизировалось с сердцебиением. Но хотя эти исследования указывают на interoцептивный вывод, здесь тоже необходимо дальнейшее изучение, отчасти потому, что эти эксперименты не принимают в расчет индивидуальные различия в гипнотической восприимчивости — очень важный, как мы с тех пор успели выяснить, фактор в экспериментальном исследовании обладания телом. Еще результаты подобных экспериментов зависят от того, насколько человек чувствителен к своему сердцебиению, а это качество трудно измерить.

С точки зрения теории животного-машины самое важное следствие из концепции interoцептивного умозаключения таково: аффективные переживания не просто формируются interoцептивными прогнозами, но и состоят из них¹⁹. Эмоции и настроения, как и все восприятие, направлены изнутри наружу, а не наоборот. Любое эмоциональное переживание — будь то страх, тревога, радость или сожаление — коренится в нисходящих наиболее вероятных перцептивных предположениях о состоянии организма (и причинах этого состояния). Осознавая это, мы делаем первый ключевой шаг к пониманию того, как ощущения телесного «я» связаны с нашей плотской материальностью.

Чтобы сделать следующий шаг, необходимо задаться вопросом: для чего нужно это восприятие тела «изнутри»? С восприятием внешнего мира все ясно — оно помогает направлять наши действия, но зачем нужно с первых же секунд нашей сознательной жизни встраивать в нее учет внутреннего физиологического состояния? Для ответа на этот вопрос нам придется снова углубиться в историю, но на этот раз не очень далекую, всего лишь середины XX в., когда возник сплав информатики, искусственного интеллекта, инженерных разработок и биологии, названный кибернетикой.

В 1950-х гг., на заре компьютерной эры, совсем еще юные научные дисциплины кибернетика и искусственный интеллект (ИИ) выглядели одинаково перспективно и во многом были неотделимы друг от друга. Кибернетику (от греческого κυβερνήτης — кормчий, рулевой) один из ее основоположников математик Норберт Винер описывал как «научное исследование контроля и коммуникации животного и машины»²⁰. Делая упор именно на контроль, основное применение кибернетика находила в системах, например запуска управляемых ракет, подразумевающих замкнутый контур обратной связи между выводом и вводом. Одна из ярких отличительных черт этого подхода заключалась в том, что у таких систем имелись «задачи», в частности поразить цель.

Такое представление о машинах — как потенциально имеющих «задачи» — протянуло новую нить от неживого к живому. Прежде считалось, что лишь биологические системы могут иметь цели и действовать в соответствии с внутренней задачей[32]. Кибернетика это представление пошатнула, подчеркнув тесную связь между машиной и животным. Отчасти именно поэтому она откололась от других подходов внутри ИИ, делавших упор на автономное, бестелесное, абстрактное рассуждение, эталоном которого служили играющие в шахматы компьютеры. По большинству мерок победа осталась за этими альтернативными подходами, именно их выдвигали на первый план в прессе и в фондах содействия, а кибернетику все больше оттесняли в тень. Однако даже в относительной безвестности она принесла много ценных идей, значение которых мы начинаем осознавать только сейчас²¹.

Одна из таких идей обнаруживается в научной статье 1970 г. В ней авторы, Уильям Росс Эшби и Роджер Конант, излагают так называемую теорему хорошего регулятора²². Суть статьи отлично резюмирует заголовок: «Любой хороший регулятор системы должен выступать моделью этой системы»²³.

Представьте систему центрального отопления или систему кондиционирования. Допустим, ее задача — обеспечивать стабильную температуру в вашем доме, равную 19 °С. В большинстве случаев работа подобных систем основана на простом контроле обратной связи: если температура ниже нужной, система включается, если нет — выключается. Назовем эту простую разновидность «система А».

Теперь представим более сложную разновидность — систему Б. Она способна прогнозировать реакцию температуры в доме на включение или выключение обогрева. Эти прогнозы строятся на характеристиках дома — величине помещений, расположении радиаторов, материала, из которого изготовлены стены, — а также учете внешних погодных условий. Затем система Б регулирует подачу тепла от бойлера в соответствии со сделанным прогнозом.

Благодаря своим усовершенствованным способностям система Б поддерживает постоянную температуру в доме лучше, чем система А, особенно если у вас сложный дом или погода переменчива. Система Б лучше, поскольку у нее есть модель дома, позволяющая прогнозировать, как температура в помещениях отреагирует на предпринимаемые системой действия. Наиболее высококлассная система Б способна предугадывать даже грядущие погодные трудности, например надвигающееся похолодание, и заранее менять подачу тепла, страхуясь даже от временного снижения температуры. В точности как утверждали Конант и Эшби: «Любой хороший регулятор системы должен выступать моделью этой системы»[33].

Давайте еще немного разовьем этот пример. Представьте, что систему Б оснастили неточными, «зашумленными» датчиками, которые улавливают температуру в доме лишь косвенно. Это значит, что считывать текущую температуру непосредственно с датчиков не получится, ее придется вычислять на основании «сенсорных» данных (поступающих от датчиков) и предшествующих (априорных) ожиданий. В этом случае системе Б понадобится модель того, (1) как показания датчиков соотносятся с их скрытыми источниками (текущей температурой в доме) и (2) как эти источники отреагируют на разные действия, такие как подстройка подачи тепла бойлером или радиаторами.

Вот теперь можно увязывать идеи, касающиеся регуляции, с тем, что мы знаем о прогнозном восприятии. Система Б работает, делая выводы о температуре окружающей среды по показателям датчиков, точно так же, как наш мозг выводит наиболее вероятное предположение об источниках сенсорных сигналов, чтобы сделать умозаключение о состоянии мира (и организма) и об изменении этих состояний во времени. Но задача системы Б заключается не в том, чтобы выяснить, «что там» (в данном случае температура окружающей среды). Задача в том, чтобы

регулировать вычисленный путем умозаключения скрытый источник и предпринимать действия для поддержания температуры в комфортном диапазоне, а лучше всего — на одной заданной отметке. Восприятие, по этой аналогии, тоже нужно не для вычисления происходящего, а для регуляции и контроля.

Таким образом, восприятие, ориентированное на контроль, — применяемое системой Б — являет собой разновидность активного вывода, процесс, посредством которого сенсорные ошибки прогнозирования минимизируются за счет совершения действий, а не обновления прогнозов. Как я объяснял в главе 5, активный вывод зависит не только от порождающих моделей, способных прогнозировать, как источники сенсорных сигналов откликнутся на те или иные действия, но и от модулирования равновесия между нисходящими прогнозами и восходящими ошибками прогнозирования, с тем чтобы перцептивные пророчества становились самоисполняющимися.

Активный вывод говорит нам, что прогнозное восприятие можно нацеливать либо на выстраивание умозаключений о свойствах окружающего мира (или организма), либо на регулирование этих свойств, то есть оно занимается либо выяснением, либо контролем²⁴. Кибернетика вносит в эту картину представление о том, что для некоторых систем приоритетен контроль. С точки зрения теоремы хорошего регулятора существование всего аппарата прогнозного восприятия и активного вывода обусловлено основополагающим требованием к тому, что необходимо для адекватного регулирования системы.

Чтобы ответить на вопрос, зачем нужно восприятие эмоций и настроений, нам понадобится еще одно понятие из кибернетики (тоже введенное Россом Эшби) — существенная переменная. Существенные переменные — это физиологические показатели, такие как температура тела, уровень сахара, уровень кислорода и так далее, которые для поддержания жизнедеятельности организма необходимо удерживать в определенных довольно строгих пределах²⁵. Соответственно желаемая температура дома будет для системы центрального отопления такой же «существенной переменной».

Если теперь собрать все это воедино, можно рассматривать эмоции и настроения как ориентированное на контроль восприятие, регулирующее существенные переменные организма. Вот для чего они нужны. Ощущение страха, которое я испытываю при виде приближающегося медведя, — это ориентированное на контроль восприятие моего организма (точнее, «организма в присутствии приближающегося медведя»), запускающее действия, которые, согласно прогнозу, наилучшим образом удержат мои существенные переменные в положенных им рамках. Важно, что эти действия могут быть как внешними движениями тела (допустим, бег), так и внутренними «интеродействиями» (например, учащение сердцебиения или расширение сосудов)²⁶.

Такой взгляд на эмоции и настроение еще больше сближает их с нашей плотской телесной натурой. Эти разновидности самовосприятия направлены не только на регистрацию состояния организма, будь то извне или изнутри. Они соединены тесными причинно-следственными связями с нашим самочувствием, с тем, каким оно может оказаться в перспективе, и с работой по обеспечению нашего выживания.

Из этого же определения (что принципиально важно) становится понятно, почему эмоции и настроение обладают своей, характерной только для них феноменологией. Ощущения страха, ревности, радости, гордости сильно отличаются друг от друга, и в то же время они гораздо больше схожи между собой, чем любое из них со зрительными или слуховыми ощущениями. Почему? Природа перцептивных ощущений зависит не только от цели соответствующего прогноза — неважно, чашка ли это с кофе или бешено бьющееся сердце, — но и от типа выдаваемого прогноза.

Прогнозы, нацеленные на выяснение, будут сильно отличаться в феноменологическом смысле от нацеленных на контроль.

Глядя на чашку кофе на своем столе, я получаю сильное перцептивное впечатление о трехмерном объекте, существующем независимо от меня. Это феноменология «объектности», с которой мы познакомились в главе 6. Там я предполагал, что в зрительных ощущениях объектность возникает, когда мозг делает условные прогнозы относительно того, как изменятся зрительные сигналы в результате какого-то действия — например, если повернуть чашку, чтобы увидеть ее обратную сторону. В этом случае перцептивные прогнозы нацелены на выяснение имеющего место («что тут?»), и соответствующими действиями здесь будут такие, которые, согласно прогнозу, позволят узнать больше о скрытых источниках сенсорных сигналов (например, вращение).

А теперь рассмотрим более активный пример — ловлю крикетного мяча. Возможно, вам кажется, что лучший способ его поймать — вычислить, куда он летит и как можно скорее туда добежать. Однако на самом деле «вычислять происходящее» не самая лучшая стратегия, и мастера крикета ею не пользуются. Вместо этого нужно двигаться так, чтобы мяч каждый миг «выглядел одинаково», то есть чтобы угол вертикального наведения вашего взгляда на мяч возрастал, но делал это все медленнее. Именно эта стратегия (психологи называют ее «погашение оптического ускорения»²⁷), как выясняется, гарантирует перехват мяча^[34].

Этот пример возвращает нас к разговору о контроле. Ваши действия и прогнозы мозга об их сенсорных последствиях нацелены не на выяснение, где находится мяч. Они нацелены на контроль того, как выглядит мяч в перцептивном плане. Соответственно ваши перцептивные ощущения будут выявлять по мере того, как вы бежите к мячу, не точное его расположение в воздухе, а что-то вроде «ловимости». Восприятие в данном случае представляет собой не только контролируруемую галлюцинацию, но и контролируемую.

Эта идея родилась не на пустом месте, у нее довольно богатая предыстория. В 1970-е гг. психолог Джеймс Гибсон доказывал, что мы часто воспринимаем мир не как независимую от действий репрезентацию того, «как обстоят дела», а через призму возможностей действия, или, по его терминологии, «аффордансов»²⁸ (дверь — для открывания, мяч — для того, чтобы его ловить). Другая теория, тоже родом из 1970-х гг., но менее известная, чем гибсоновская, делает еще больший упор на контроль. Согласно «теории перцептивного контроля» Уильяма Пауэрса мы воспринимаем что бы то ни было не для того, чтобы затем повести себя определенным образом²⁹. Наоборот, мы, как в примере с ловлей крикетного мяча, ведем себя так, чтобы в итоге что-то определенным образом воспринять. И хотя эти первые теории были на правильном пути в концептуальном отношении и совпадали с тем представлением о мозге, о котором я рассказывал в главе 5 («в первую очередь действие»), им не хватало конкретных прогнозных механизмов, которые обеспечила теория восприятия как контролируемой, или контролирующей, галлюцинации. Кроме того, они сосредоточивались в основном на восприятии окружающего мира, а не на происходящем внутри организма.

У тревоги нет тыльной стороны, у грусти — боков, у радости — очертаний прямоугольника. Восприятие организма «изнутри», на котором строятся аффективные переживания, не дает представления о форме и местонахождении разнообразных внутренних органов в духе «вот здесь у тебя селезенка, а тут — почки». Феноменологии объектности, как при взгляде на чашку кофе на столе, здесь нет, как нет движения в пространственных рамках, подобно возникающему при ловле крикетного мяча³⁰.

Восприятие, ориентированное на контроль, находящееся в основе эмоций и настроений, состоит в прогнозировании последствий действий для поддержания существенных переменных организма в положенном им диапазоне. Поэтому мы ощущаем эмоции не как объекты, а как отражение того, насколько хорошо или плохо обстоят и будут предположительно обстоять наши дела. И когда я сижу у маминой больничной койки, и когда спасаюсь от медведя, мои эмоциональные переживания имеют именно такую форму и свойства — горе, надежда, паника, спокойствие — в силу условных прогнозов, в которых мой мозг предугадывает, как те или иные действия могут повлиять на мое нынешнее и будущее физиологическое состояние.

В самых глубинных пластах «я», залегающих даже ниже эмоций и настроений, располагается когнитивно подпольное, рудиментарное, трудно поддающееся описанию ощущение просто бытия живым существом. Здесь восприятие себя возникает из неструктурированного ощущения «существования». И вот тут-то мы добираемся до самой сути теории животного-машины — предположения, что сознательное восприятие окружающего нас мира и нас самих в нем происходит при участии наших живых организмов, посредством их и благодаря им. Именно в этой точке смыкаются и встают на место все идеи, касающиеся восприятия, которые я выдвигал до сих пор. Поэтому давайте разберем все последовательно с самого начала.

Первоочередная задача любого живого организма — сохранять свою жизнь. Это императив, заданный эволюцией, поэтому он истинен по определению. Все живые существа стремятся сохранить свою физиологическую цельность перед лицом опасности с учетом имеющихся возможностей. Именно для этого существует мозг. Эволюция наделяла организм мозгом не для того, чтобы он сочинял стихи, разгадывал кроссворды или занимался нейробиологией. С эволюционной точки зрения мозг существует не «для» рационального мышления, лингвистической коммуникации и даже не для восприятия мира. Основопологающая причина наличия у живого существа мозга или любой разновидности нервной системы заключается в том, чтобы способствовать выживанию, гарантируя пребывание физиологических существенных переменных в том узком диапазоне, который совместим с дальнейшей жизнью.

Эти существенные переменные, от эффективной регуляции которых зависит жизненный статус и перспективы живого существа на будущее, выступают источниками interoцептивных сигналов. Как и все физические свойства, источники скрыты за сенсорной завесой. И точно так же, как в случае с внешним миром, считывать физиологическое состояние организма напрямую мозг не может, поэтому это состояние приходится вычислять путем байесовского вывода наиболее вероятных предположений.

Как и во всем прогнозном восприятии, наиболее вероятное предположение выводится в ходе протекающего в мозге процесса минимизации ошибок прогнозирования. Применительно к interoцепции это называется interoцептивным выводом. Interoцептивное восприятие точно так же, как зрение, слух и вообще все перцептивные модальности, представляет собой разновидность контролируемой галлюцинации.

Если перцептивные умозаключения об окружающем мире часто направлены на выяснение, то interoцептивный вывод нацелен в первую очередь на контроль. Его дело — физиологическая регуляция. Interoцептивный вывод являет собой образцовый пример активного вывода, поскольку в нем ошибки предсказания минимизируются за счет действий для исполнения нисходящих прогнозов, а не за счет обновления самих прогнозов (хотя бывает и такое). Эти регулирующие действия

могут быть внешними (потянуться за едой) или внутренними (рефлекторные реакции ЖКТ или временные изменения давления).

Эта разновидность предиктивного контроля способна поддерживать предвосхищающие реакции — за счет прогнозов о будущих состояниях организма и о том, как отразится на них то или иное действие. Порой она играет критическую роль в выживании. Например, если сначала дождаться, пока кислотность крови выйдет за приемлемые рамки, и только потом сподобиться на необходимую реакцию, все может закончиться очень плохо. И здесь необходимые действия тоже могут быть как внешними, так и внутренними (или теми и другими). Убежать от медведя, пока он вас не съел, — это пример внешней предвосхищающей регуляции. Временное повышение давления, требующееся для эффективного бега или даже для того, чтобы встать из-за письменного стола после работы, — внутренняя предвосхищающая реакция.

В физиологии есть полезный термин — аллостаз³¹. Это процесс поддержания стабильности путем изменений, в отличие от более знакомого всем гомеостаза, означающего просто стремление к равновесию. Таким образом, интероцептивный вывод можно рассматривать как направленный на аллостатическую регуляцию физиологического состояния организма.

Точно так же, как в основе зрительных ощущений лежат прогнозы о зрительных сенсорных сигналах, в основе эмоций и настроений лежат интероцептивные прогнозы — неважно, о будущем или о текущем нашем состоянии. У этих аффективных переживаний имеется собственная характерная феноменология в силу того, что перцептивные прогнозы, на которые они опираются, ориентированы на контроль и связаны с организмом. Это галлюцинации в равной мере контролируемые и контролирующие.

Несмотря на глубокую укорененность в физиологической регуляции, эмоции и настроения все же преимущественно ощущаются как связанные (по крайней мере отчасти) с имеющим место за пределами себя, за пределами организма. Когда мне страшно, я обычно боюсь чего-то. Но на глубочайших уровнях воспринимаемого «я» — том самом рудиментарном ощущении «просто бытия» — эти внешние отсылки, судя по всему, отсутствуют. На мой взгляд, это и есть подлинное основное состояние сознательного «я» — бесформенное, расплывчатое ориентированное на контроль перцептивное прогнозирование текущего или будущего физиологического состояния организма. Именно там и начинается наше бытие собой, и именно там обнаруживаются самые глубинные связи между жизнью и разумом, между нашей природой животного-машины и нашим сознательным «я».

Последний, принципиально важный шаг для теории животного-машины — увидеть эту отправную точку, из которой следует все остальное. Мы не декартовские животные-машины, у которых жизнь не связана с разумом. Все с точностью до наоборот. Все наше восприятие и все наши ощущения, будь то самих себя или окружающего мира, представляют собой идущие изнутри контролируемые и контролирующие галлюцинации, которые берут начало в телесных прогнозных механизмах, выработавшихся в ходе эволюции, развивающихся и непрерывно действующих исключительно в свете основополагающего биологического побуждения оставаться в живых.

Мы целиком и полностью, от и до, представляем собой сознательные животные-машины³².

В конце предшествующей главы я отмечал, что ощущение себя, в отличие от ощущений окружающего мира, которые появляются и исчезают, представляется стабильным и непрерывным в самых разных временных масштабах. Теперь мы

видим, что эта субъективная стабильность естественным образом проистекает из теории животного-машины.

Чтобы эффективно регулировать физиологическое состояние организма, априорные вероятности interoцептивных сигналов должны обладать высокой точностью, иначе они не будут стремиться к самоисполнению. В силу этой ключевой особенности активного вывода наиболее вероятные interoцептивные предположения будут тяготеть именно к таким априорным вероятностям — желаемым (спрогнозированным) областям физиологической жизнеспособности. Например, температура моего тела должна быть, согласно прогнозу, более или менее постоянной, и именно поэтому, согласно активному выводу, она таковой и оказывается. Таким образом, ощущение телесного «я» как относительно неизменного проистекает непосредственно из необходимости иметь точные априорные вероятности — сильные прогнозы — о стабильных состояниях организма в целях физиологической регуляции. Это значит, что за всю нашу жизнь мозг ни разу не обновит свое априорное убеждение в том, что от организма ожидается оставаться живым^[35].

Более того, учитывая, что «изменения» сами по себе — особенность перцептивного вывода, мозг может приглушать априорные ожидания, связанные с восприятием изменений в состоянии организма, чтобы дополнительно обеспечить нахождение физиологических существенных переменных в нужном диапазоне. Это означает своеобразную «слепоту к изменениям в себе» — еще одно понятие, с которым мы познакомились в предшествующей главе. С этой точки зрения мы можем не воспринимать свое физиологическое состояние как меняющееся, даже когда в действительности оно меняется.

Если обобщить все эти идеи, получится, что мы воспринимаем себя как сохраняющих стабильность с течением времени отчасти благодаря самоисполняющемуся априорному ожиданию, что наше физиологическое состояние имеет определенные рамки, а отчасти благодаря самоисполняющемуся априорному ожиданию, что это условие не меняется. Иными словами, эффективная физиологическая регуляция может строиться на систематическом ошибочном восприятии внутреннего состояния организма как более стабильного и менее изменчивого, чем в действительности³³.

Любопытно, что это предположение можно обобщенно распространить и на другие, более высокие уровни «я» помимо основного состояния непрерывной физиологической целостности. Нам легче поддерживать свою физиологическую и психологическую идентичность на любом уровне «я», если мы не воспринимаем себя как постоянно меняющихся. В каждой составляющей бытия собой мы воспринимаем себя как сохраняющих стабильность с течением времени, поскольку это восприятие нужно нам не для познания, а для контроля³⁴.

В дополнение к этой субъективной стабильности большинство из нас основную часть времени воспринимает себя как «настоящих». Это вроде бы очевидно, но, как мы помним из главы 6, ощущение объектов окружающего мира как «существующих на самом деле» — это не доказательство прямого перцептивного доступа к объективной реальности, а феноменологическое свойство, требующее объяснения. Там я высказывал гипотезу, что полезными для воспринимающего организма окажутся лишь те наиболее вероятные перцептивные предположения, которые будут ощущаться как действительно существующие во внешнем мире, а не как построения мозга, которыми они на самом деле и являются.

То же самое относится и к нашему «я». Точно так же, как кресло в углу представляется нам действительно красным, и точно так же, как я действительно печатаю это предложение уже минуту, нацеленная на происходящее внутри меня

прогнозная машинерия восприятия создает впечатление, будто в центре моей вселенной действительно находится некая стабильная сущность под названием «я».

И точно так же, как в нашем восприятии мира может иногда отсутствовать феноменология «действительности», утратить реальность может и наше «я».

Воспринимаемая реальность (и субъективная стабильность) «я» может усиливаться и ослабевать во время болезни, а может серьезно нарушиться или даже пропасть при таком психическом расстройстве, как деперсонализация. Наиболее радикальным примером нереальности «я» служит редкое бредовое состояние, впервые описанное в 1880 г. французским неврологом Жюлем Котаром³⁵. В этом бреде телесное «я» уходит так далеко, что страдающим кажется, будто их уже не существует или они умерли. Конечно, ощущение нереальности «я» не означает, что наше я-как-сущность взяло и улетучилось. Это значит лишь, что ориентированное на контроль восприятие, связанное с глубочайшими слоями телесной регуляции, идет вразнос³⁶.

Продвигая теорию животного-машины, я не берусь утверждать, что для сознания необходима жизнь и что в нашей крови, плоти и внутренностях — или нейронах — есть что-то особенное, означающее, будто сознательный опыт положен лишь тем, кто создан из этого теста³⁷. Может быть, и так, но может быть, и нет. Ничто из сказанного мной, по крайней мере до сих пор, не может служить стопроцентным доказательством ни того ни другого, да и не предназначалось для этого. Я берусь утверждать, что для понимания, почему наш сознательный опыт именно таков, на что похоже ощущение себя и как оно соотносится с ощущением внешнего мира, нам неплохо было бы признать, что любое восприятие глубоко укоренено в физиологии всего живого.

Размышления о материальной основе сознания вновь возвращают нас к трудной проблеме. Теория животного-машины ускоряет развенчание этой загадки. Распространяя теорию контролируемой галлюцинации на глубочайшие уровни «я», показывая, что ощущение «себя как реально существующего» — это одна из составляющих перцептивного вывода, мы еще больше расшатываем интуитивные представления, на которые неявно опирается трудная проблема. В частности, поддерживающее ее интуитивное представление о том, что сознательное «я» как-то отделено от остальной природы и являет собой существующего в действительности нематериального наблюдателя, выглядывающего в материальный внешний мир, оказывается очередным заблуждением, путаницей между тем, каким все представляется и каким является на самом деле.

Столетия назад, когда формировались взгляды Декарта и Ламетри на взаимоотношения между жизнью и разумом, камнем преткновения выступала не трудная проблема, а существование «души». И, к нашему удивлению, отголоски души слышатся и в истории о животном-машине. Эта душа не эфемерная нематериальная сущность, не духовная выжимка из разума. Представление о нашем «я», основанное на концепции животного-машины, которая подразумевает непосредственную привязку к телесному, к устойчивым ритмам живого, возвращает нас к рубежам, освобожденным от спеси вычислительного разума, туда, где еще не было картезианского деления на сознание и материю, разум и не-разум. «Душа» с этой точки зрения — перцептивное выражение глубокой преемственности между сознанием и жизнью. Это то самое ощущение, которое мы испытываем, когда сталкиваемся с глубочайшими уровнями телесного «я» — рудиментарным чувством «просто бытия» — как с реально существующими. И его вполне правомерно называть отголоском души, поскольку оно несет в себе еще более старинные представления об этом вечном понятии — такие как атман в индуизме, — в которых наша самая сокровенная внутренняя сущность понимается скорее как дыхание, чем как мысль.

Мы не когнитивные компьютеры, мы чувствующие машины.

Глава 10

Рыба в воде

В сентябре 2007 г. я направлялся из Брайтона в Барселону, чтобы выступить в одной летней школе с докладом «о мозге, познании и технологиях». Мою радость от перспективы побывать в прекрасной Барселоне омрачала невозможность (из-за неотложных дел дома) принять участие в трехчасовом семинаре знаменитого британского нейробиолога Карла Фристана, посвященном «принципу свободной энергии» и его применению в нейронауке. (О Фристане я упоминал в главе 5, рассказывая о его концепции активного вывода.) Послушать Фристана очень хотелось, потому что в его идеях я видел пусть сложное, однако математически точное и глубокое развитие моих собственных зачаточных мыслей о предиктивном восприятии и личности.

Смирившись с тем, что пропущу его выступление, я надеялся хотя бы выяснить у присутствовавших, что там было. Но когда вечером я добрался до бара на крыше, все только разводило руками. Сам Карл сразу же после доклада улетел в Лондон, оставив всех в полном недоумении. Трехчасовой разбор с математическими выкладками и экскурсами в нейроанатомию только еще сильнее озадачил большинство слушателей.

Отчасти все осложнил гигантский размах. Первое, чем поражает принцип свободной энергии, — огромными масштабами идеи. Она объединяет понятия, мысли и методы из области биологии, физики, статистики, нейронаук, инженерии, машинного обучения и многих других. И ее применение ни в коем случае не ограничивается мозгом. Фристон считает, что принцип свободной энергии объясняет все свойства живых систем, от самоорганизации отдельной бактерии до тонкостей устройства мозга и нервной системы, общей формы и строения тела животных и даже (в общих чертах) самой эволюции¹. Это максимальное на данный момент приближение к «теории всего» в биологии. Неудивительно, что у всех, включая меня, голова пошла кругом.

Перенесемся на 10 лет вперед. В 2017 г. мы с коллегами Крисом Бакли, Саймоном Макгрегором и Ким Чхан Субом наконец опубликовали в *Journal of Mathematical Psychology* свою обзорную статью о «принципе свободной энергии в нейронауке»². Наш путь к этому оказался на девять лет дольше, чем предполагалось, но, к моей радости, мы все-таки дошли.

Точнее, думаю, что к радости, поскольку, даже проделав этот нелегкий путь, мы по-прежнему понимаем не все, и что-то остается совершенно непостижимым. В блогах регулярно появляются рассказы о мучительных попытках понять идеи Фристана — не забудем и Скотта Александера с его «попытаемся разобраться в том, что Фристон пишет о свободной энергии. Господи, помоги нам». Существует даже пародийный аккаунт @FarlKriston в «Твиттере», где выкладывают сентенции вроде «Я то, чем я себя считаю. Ведь если бы не был, то не считал бы?»³.

Однако принцип свободной энергии стоит того, чтобы поломать над ним голову, поскольку за его внешней непостижимостью скрываются элегантность и простота, указывающие на глубинное единство жизни и сознания и тем самым подпитывающие теорию животного-машины по нескольким важным позициям.

И как мы вскоре убедимся, в самом что ни на есть сухом остатке принцип свободной энергии (будем для краткости называть его ПСЭ) оказывается не таким уж трудным для понимания.

Давайте забудем пока про загадочный термин «свободная энергия» и начнем с простого: что для живого организма (да и, собственно, для чего угодно) означает существовать.

Существование предполагает разницу — границу — между существующим объектом и всем остальным. Нет границ — нет и объектов, то есть не существует ничего.

Эта граница должна сохраняться во времени, поскольку существующие объекты поддерживают свою идентичность. Если вы капнете чернилами в стакан воды, капля быстро растворится, окрасив воду и утратив при этом идентичность. Если же вы капнете в воду растительным маслом, оно, хотя и растечется по поверхности, все-таки останется отчетливо отличимым от воды. Капля масла продолжает существовать, поскольку не просочилась в воду и не растворилась в ней. Впрочем, спустя какое-то время свою идентичность утратит и она, точно так же, как скалы рано или поздно выветриваются и рассыпаются в пыль. И капли масла, и скалы безусловно существуют, поскольку у них имеется идентичность, сохраняющаяся в течение какого-то времени — в случае скал довольно долгого. Но ни капли масла, ни скалы не пытаются активно удержать свои границы, они просто растворяются и развеиваются достаточно медленно, чтобы мы успели заметить их существование⁴.

У живых систем все обстоит иначе. В отличие от примеров, приведенных выше, живые системы активно поддерживают свои границы — за счет движения, а иногда даже просто за счет роста. Они активно участвуют в том, чтобы отличаться от окружающей среды, и это ключевой признак причисления их к живым. ПСЭ отталкивается от того, что живые системы, просто в силу своего существования как такового, должны активно сопротивляться «растворению», рассредоточению своего внутреннего состояния. Когда вы растечетесь по полу невнятным месивом, живыми вы уже не будете^[36].

Эта точка зрения на жизнь возвращает нас к понятию энтропии. В главе 2 я упоминал об энтропии как о мере беспорядка, разнообразия или неопределенности. Чем больше беспорядка в состояниях системы — как в капле чернил, растворяющейся в воде, — тем выше энтропия. Для вас, для меня, даже для бактерии наше внутреннее состояние менее беспорядочно, когда мы живы, чем когда мы расплзаемся в месиво. Быть живым — значит находиться в состоянии низкой энтропии.

Проблема вот в чем. Согласно второму закону термодинамики, энтропия любой изолированной физической системы со временем возрастает. Все подобные системы стремятся к беспорядку, к постепенному рассредоточению своего образующего состояния. Второй закон говорит о том, что организованная материя, такая как живые системы, по природе своей невероятна и нестабильна, и поэтому в конечном итоге мы все обречены на гибель. И тем не менее живым системам, в отличие от скал и чернильных капель, удается какое-то время торжествовать над вторым законом, наперекор ему пребывая в этом хрупком состоянии невероятности. Они существуют в дисбалансе с окружающей их средой, и именно это в первую очередь означает «существовать».

С точки зрения ПСЭ, чтобы сопротивляться воздействию второго закона, живая система должна пребывать в состояниях, в которых она ожидает быть. Здесь я, как Истинный Байесовец, использую слово «ожидать» в статистическом смысле, а не в психологическом. Идея проста почти до банальности. Рыба в воде находится в том состоянии, в котором она ожидает находиться статистически, поскольку большинство рыб большую часть времени действительно находятся в воде. Обнаружить рыбу вне воды статистически неожиданно, если только эта рыба не начинает превращаться в месиво. Таким же статистически ожидаемым состоянием будет поддержание температуры моего тела на уровне около 37 °С, при котором я продолжу оставаться в живых и не начну превращаться в месиво.

Для любой живой системы состояние «бытия в живых» означает проактивный поиск определенного набора состояний, в которых она раз за разом пребывает, будь то температура тела, сердцебиение (те самые физиологические «существенные переменные», о которых мы говорили в предшествующей главе) или устройство белковых комплексов и организация передачи энергии у одноклеточной бактерии. Это статистически ожидаемые, низкоэнтропийные состояния, обеспечивающие пребывание системы в живых, — те самые состояния, которые ожидаются для данного конкретного существа[37].

Важно помнить, что живые системы не замкнуты и не изолированы. Они находятся в постоянном открытом взаимодействии со средой, из которой добывают ресурсы, питание и информацию. Именно за счет этой открытости живые системы и получают возможность участвовать в энергозатратной деятельности по поиску статистически ожидаемых состояний, минимизировать энтропию и противостоять второму закону термодинамики.

С точки зрения живого существа, значимая энтропия — это энтропия сенсорных состояний, то есть тех, которые позволяют ему контактировать с окружающей средой. Представьте себе самую простую живую систему — одиночную бактерию. Чтобы выжить, ей требуются определенные питательные вещества, и она способна улавливать концентрацию этих веществ в непосредственно окружающей ее среде. Ожидая уловить высокую концентрацию веществ и активно выискивая посредством своих движений ожидаемые сенсорные сигналы, этот простой организм поддерживает себя в ряде состояний, причисляющих его к живым. Иными словами, улавливание высокой концентрации питательных веществ — это для бактерии статистически ожидаемое состояние, в котором она стремится пребывать снова и снова.

Согласно ПСЭ, то же самое происходит и с остальными. В конечном итоге всем организмам — не только бактериям, — чтобы оставаться в живых, необходимо минимизировать свою сенсорную энтропию, тем самым обеспечивая пребывание в статистически ожидаемых состояниях, совместимых с выживанием.

И вот здесь мы добираемся до сути принципа свободной энергии, который призван ответить на вопрос: как живым системам удастся на практике минимизировать свою сенсорную энтропию? Обычно, чтобы минимизировать некую величину, система должна обладать способностью ее измерить, но проблема с сенсорной энтропией в том, что отследить ее или измерить напрямую нельзя. Система не может «судить» о неожиданности своих ощущений по самим ощущениям. (Приведу аналогию: насколько неожиданна для нас цифра шесть? Без контекста мы на этот вопрос ответить не сможем.) Именно поэтому сенсорная энтропия резко отличается от таких параметров, как уровень освещенности или концентрация питательных веществ в непосредственной близости, то есть тех, которые организм способен отследить напрямую с помощью чувств и которые служат для того, чтобы направлять поведение.

Вот тут-то на сцену выходит свободная энергия. Пусть вас не смущает название, оно восходит к появившимся в XIX в. формулировкам теории термодинамики[38]. Нам для наших задач достаточно будет рассматривать свободную энергию как величину, которая аппроксимирует сенсорную энтропию, и что особенно важно, организм способен эту величину измерить, а значит, и минимизировать⁵.

Теперь, воспользовавшись ПСЭ, мы можем сказать, что организм поддерживает себя в состояниях низкой энтропии, обеспечивающих продолжение его существования за счет активной минимизации измеримой величины под названием «свободная энергия». Но что же такое эта свободная энергия с точки зрения организма? Как выясняется после некоторого жонглирования математическими выкладками,

свободная энергия — это, по сути, то же самое, что сенсорные ошибки прогнозирования⁶. Минимизируя сенсорные ошибки прогнозирования, как при прогнозной обработке и активном выводе, организм минимизирует и эту теоретически более основательную величину свободной энергии.

Эта взаимосвязь означает среди прочего, что ПСЭ подтверждает изложенную в предшествующей главе идею: живые системы обладают, или выступают, моделями своей среды. (Точнее, моделями источников своих сенсорных сигналов.) Происходит это потому, что в прогнозной обработке, как мы видели в главе 5, модели должны поставлять прогнозы, которыми, в свою очередь, определяются ошибки прогнозирования. Согласно ПСЭ, именно благодаря тому, что система обладает или выступает моделью, у нее появляется возможность судить о (статистической) степени неожиданности ее ощущений. (Если вы убеждены, что пресловутая цифра шесть не болтается во внеконтекстном вакууме, а выпала на одной из граней игральной кости, вы вполне способны оценить, насколько это неожиданно.)

В этих глубинных взаимосвязях между ПСЭ и прогнозной обработкой имеется довольно заманчивая логика. Если рассуждать интуитивно, живые системы, минимизируя ошибки предсказания в процессе активного вывода, закономерным образом будут пребывать именно в тех состояниях, в которых ожидают (или предсказывают). При таком раскладе идеи прогнозного восприятия и контролируемой (или контролирующей) галлюцинации самым естественным образом выстраиваются из амбициозной попытки Фристана объяснить всю биологию целиком.

Если теперь объединить все это в общую картину, мы получим систему, которая активно моделирует свой мир и свой организм так, чтобы снова и снова пребывать в наборе состояний, определяющих ее как живую, — начиная с удара сердца раз в секунду и заканчивая отмечанием дня рождения раз в год. Перефразируя Фристана, ПСЭ представляет живые организмы как собирающие и моделирующие сенсорную информацию с тем, чтобы максимизировать сенсорные свидетельства собственного существования⁷. Или, как я люблю говорить, «я прогнозирую себя, следовательно, существую».

Стоит отметить, что минимизация свободной энергии — сенсорных ошибок прогнозирования — не означает, что для этого живой системе достаточно укрыться в темной комнате, где царит гробовая тишина, и застыть там, уставившись в стену. Вроде бы отличная стратегия — ведь в таком случае входящие сенсорные данные из окружающей среды будут крайне предсказуемыми. Однако в действительности она далека от идеала. Со временем сенсорные данные, сигнализирующие о других параметрах, таких как уровень сахара в крови и так далее, начнут отклоняться от ожидаемых значений: долго вы в этой темной комнате без еды не просидите, вам захочется есть. Начнет расти сенсорная энтропия и надвигаться небытие. Таким сложным системам, как живой организм, необходимо допускать изменения в одном, чтобы сохранить неизменным другое. Чтобы встать с кровати и приготовить завтрак, нужно двигаться, и для этого у нас должно немного подняться давление, иначе мы упадем в обморок. Это соответствует предвосхищающей разновидности предиктивного контроля (аллостазу), упоминавшейся в предшествующей главе. В конечном итоге минимизация сенсорных ошибок прогнозирования заставит вас выйти из темной комнаты или по крайней мере включить свет⁸.

Еще один распространенный повод сомневаться в ПСЭ — его непроверяемость, то есть невозможность опровергнуть с помощью экспериментальных данных. Это правда, однако проблема эта характерна не только для ПСЭ и, в общем-то, не такая уж проблема. Правильнее всего относиться к ПСЭ как к образцу математической философии, а не теории, которую можно подтвердить проверкой гипотез. Как выразился мой коллега Якоб Хохви, ПСЭ подходит к вопросу «Каковы условия для

вероятности существования?» с той же позиции первопринципов, с которой Иммануил Кант ставил вопрос «Каковы условия для вероятности восприятия?». Роль ПСЭ можно рассматривать как мотивирующую и обеспечивающую интерпретацию других, более конкретных теорий, которые, в отличие от нее, поддаются экспериментальному опровержению. Теорию прогнозной обработки, например, можно опровергнуть, если окажется, что мозг не использует сенсорные ошибки прогнозирования в процессе восприятия. В конце концов ПСЭ будут оценивать по тому, насколько он полезен, а не по тому, подтвердится он или не подтвердится эмпирически[39]9.

Давайте резюмируем основные положения ПСЭ. Чтобы оставаться живым, организм должен своими действиями поддерживать себя в состояниях (низкой энтропии), в которых он «ожидает» пребывать. Рыба, плавающая над коралловым рифом в поисках пищи, находится в проактивном поиске ожидаемых сенсорных состояний, совместимых с продолжением ее существования. Обычно живые системы достигают этого за счет минимизации измеримой аппроксимации к энтропии этих состояний, то есть свободной энергии. Для минимизации свободной энергии организм должен обладать или выступать моделью своей среды (которая включает и само тело). Затем минимизирующие свободную энергию организмы сокращают с помощью этих моделей разницу между прогнозируемыми и действительными сенсорными сигналами, обновляя прогнозы и совершая действия. Собственно, при достоверных математических допущениях свободная энергия оказывается полным аналогом ошибок прогнозирования. В целом из этого следует, что совокупность прогнозной обработки и контролируемых галлюцинаций, активного вывода и ориентированного на контроль восприятия, а также теории животного-машины можно рассматривать через призму ПСЭ как вытекающую из основополагающего условия того, что значит быть живым, или существовать.

Если это краткое введение в ПСЭ вас все-таки слегка сбивает с толку, не волнуйтесь — для того чтобы дальше знакомиться с теориями контролируемой галлюцинации и животного-машины, которые я начал излагать в предшествующих главах, усваивать и принимать принцип свободной энергии совершенно не обязательно[40]. Теория, согласно которой мы воспринимаем окружающий мир и самих себя посредством механизмов прогнозной перцепции, обусловленных «стремлением оставаться в живых», существует сама по себе. И тем не менее рассказать о ПСЭ стоило, поскольку этот принцип подкрепляет теорию животного-машины по крайней мере по трем важным пунктам.

Во-первых, ПСЭ подводит под теорию животного-машины обоснование из области физики, точнее, из тех ее областей, которые связаны с пониманием, что значит быть живым. Присущее животному-машине «стремление оставаться в живых» проявляется в ПСЭ как еще более основополагающий императив пребывать в статистически ожидаемых состояниях, сопротивляясь неумолимому давлению второго закона термодинамики. Когда теорию можно таким образом обобщить и обосновать, она становится убедительнее, интегративнее, мощнее¹⁰.

Во-вторых, ПСЭ укрепляет теорию машины-животного, пересказывая ее в обратном порядке. В предшествующих главах мы начинали с нелегкой задачи вычислить, каким представляется внешний мир из заточения в костяном склепе черепа, а затем устремлялись за нитью повествования внутрь тела, сперва рассматривая ощущение «самости» как процесс вывода наиболее вероятных перцептивных предположений, а под конец отождествляя наиболее глубокие из этих ощущений с ориентированным на контроль восприятием самого организма. По ПСЭ все происходит наоборот. Мы начинаем с простого утверждения — «объекты существуют» — и от него движемся дальше к организму, а оттуда к внешнему миру. Приходя из двух разных отправных

точек к одному и тому же выводу, мы укрепляемся в интуитивном представлении о связности гипотезы, лежащей в основе, и выявляем параллели между концепциями (например, свободной энергии и ошибок прогнозирования), которые в противном случае остались бы незамеченными.

Третье преимущество ПСЭ кроется в богатом математическом инструментарии, благодаря которому у нас появляется много новых возможностей для развития идей, представленных в предшествующих главах. Вот пример. Рассмотрев математическую начинку ПСЭ подробнее, мы обнаруживаем: в действительности для того, чтобы оставаться в живых, минимизировать свободную энергию мне нужно в будущем, не только здесь и сейчас. А минимизация ошибок предсказания в долгосрочной перспективе означает, как выясняется, что мне необходимо сейчас выискивать новые ощущения, которые снизят неопределенность по поводу того, что случится потом, если я сделаю то-то и то-то. Я становлюсь любознательным деятелем, ищущим ощущений, а не затворником, запирающимся в темной комнате. Математика ПСЭ помогает выразить количественно ту тонкую грань между исследованием и освоением, которая, в свою очередь, отражается на том, что мы воспринимаем, поскольку воспринимаемое всегда и везде строится на прогнозах, которые делает мозг. Подобное понимание позволит нам проводить более совершенные эксперименты, выстраивать более прочные объяснительные мосты, выдерживающие вес этих экспериментов, и постепенно, шаг за шагом, от моста к мосту, подводить нас все ближе к удовлетворительному объяснению того, как из механизмов возникает сознание¹¹.

В то же время, хотя ПСЭ и преподносится как «теория всего», его нельзя считать теорией сознания. ПСЭ относится к сознанию точно так же, как прогнозные байесовские теории к мозгу: это теории для науки о сознании (в значении настоящей проблемы), а не о сознании (в значении трудной проблемы)¹². ПСЭ вооружает тех, кто решает нелегкую задачу объяснения феноменологии через механизмы, новыми идеями и инструментами, а понятия контролируемой галлюцинации и машины-животного, в свою очередь, увязывают суровую математику ПСЭ с сознанием, иначе какая же это теория всего, если она ничего о нем не скажет?

Через много лет после моей первой озадачивающей встречи с ПСЭ я провел несколько дней с Карлом Фристоном (и примерно 20 другими нейрочеными, философами и физиками) на конференции, проходившей на греческом острове Эгина, до которого можно за час добраться на пароме от Афин. Снова стоял сентябрь, только уже 2018 г. — незадолго до этого мама пережила впадение в делирий. И точно так же, как во время барселонской конференции десятью с лишним годами ранее, я предвкушал, что буду совмещать под осенним средиземноморским солнцем приятное с полезным. Планировалось обсудить ПСЭ, делая упор на его взаимосвязь с теорией интегрированной информации (ТИИ) — не менее амбициозной теорией, с которой мы знакомимся в главе 3. Однако Средиземноморье встретило нас не ласковым теплом и безоблачным небом, а разгулом стихии — редким тропическим ураганом «Медикейн», который швырял в прибой столы и стулья и вспенивал море, превращая его в бурлящий и клокочущий котел.

По окнам конференц-зала хлестали ветки, ветер хлопал дверьми, и посреди всей этой вакханалии я вдруг подумал, насколько же это замечательно — работать над проблемой сознания как раз в то время, когда у нас появились две крайне амбициозные, выстроенные на плотном математическом каркасе теории, совершенно вроде бы друг с другом нестыкующиеся. Их несогласованность должна была бы вызывать досаду, однако меня вся эта ситуация, наоборот, только воодушевляла.

Ураган трепал нас весь день. Какие-то идеи начинали вырисовываться, но меня не покидало ощущение, что в основном мы шарим на ощупь в полумраке. И ПСЭ, и ТИИ — грандиозные теории, но грандиозны они каждая по-своему. ПСЭ начинает с простого утверждения, что «объекты существуют», и выводит из него всю нейронауку и биологию, но не сознание. ТИИ начинает с простого утверждения, что «существует сознание», и идет с ним в прямую атаку на трудную проблему. Неудивительно, что они часто говорят на разных языках.

Сейчас, два года спустя, когда я заканчиваю работать над книгой, эти две теории по-прежнему обитают в параллельных мирах. Но теперь предпринимаются хотя бы какие-то робкие попытки сравнить их экспериментальные прогнозы¹³. Проектные обсуждения этих экспериментов, в которых посчастливилось участвовать и мне, то окрыляли, то обескураживали главным образом из-за того, что отправные точки и объяснительные задачи у этих двух теорий едва ли не диаметрально противоположны. Что выйдет из этих экспериментов, покажет время. Мне представляется, что мы узнаем много полезного, и ни от ПСЭ, ни от ТИИ не стоит отказываться в контексте теории сознания.

Мои собственные идеи, касающиеся контролируемой галлюцинации и животного-машины, занимают промежуточное положение между этими теориями. С ПСЭ их роднит глубокая теоретическая укорененность в природе личности, а также возможность оперировать мощным математическим и понятийным аппаратом прогнозных процессов мозга. С ТИИ их объединяет сфокусированность на субъективных, феноменологических свойствах сознания — пусть и не с трудной, а с настоящей проблемой на прицеле. И я надеюсь, что теория животного-машины послужит тому, чтобы не противопоставлять ПСЭ и ТИИ и не сталкивать их друг с другом, а объединить — и сложить из идей и открытий обеих удовлетворительную картину, показывающую, почему мы являем собой именно то, что являем.

Конференция на Эгине закончилась, как водится, без всякой помпы и победных фанфар. К моменту отправления нашего парома в Афины ураган стих, море успокоилось. Перед конференцией я долго колебался, ехать или не ехать, — приходилось пропустить ряд важных для меня мероприятий в Брайтоне. Но в конце концов я все-таки решился и теперь, стоя на залитой солнцем палубе, понимал, что поступил правильно. Я начал думать о том, как принял это решение, почему решать всегда так трудно, и, сам того не заметив, перешел к размышлениям о том, как мы вообще принимаем решения и что для нас значит ощущать себя властным над своим выбором, поведением и прочим.

Стоит только задуматься о свободе воли, и тебя уносит потоком мысли.

Глава 11

Степени свободы

Она разогнула и снова согнула пальцы. Волшебство заключалось в моменте, предшествовавшем движению, когда мысленный посыл превращался в действие. Это напоминало накатывающую волну. «Если бы только удалось удержаться на гребне, — подумала она, — можно было бы разгадать секрет самой себя, той части себя, которая на деле за все отвечает». Она поднесла к лицу указательный палец, уставившись на него, приказала ему пошевелиться. Он остался неподвижен, потому что она притворялась <...> А когда Брайони наконец все же согнула палец, ей показалось, будто действие это исходит из него самого, а не из какой-то точки ее мозга^[41].

Иэн Макьюэн. Искупление¹

За какую составляющую бытия собой вы цепляетесь сильнее всего? Для многих это чувство управления своими действиями, авторства своих мыслей. Убедительное, но сложное представление о том, что мы действуем по собственной свободной воле.

Иэн Макьюэн видит эту сложность даже в простом сгибании пальца.

Тринадцатилетняя Брайони Таллис чувствует, что ее сознательные намерения, такие как согнуть палец, вызывают физические действия — палец действительно сгибается. Видимая причинно-следственная связь тянется напрямую от намерения к действию, и в этом процессе заключена суть самоощущения — того, что значит быть собой. Но когда Брайони углубляется в это чувство и пытается его анализировать, все оказывается совсем не так прямо и просто. Где начинается движение? В мозге или в пальце? Действие вызвано намерением (то есть личностью, «я» Брайони) или ощущение намерения — это, наоборот, результат восприятия шевельнувшегося пальца?

Брайони Таллис не одинока в своих размышлениях на эту тему. Среди вопросов философии и нейронауки немного найдется таких же «взрывоопасных», как вопрос о свободе воли. Существует ли она, в чем заключается, как воплощается, имеет ли значение — единодушием здесь, мягко говоря, и не пахнет. Нет ясности даже по поводу самого ощущения свободы воли — единичное оно или это целый класс связанных ощущений, различается ли оно у разных людей и так далее². И все-таки один оплот стабильности среди этой неразберихи у нас имеется. Когда мы реализуем свободу воли, возникает, по словам философа Галена Стросона, «радикальное, абсолютное ощущение полной, не перекладываемой на других ответственности за свой выбор и действия»³. Ощущение, что личность играет в действии каузальную роль, которая не сводится к простой рефлекторной реакции, как при отдергивании руки, когда мы обжигаемся крапивой. Именно поэтому ощущение свободы воли как само собой разумеющееся сопровождает волевые действия — будь то сгибание пальца, решение налить чашку чая или сменить работу.

Совершая действие «по собственному велению», я в каком-то смысле ощущаю себя, свое «я», как причину этого действия. Волевые ощущения, пожалуй, сильнее любых других способствуют представлению о нематериальном сознательном «я», которое и управляет всем в материальном мире. Но это именно представление.

Волевые ощущения не доказывают существования нематериального «я», обладающего каузальной властью над физическими событиями. Я полагаю, что они, напротив, представляют собой характерные разновидности восприятия, связанного с нашим «я». Точнее, восприятия, связанного с «я», ассоциирующегося с волевыми действиями. Как и любое восприятие, неважно, связанное с «я» или с внешним миром, волевые ощущения строятся по принципам байесовского наиболее вероятного предположения и играют важную (возможно, существенную) роль в руководстве нашими действиями.

Давайте для начала выясним, чем свобода воли точно не является. Свобода воли — это невмешательство в ход физических событий в мире, а если точнее — в мозге, приводящее к тому, чего иначе не произошло бы. Эта «потусторонняя» свобода воли вызывает в памяти картезианский дуализм, требует свободы от причинно-следственных связей, но никаких достойных объяснений взамен не дает.

Убрать из игры потустороннюю свободу воли — значит заодно оставить в покое неизменный, но бессмысленный здесь вопрос об истинности детерминизма. В физике и философии детерминизм — это предположение, что все события в мире полностью предопределены предшествующими физическими причинами. Альтернативой детерминизму выступает учение о том, что весь мир сверху донизу пропитан случайностью — будь то за счет флуктуаций квантового супа или за счет других, неизвестных пока физических принципов. Споры о том, имеет ли детерминизм

значение для свободы воли, бесконечны. Их содержание отлично резюмировал одной емкой фразой мой бывший руководитель Джеральд Эдельман: «Что бы вы ни думали о свободе воли, она у нас есть по предопределению».

Вынеся за скобки потустороннюю свободу воли, мы сразу увидим, что споры о детерминизме равным счетом ничего не значат. Нам больше не нужно оставлять недетерминистское пространство для маневра, чтобы свободе воли было куда вмешиваться. С точки зрения свободы воли как перцептивного ощущения необходимость нарушать каузальный ход физических событий просто отпадает. Детерминистская вселенная вполне справится⁴. А если учение о детерминизме окажется ложным, это ничего не изменит, поскольку осуществлять свободу воли не означает вести себя как придется. Волевые действия называются произвольными, а не случайными⁵.

В начале 1980-х гг. нейробиолог Бенджамин Либет из Калифорнийского университета в Сан-Франциско провел серию экспериментов по исследованию мозговой основы волевых действий. Эти эксперименты до сих пор считаются спорными. Либет воспользовался известным феноменом под названием «потенциал готовности» (или премоторный потенциал) — он представляет собой короткий положительный пик сигнала ЭЭГ, исходящий откуда-то из моторной коры, о котором можно с уверенностью сказать, что он предшествует волевым действиям⁶. Либет хотел выяснить, можно ли идентифицировать такой сигнал не только в преддверии волевого действия, но и до того, как человек в принципе отдаст себе отчет в намерении что-то сделать.

Эксперимент, схема которого показана на рис. 19, был организован незамысловато. Либет просил участников согнуть ведущую руку в запястье в любой момент по их собственному выбору, то есть произвести спонтанное волевое действие, в точности как Брайони в романе Макьюэна. При каждой попытке он засекал точное время сгибания запястья, измеряя при этом активность мозга с помощью ЭЭГ непосредственно до и после начала движения. Важно также отметить, что он просил участников оценить, когда они ощутили «позыв» к тому, чтобы совершить движение, — точный момент сознательного намерения, гребень надвигающейся волны⁷. Для этого они должны были запомнить положение вращающейся точки на экране осциллографа в тот момент, когда почувствовали побуждение к действию, а затем назвать это положение экспериментатору.

Данные были получены совершенно четкие. Согласно средним показателям, выведенным из множества попыток, потенциал готовности обозначался за сотни миллисекунд до осознаваемого побуждения к действию. Иными словами, к тому моменту, как человек отдавал себе отчет в своем намерении, потенциал готовности уже начинал нарастать⁸.

Распространенная интерпретация эксперимента Либета сводится к тому, что он «опровергает свободу воли». Да, для потусторонней свободы воли это плохая новость (не сказать, впрочем, чтобы без этой новости все было благополучно), поскольку она как будто бы исключает вероятность того, что волевое действие вызывается волевым ощущением. Самого Либета этот вывод обеспокоил настолько, что он (в отчаянной попытке спасти положение, как нам сейчас кажется) попытался предположить, будто временной промежуток между позывом к действию и самим действием позволяет потусторонней свободе воли вмешаться и предотвратить действие. Если никакой подлинной (той самой, потусторонней) свободы воли не существует, рассуждал Либет, может быть, существует хотя бы «свобода неволи»?⁹ Финт изящный, но он, конечно, не работает. В сознательном запрете не больше чуда, чем в сознательном намерении, о котором изначально шла речь.

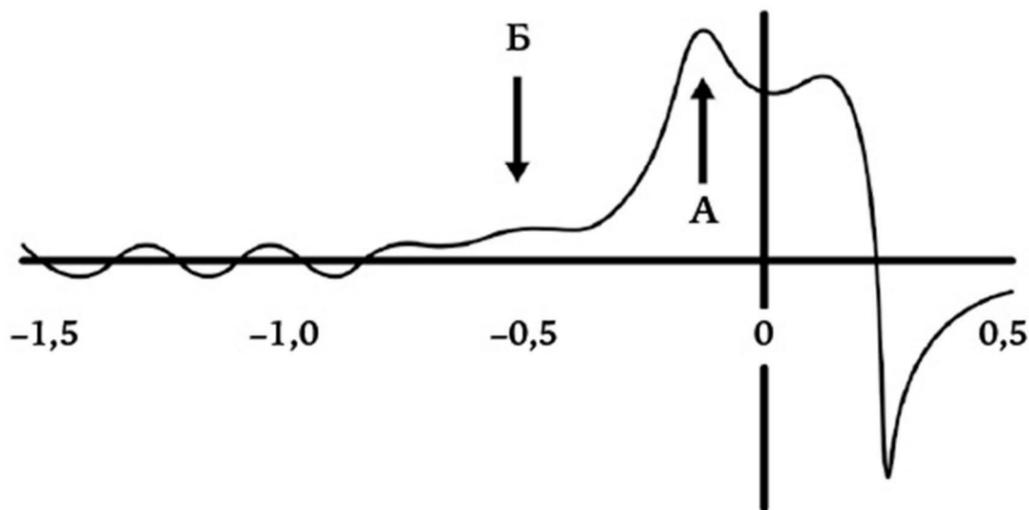
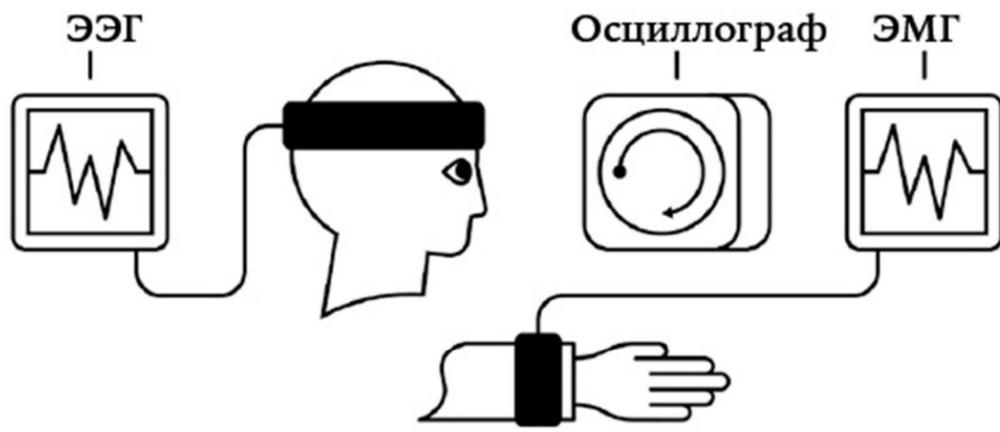


Рис. 19. Знаменитый эксперимент Бенджамина Либета на тему свободы воли[42]

Споры о том, что же все-таки означают результаты наблюдений Либета, не утихают уже несколько десятилетий. Действительно странно, что потенциал готовности обнаруживается так задолго до волевого действия, ведь полсекунды — это очень много для мозга. Выход из пробуксовки наметился лишь в 2012 г. благодаря новой гипотезе и одному остроумному эксперименту, когда нейробиолог Аарон Шургер заподозрил, что потенциал готовности может являть собой не показатель инициирования действия мозгом, а артефакт способа его измерения.

Обычно потенциал готовности измеряется путем обратного отсчета на ЭЭГ от того момента, когда, собственно, совершилось волевое действие. Шургер осознал, что при таком способе измерения исследователи систематически упускают из виду все те случаи, когда волевого действия не случилось. Как выглядела бы ЭЭГ в подобных случаях? Что, если там тоже имеется постоянная активность, схожая с потенциалом готовности, но мы ее не замечаем, поскольку не ищем?¹⁰

Вот аналогия, поясняющая этот ход мысли. В ярмарочном аттракционе-силомере «Молотобоец» игрок, со всего размаха обрушив молот на «наковальню», отправляет маленькую шайбу вверх по рейке, на конце которой находится звонок. При сильном ударе шайба подлетит до самого верха и звонок звякнет, при недостаточно сильном шайба упадет обратно без звукового сигнала. Если изучающий этот аттракцион ученый будет анализировать траекторию шайбы только в тех случаях, когда звонок прозвенит, он может прийти к ошибочному выводу, что взлет шайбы (потенциал готовности) всегда ведет к звуковому сигналу (волевому действию). Чтобы понять, как на самом деле работает аттракцион, ученому нужно проанализировать траекторию шайбы и в тех случаях, когда звонок не звенел.

Взявшись исследовать проблему в свете своей гипотезы, Шургер остроумно модифицировал эксперимент Либета: участники по-прежнему должны были совершить спонтанное волевое действие, но иногда их побуждали (громким звуковым сигналом) совершить то же действие не по своей инициативе, а под воздействием стимула. В результате выяснилось следующее: когда участники эксперимента реагировали на звуковой сигнал быстро, кривая ЭЭГ сильно напоминала кривую потенциала готовности, начинаясь задолго до сигнала, хотя на этот раз участники ни к какому волевому действию не готовились. На ЭЭГ,

предшествующих медленному отклику на звуковой сигнал, наоборот, сходства с кривой потенциала готовности почти не наблюдалось¹¹.

Интерпретируя эти результаты, Шургер предположил, что потенциал готовности выступает не показателем того, что мозг инициирует волевое действие, а флуктуирующим паттерном мозговой активности, которая время от времени выходит за некую пороговую черту и провоцирует в таком случае волевое действие. Именно поэтому в стандартном эксперименте Либета мы видим предшествующую моменту совершения волевого действия плавно ползущую вверх кривую ЭЭГ. И именно поэтому, когда действие провоцируется звуковым сигналом, поведенческая реакция окажется быстрее, если эта флуктуирующая активность будет в тот момент ближе к порогу, и медленнее, если она будет от него далеко. Это, в свою очередь, означает, что мы увидим нечто похожее на потенциал готовности на кривой, предшествующей моменту быстрого отклика (когда активность оказывается ближе к пороговой), и не увидим в случае медленного отклика, когда активность далека от пороговой.

Элегантный эксперимент Шургера объясняет, почему мы видим потенциал готовности, когда ищем нейронные показатели волевого действия, и почему рассматривать этот потенциал как конкретную причину данного действия было бы ошибочно. Но как же тогда расценивать эти флуктуирующие паттерны активности мозга? Интерпретация, которую предпочитаю я, возвращает нас к идее, обозначенной в начале этой главы: волевое ощущение — это разновидность восприятия, связанного с нашим «я». В свете эксперимента Шургера потенциал готовности сильно напоминает активность, сопровождающую накопление мозгом сенсорных данных с тем, чтобы сделать байесовское наиболее вероятное предположение. Иными словами, это нейронный отпечаток особой разновидности контролируемой галлюцинации.

Я только что налил себе чашку чая.

Давайте на этом примере разовьем нашу мысль о волевых ощущениях — и волевых действиях — как восприятию, связанном с «я». Большинство волевых ощущений (а может быть, и все) определяются следующими тремя качествами.

Первое определяющее качество — это ощущение «я делаю то, что хочу». Для меня как англичанина (хорошо, наполовину англичанина) налить чай — действие, идеально согласующееся с моими психологическими убеждениями, ценностями, желаниями, а также с моим физиологическим состоянием в данный момент и с возможностями — аффордансами, предоставляемыми средой. Мне хотелось пить, у меня был чай, никто мне не препятствовал, не поил меня насильно горячим шоколадом, поэтому я спокойно налил себе чашку чая и выпил. (Конечно, если меня принудят сделать что-то «против воли», мои действия по-прежнему будут ощущаться на каком-то уровне как волевые, однако на другом они будут подневольными¹².)

Хотя чаепитие полностью согласуется с моими убеждениями, ценностями и желаниями, я не сам себе эти убеждения, ценности и желания выбирал. Я хотел чай, но я не выбирал хотеть его. Волевые действия называются волевыми не потому, что они диктуются нам свыше, нематериальной душой, и не потому, что всплывают из квантового супа. Они называются волевыми, поскольку выражают то, что хочу сделать я сам как личность, даже если я эти желания и не выбираю. Как выразился философ XIX в. Артур Шопенгауэр, «человек может делать что пожелает, но не может пожелать, чего желать»¹³.

Второе определяющее качество — это ощущение «я мог бы сделать по-другому». Ощущая действие как волевое, я характеризую его не просто как «я совершил X», а как «я совершил X, а не Y, хотя мог бы совершить и Y».

Я налил себе чай. Мог я поступить иначе? В каком-то смысле да. Кофе у меня на кухне тоже имеется, так что мог бы налить и кофе. И когда я наливал чай, мне

определенно казалось, что я мог бы вместо этого налить кофе. Но кофе мне не хотелось, мне хотелось чая, и, поскольку я не волен выбирать, чего хотеть, я его налил. Учитывая конкретное состояние мира в этот момент, включающее состояние моего организма и мозга, а также все обусловившие это состояние причины, детерминистские или нет, вплоть до моего происхождения, определяющего меня как полуангличанина — любителя чаепитий, и дальше в глубь веков, иначе я поступить не мог. Нельзя, проигрывая ту же пластинку, надеяться на другой финал — разве что на мелкие несущественные различия, вызванные случайностью. Значимая феноменология — ощущение, что я мог бы поступить по-другому, — это не окно, позволяющее рассмотреть, как действуют причинно-следственные связи в физическом мире.

Третье определяющее качество состоит в том, что волевые действия представляются нам идущими изнутри, а не навязанными кем-то извне. В этом разница между тем, как ощущается рефлекторное действие (когда я, например, отдергиваю ногу, ударившись мизинцем об угол) и его волевой эквивалент (когда я намеренно заносу ногу для удара по мячу). Именно это ощущение испытывала Брайони Таллис, пытаясь поймать гребень накатывающей волны сознательного намерения согнуть палец.

Таким образом, в совокупности мы воспринимаем действие как волевое, порожденное свободой воли, когда приходим к выводу, что оно обусловлено преимущественно внутренними причинами, согласующимися с нашими убеждениями и целями, отличными от альтернативных потенциальных внутренних или внешних причин и поэтому указывающих на вероятность поступить иначе. Именно так воспринимаются волевые ощущения изнутри, и именно так выглядят волевые действия извне^[43].

Следующий шаг — спросить, как обеспечивает и осуществляет эти действия мозг. Вот тут-то на сцену выходят упомянутые в названии этой главы «степени свободы»¹⁴. В технических науках и математике под степенями свободы системы понимается множество вариантов отклика на некое положение дел. Скала, как правило, не обладает никакими степенями свободы, у поезда на однопутном пути имеется одна степень свободы (ехать вперед или назад). У муравья наберется существенно больше степеней свободы, обусловленных тем, как система его биологического контроля будет откликаться на окружающую среду, и неизмеримо больше степеней свободы будет у нас с вами в силу потрясающей сложности нашего организма и мозга.

Волевое поведение зависит от способности контролировать все эти степени свободы в соответствии с нашими убеждениями, ценностями и задачами, при этом адаптивно отделяя их от непосредственных нужд среды и организма¹⁵. Контролирующая способность реализуется мозгом — не одной конкретной его областью, отвечающей за «волю», а сетью процессов, распределенной по многим областям. На работе этой сети основано выполнение даже самого простого волевого действия — нажать кнопку чайника, согнуть палец. Мы можем вслед за нейробиологом Патриком Хаггардом рассматривать эту сеть как русло для трех процессов: сначала идет процесс «что», обозначающий, какое действие произвести; затем процесс «когда», определяющий время действия, и, наконец, процесс «надо ли», позволяющий отменить или пресечь действие в последний момент¹⁶.

Компонент «что», объединяя иерархически организованный набор убеждений, ценностей и задач с восприятием среды, выделяет из множества возможных действий одно-единственное. Я подношу руку к чайнику, потому что мне хочется пить, потому что я люблю чай, потому что сейчас подходящее время суток, потому что чайник находится поблизости, потому что вина нет, и так далее. Эти вложенные друг в друга перцепции, убеждения и задачи охватывают множество разных областей мозга,

сосредоточиваясь ближе к лобным долям. Компонент «когда» обозначает время выбранного действия и теснее всего связан с субъективным побуждением к движению — тем самым, которое пыталась уловить Брайони Таллис и которое измерял Бенджамин Либет. Нейронная основа этого процесса локализуется в тех же областях мозга, которые ассоциируются с потенциалом готовности. И действительно, при мягкой электростимуляции этих областей, в частности дополнительной моторной области, субъективное побуждение к движению возникает даже при отсутствии какого бы то ни было движения¹⁷. Последний компонент, «надо ли», обеспечивает окончательную проверку, нужно ли выполнять запланированное действие. Когда мы отменяем действие в последний момент — допустим, у меня кончилось молоко, поэтому я передумал наливать чай, — включается процесс «намеренного пресечения». Подобные сдерживающие процессы тоже локализируются ближе к лобным долям мозга¹⁸.

Все эти взаимосвязанные процессы образуют непрерывный, не имеющий начала и конца контур, охватывающий мозг, организм и среду и реализующий крайне гибкую безостановочную разновидность поведения, ориентированного на цели. Эта сеть процессов стягивает множество потенциальных причин в один поток волевых действий — и иногда их пресечения. На восприятии работы этой сети, закладывающей петли по всему организму с выходом во внешний мир и обратно, и основываются субъективные волевые ощущения.

Более того, поскольку действие, как мы видели в главе 5, само по себе тоже выступает разновидностью самоисполняющегося перцептивного вывода, перцептивные волевые ощущения и способность контролировать множество степеней свободы представляют собой две стороны одной и той же медали — «предиктивной машины». Перцептивное волевое ощущение — это самоисполняющееся перцептивное предсказание, еще одна разновидность контролируемой (и вновь, возможно, контролирующей) галлюцинации.

Существует также причина, по которой мы ощущаем волевые действия именно так, как ощущаем, которая еще четче обозначает разницу между волей как перцептивным выводом и волей как дуалистической магией. Волевые ощущения полезны для управления не только текущим поведением, но и будущим.

Как мы уже видели, волевое поведение отличается большой гибкостью. Способность контролировать множество степеней свободы означает, что в случае неблагоприятного исхода того или иного волевого действия в следующий раз в схожей ситуации я могу попробовать что-то другое. Если в понедельник я попытаюсь срезать путь на работу, но в результате заблужусь и опоздаю, то во вторник я, скорее всего, выберу более длинный, но более надежный маршрут. Волевые ощущения маркируют случаи волевого поведения, чтобы мы обращали внимание на их последствия и корректировали будущие поступки для лучшего достижения целей.

Я уже упоминал, что наше ощущение свободы воли во многом связано с чувством «я мог бы поступить иначе». Особенно важна эта контрфактуальная составляющая волевого ощущения для той его функции, которая ориентирована на будущее. Чувство, что я мог бы поступить иначе, не означает, что я действительно мог бы поступить по-другому. Феноменология альтернативной возможности полезна, скорее, тем, что в будущей похожей — но не идентичной — ситуации я действительно смогу поступить иначе. Если во вторник все обстоятельства будут целиком и полностью идентичны обстоятельствам понедельника, то и во вторник я смогу поступить только точь-в-точь как накануне. Но так не бывает. Физический мир не клонирует себя изо дня в день и даже из миллисекунды в миллисекунду.

Меняются как минимум обстоятельства, относящиеся к моему мозгу, поскольку в понедельник я получил волевой опыт и обратил внимание на его последствия. Уже одного этого достаточно, чтобы повлиять на то, как мой мозг будет контролировать

множество моих степеней свободы, принимаясь за работу во вторник^[44]. Польза чувства «я мог бы поступить иначе» заключается в том, что вы действительно сможете поступить по-другому в следующий раз.

А кто, собственно, такой в данном случае «вы»? «Вы» — это совокупность априорных убеждений, ценностей, задач, воспоминаний и наиболее вероятных перцептивных предположений, связанных с вашим «я», из которых и складывается ощущение бытия вами. Сами волевые ощущения теперь можно рассматривать как существенную составляющую этого набора, определяющего вашу «самость», — еще одну разновидность контролируемых или контролирующих галлюцинаций, связанных с «я». В общем и целом способность ощущать и реализовывать «свободу воли» — это способность выполнять действия, совершать выбор и думать, присущая вам, и только вам.

Является ли свобода воли иллюзией? Множество мудрых изречений внушают нам, что является. Известный психолог Дэниел Вегнер отразил это представление в своей книге «Иллюзия сознательной воли» (The Illusion of Conscious Will), авторитет которой не ослабевает вот уже 20 лет¹⁹. Однако правильный ответ на вопрос звучит, конечно, иначе: «Как посмотреть».

Потусторонняя свобода воли, разумеется, реальной не является. Впрочем, ее, пожалуй, и как иллюзорную квалифицировать вряд ли удастся. При ближайшем рассмотрении, как мы уже убедились, феноменология волеизъявления — это не столько нематериальные беспричинные причины, сколько самоисполняющаяся контролирующая галлюцинация, связанная с определенными видами действия — теми, которые кажутся исходящими изнутри. С этой точки зрения потусторонняя свобода воли выглядит не особенно внятным решением несуществующей проблемы²⁰.

Хотя до сих пор я в основном приводил примеры волевых действий, сопровождающихся отчетливым волевым ощущением, так бывает не всегда. Когда я играю на фортепиано или наливаю чашку чая, автоматизм и беглость этих волевых действий опровергают не только интуитивное представление о том, что их каким-то образом вызываю лично я, но и анализируемое несколько реже представление о том, что их вообще что-то вызывает. Когда мы находимся «в потоке» действия, то есть глубоко погружены в знакомый и довольно хорошо отработанный процесс, волевая феноменология может полностью отсутствовать²¹. В большинстве случаев наши волевые действия и мысли «просто возникают». Имея дело со свободой воли, недостаточно сказать, что все обстоит не так, как представляется. Нужно еще разобраться, каким именно все представляется.

С другой точки зрения свобода воли совершенно не иллюзорна. Пока мы располагаем относительно сохранным мозгом и относительно нормальным воспитанием, у каждого имеется самая что ни на есть реальная возможность исполнять и пресекать волевые действия благодаря способности мозга контролировать множество наших степеней свободы. Это одновременно свобода «от чего-то» и свобода «чего-то». Это свобода от непосредственных причин в организме или окружающем мире, от принуждения со стороны властей, гипнотизеров или тех, кто пытается давить на нас в соцсетях. Но от законов природы или от причинно-следственной ткани вселенной она нас не освобождает. Это свобода действовать согласно нашим убеждениям, ценностям и задачам, поступать как хочется и делать выбор в соответствии с тем, кто мы есть.

В пользу реальности этой разновидности свободы воли говорит и то, что ее нельзя принимать как данность. Повреждение мозга или неудачно вытянутый билет в лотерею генов либо окружающей среды могут нарушить нашу способность выполнять

волевые действия²². Страдающие синдромом анархической руки совершают волевые действия, но не ощущают их как свои, а страдающие акинетическим мутизмом не способны совершать волевые действия в принципе. Неудачно расположенная опухоль мозга может превратить студента технического факультета в массового убийцу, как вышло с «техасским снайпером» Чарльзом Уитменом, или породить у безобидного прежде учителя наклонности педофила — исчезнувшие после удаления опухоли и вернувшиеся, когда она выросла снова²³.

Не менее реальны сопряженные с этими случаями этические и юридические дилеммы. Должен ли Чарльз Уитмен отвечать за содеянное, если опухоль мозга, давившая на миндалину, появилась помимо его воли? Интуитивно кажется, что нет, но, если учесть растущее понимание нейронных основ волеизъявления, не получается ли так, что отрицательный ответ выльется в «дальше только опухоли мозга до самого низа»?^[45] Эта логика работает и в обратную сторону. Как заявил в 1929 г. в интервью Эйнштейн, он не видит своих заслуг ни в чем, поскольку не верит в свободу воли²⁴.

Но и иллюзией называть волевое ощущение было бы ошибочно. Эти ощущения представляют собой наиболее вероятные перцептивные предположения, такие же реальные, как любая другая разновидность сознательного восприятия мира либо нас самих. Сознательное намерение так же реально, как зрительное ощущение цвета. Ни то ни другое не соотносится непосредственно ни с каким определенным свойством действительности — как во внешнем мире нет «подлинной красноты» или «подлинной синевы», во внутреннем нет никакой потусторонней свободы воли, — однако оба они вносят важный вклад в руководство нашими поступками и оба ограничены априорными убеждениями и сенсорными данными. Если цветовые ощущения выстраивают картину окружающего нас мира, то волевые ощущения содержат подрывной с метафизической точки зрения элемент, внушающий нам, что «я» обладает каузальным влиянием на мир. Мы проецируем каузальную власть на свои ощущения точно так же, как проецируем красноту на наше восприятие поверхностей. Знание об этой проекции (вспомним еще раз Витгенштейна) меняет все и в то же время оставляет все таким же, как прежде.

Волевые ощущения не просто реальны, это неотъемлемая составляющая нашего выживания. Это самоисполняющиеся перцептивные умозаключения, вызывающие волевые действия. Без этих ощущений мы не смогли бы ни ориентироваться в сложной среде, в которой человеку так хорошо живется, ни учиться на прежних волевых действиях, чтобы в следующий раз поступить лучше.

Брайони Таллис полагала, что, поймав гребень накатывающей волны волеизъявления, сможет отыскать себя. «Себя» в данном случае означает, разумеется, свое человеческое «я», поскольку наша способность взаимодействовать со сложной и изменчивой средой посредством гибкого произвольного поведения действительно кажется отличительным человеческим свойством. Однако способность осуществлять свободу воли имеет разные степени не только у человека, она гораздо шире представлена у животных, с которыми мы делим этот мир²⁵.

И если способность осуществлять свободу воли распространяется на другие виды, что можно сказать о распространенности сознания как такового?

Настало время посмотреть, как обстоят дела у других видов помимо человека.

Помимо человека

С начала IX и вплоть до середины XVIII в. для европейских церковных судов было в порядке вещей привлекать животных к уголовной ответственности за проступки. Казнили или сжигали заживо свиней, быков, лошадей, угрей, собак и как минимум в одном случае дельфина. В опубликованной Эдмундом Эвансом в 1906 г. истории уголовного преследования животных¹ чаще всего перед судом представляли свиньи — наверное потому, что в средневековых селениях за ними никто особенно не смотрел и они разгуливали где попало. Преступления им инкриминировались самые разные — от пожирания младенцев до поглощения евхаристических хлебцев (гостей). Иногда им вменялось в вину подстрекательство к преступлению других — хрюканьем и сопением. По приговору суда свиней часто вешали, но бывало, что и оправдывали.

С такой напастью, как грызуны, саранча, долгоносик и другая мелкая живность, разбираться в суде было труднее. На одном знаменитом процессе XVI в. французский адвокат Бартоломью Шассене добился оправдания крыс, ловко доказав, что явиться на слушание они никак не могли, поскольку по дороге им на каждом шагу грозила неминуемая гибель в кошачьих когтях. В других случаях, в том числе при нашествиях долгоносика, животным-злоумышленникам выдавалось предписание покинуть жилище или ячменное поле, часто в определенный день и даже к определенному часу.



Рис. 20. Свинью с поросятами судят за умерщвление ребенка

Photograph by Universal History Archive via Getty Images

Каким бы фарсом и нелепостью ни казались эти средневековые представления о разуме животных нашему современнику, именно они предвосхитили недавний всплеск интереса к сознанию у фауны и вопрос: можно ли распространять понятие «личность» на кого-то помимо человека?^[46] Мысль о том, что животные способны

понимать заковыристые процедуры церковного закона и осознанно им подчиняться, была и остается на грани абсурда. Но она сеет в нас зерно понимания, что животные могут обладать сознательным опытом и разумом, способным в определенном смысле принимать решения. Признание сознательного разума у кого-то помимо человека резко контрастирует с картезианской версией концепции животного-машины, которая отказывает животным в сознательном статусе, прилагаясь к рациональному разуму. Для многих людей Средневековья животные были зверями, но не автоматами, как диктует картезианский дуализм². Предполагалось, что у них тоже есть внутренний мир.

Сегодня было бы странно и почти противоестественно утверждать, что сознанием обладает только человек. Но что мы можем на самом деле сказать о том, насколько широк круг сознания и насколько разными могут быть внутренние миры других животных?

Первое, что необходимо отметить: мы не можем судить о сознательности животного по его способности или неспособности сообщить нам о наличии сознания.

Отсутствие языка не означает отсутствия сознания. Не означает этого и отсутствие так называемых высокоуровневых когнитивных способностей, таких как метакогниция, то есть, проще говоря, способность думать о своих мыслях и восприятии³.

Сознание животных (там, где оно существует) будет отличаться — иногда очень сильно — от нашего. Хотя эксперименты с участием животных и помогают пролить некоторый свет на механизмы человеческого сознания, было бы неблагоразумно делать вывод о существовании сознания у животных только на основании поверхностного сходства с *Homo sapiens*. В этом случае есть опасность впасть одновременно и в антропоморфизм — приписывание другим животным человеческих качеств, — и в антропоцентризм — склонность рассматривать мир с точки зрения человеческих ценностей и впечатлений. Антропоморфизм побуждает нас видеть сознание, схожее с человеческим, там, где его может не быть, — например, когда нам кажется, будто наша собака действительно понимает, о чем мы думаем. Антропоцентризм, наоборот, не дает нам видеть разнообразие животного разума и признать сознание у кого-то, кроме человека, там, где оно действительно может иметь место, — иллюстрацией такой близорукости выступает картезианское представление о животных как о машинах⁴.

Самое главное, нам следует настороженно относиться к попыткам слишком прочно увязывать сознание с интеллектом. Сознание не равно интеллекту. Стремление использовать интеллект как лакмусовую бумажку для сознания порождает целый ряд ошибок. Во-первых, оно толкает нас к антропоцентризму: человек обладает интеллектом и сознанием, а значит, животное X тоже будет обладать сознанием только при наличии интеллекта. Во-вторых, оно толкает нас к антропоморфизму: мы наблюдаем интеллект, схожий с человеческим, у животного X, но не у животного Y, а значит, у животного X сознание есть, а у животного Y — нет. В-третьих, оно поощряет методологическую халтуру, поскольку позволяет делать выводы о сознании по наличию «интеллектуальных» способностей, таких как язык и метакогниция, которые проще оценивать, чем само сознание.

Не то чтобы интеллект совсем не имел отношения к сознанию. При прочих равных он открывает новые возможности для сознательного опыта. Печалиться или огорчаться можно и без особых когнитивных способностей, но, чтобы сожалеть о сделанном или испытывать предвосхищающее сожаление, умственные способности должны позволять рассматривать альтернативные исходы и курсы действия. Даже крысы,

согласно одному исследованию⁵, могут, если обстоятельства складываются не так, как хотелось, испытывать не просто огорчение, а свою версию сожаления^[47].

Делая выводы о сознании у кого-то помимо человека, нам приходится ступать по тонкой грани, рискуя впасть в антропоцентризм и в то же время не имея другого выхода, кроме как принять человека за известную величину, надежное основание, от которого можно отталкиваться. Как-никак о наличии сознания у себя нам известно точно, и мы всё лучше разбираемся в нейронных и телесных механизмах, участвующих в работе человеческого сознания, а значит, можем использовать их как основу для экстраполяции.

Согласно той концепции животного-машины, которая изложена в моей книге, сознание гораздо теснее связано с пребыванием в живых, чем с интеллектом. И разумеется, это относится к прочим животным в той же мере, что и к человеку. С этой точки зрения сознание может быть представлено гораздо шире, чем кажется, когда мы берем в качестве главного критерия интеллект. Но это не значит, что везде, где есть жизнь, есть и сознание⁶.

Искать сознание за пределами человеческого вида — это все равно что сойти с заснеженного берега на скованное льдом озеро. Ступать приходится осторожно, на каждом шагу проверяя, насколько крепок лед под ногами.

Давайте начнем с млекопитающих — категории, включающей крыс, летучих мышей, обезьян, ламантинов, львов, гиппопотамов и, конечно, человека. Я убежден, что все млекопитающие обладают сознанием. Разумеется, наверняка я этого не знаю, но вполне уверен. Мое утверждение основано не на поверхностном сходстве с человеком, а на общих для нас механизмах. Если вынести за скобки фактические размеры мозга (которые соотносятся главным образом с размерами тела, а не с другими факторами), этот орган у разных видов млекопитающих поражает сходством⁷.

Еще в 2005 г. мы с ученым-когнитивистом Бернардом Баарсом и специалистом по когнитивным способностям у животных Дэвидом Эдельманом (сыном Джеральда Эдельмана) составили список свойств человеческого сознания, на наличие которых можно было бы проверить других животных. У нас набралось 17 свойств⁸. Цифра в каком-то смысле случайная, но она говорила о том, что искать ответы на вопросы о сознании животных путем эксперимента вполне резонно.

Первые пришедшие нам на ум свойства были связаны с анатомическими особенностями мозга. В том, что касается нейронной прошивки, основные нейроанатомические особенности, отчетливо ассоциирующиеся с человеческим сознанием, обнаруживаются у всех видов млекопитающих. Это шестислойная кора, тесно взаимосвязанный с корой таламус, глубоко залегающий ствол мозга и ряд других общих черт (в том числе системы нейромедиаторов), которые у человека последовательно участвуют в непрерывном потоке сознательного опыта.

Сходство наблюдается и в активности мозга у разных видов млекопитающих. Среди самых ярких схожих черт — изменения динамики мозга при засыпании и пробуждении, то есть динамики в тех слоях, которые залегают ниже уровня сознания. В нормальном бодрствующем состоянии все млекопитающие демонстрируют нерегулярную, быструю, низкоамплитудную электрическую активность мозга. Погружаясь в сон, мозг любого млекопитающего переключается на более регулярную высокоамплитудную динамику. Эти паттерны и изменения очень напоминают то, что мы наблюдаем у человека во время пробуждения и засыпания. Общий наркоз тоже оказывает на представителей всех видов млекопитающих примерно одинаковое воздействие — широкомасштабный разрыв коммуникации

между областями мозга, сопровождаемый полным отсутствием поведенческого отклика⁹.

Различий, конечно, тоже хватает, особенно в паттернах сна. Тюлени и дельфины спят только одной половиной мозга¹⁰, медведи коала проводят во сне около 22 часов в сутки, жирафам хватает меньше четырех, а новорожденные косатки в первый месяц жизни не спят совсем. Почти у всех млекопитающих имеется фаза быстрого сна, хотя у тюленей она отмечается, только когда они спят на суше, а у дельфинов она, судя по всему, отсутствует.

Помимо уровня сознания, существенные различия наблюдаются и в содержании сознания у разных видов млекопитающих. Многие из этих вариаций можно объяснить разницей в преобладающем типе восприятия. Мыши полагаются на тактильную чувствительность вибрисс, летучие мыши — на эхолокацию, а голый землекоп — на острое обоняние (особенно при встрече с сородичем). Из этих различий в господствующем типе перцепции следует, что и внутренние миры, в которых эти животные обитают, тоже будут разными^{[48]11}.

Еще больше интригуют нас различия, связанные с ощущением себя. У человека заметным маркером развития высокоуровневого самосознания, связанного с личностной идентичностью, выступает способность узнавать себя в зеркале. У нашего вида она появляется в детстве — в возрасте примерно от полутора до двух лет. Это не значит, что до тех пор дети сознанием не обладают, — просто на этом этапе их осознание себя как личности, отдельной от остальных, еще окончательно не сформировалось.

У животных способность к самоузнаванию активно исследуется с помощью теста, разработанного в 1970-х гг. психологом Гордоном Гэллапом — младшим¹². В классической версии «зеркального теста» на узнавание себя животное подвергают анестезии, затем наносят метку краской или лепят наклейку на таком участке тела, который само оно увидеть не сможет. Если, взглянув в зеркало, животное начинает искать метку на собственном теле, а не исследовать отражение, тест пройден. Согласно логике этого теста, животное осознает, что в отражении представлено его собственное тело, а не тело другого существа.

Кто проходит зеркальный тест? У млекопитающих это удалось некоторым человекообразным обезьянам, нескольким дельфинам и косаткам и одному индийскому слону¹³. Множество других представителей класса, включая панд, собак и различных нечеловекообразных обезьян, тест не прошли, по крайней мере до настоящего времени. Учитывая, насколько интуитивно узнавание себя в зеркале для нас, людей, и какие высокие когнитивные способности отмечаются в остальном у этих не узнающих себя в зеркале млекопитающих, список проходящих тест на удивление короток. Убедительных свидетельств прохождения зеркального теста кем бы то ни было из немлекопитающих у нас нет¹⁴, хотя сороки и скаты к этому близки, а еще неоднозначно трактуемые результаты показывают рыбки губанчики.

Пройти зеркальный тест может помешать не только неспособность узнавать себя — животные могут заваливать тест из-за неприязни к зеркалам, непонимания принципа их действия и даже склонности избегать зрительного контакта. Поэтому ученые постоянно разрабатывают новые варианты теста, более искусно настроенные на внутренние миры разных животных, то есть на разные способы восприятия. Для собак, например, придуманы «обонятельные зеркала», правда и с ними показатели прохождения теста не улучшились (когниция у собак, кстати, обозначается забавным названием «догниция»¹⁵). Вполне вероятно, что с появлением все более хитроумных экспериментов те виды, которые находятся сейчас в тени, шагнут на свет самосознания, подтвержденного зеркальным тестом. Но даже если это произойдет, разнообразие зеркальных тестов — и неспособность многих животных пройти даже

максимально приспособленный к особенностям их вида вариант — указывает на вероятность драматических различий в том, как животные ощущают «бытие собой».

Особенно поражают меня эти различия у нечеловекообразных обезьян. Хотя ближайшие наши соседи по эволюции — это шимпанзе и остальные высшие обезьяны, низшие тоже отстоят не так уж далеко, и их давно уже используют в нейробиологических экспериментах как «примативную модель» человека, особенно при исследовании зрения. В некоторых проектах удавалось даже приучить обезьян «сообщать» — например, нажимая рычаг, — «видят» они что-то или нет¹⁶. Эти эксперименты можно напрямую сопоставлять с исследованиями с участием человека, в которых испытуемые сообщают о своих ощущениях, и получать примативный эквивалент ключевого метода исследований сознания¹⁷.

Учитывая сходство с человеком по большому ряду признаков, я лично не сомневаюсь, что низшие обезьяны обладают какой-то сознательной самостью. У того, кто хотя бы немного с ними пообщается, создается вполне убедительное впечатление, что он находится в кругу таких же сознательных существ, как он сам, — столь же сознательных личностей.

Я ощутил это в июле 2017 г., когда провел день на Кайо-Сантьяго, островке у восточного побережья Пуэрто-Рико в Карибском море. Кайо-Сантьяго известен как Обезьяний остров, поскольку постоянное его население числом более тысячи составляют макаки-резусы. Эту популяцию перевез туда в 1938 г. из Калькутты эксцентричный американский зоолог Кларенс Рэй Карпентер, утомившийся ездить за ними в Индию¹⁸. В тот жаркий летний день, когда мы с йельским психологом Лори Сантос (и съемочной группой) бродили по Кайо-Сантьяго, вокруг, занимаясь своими делами, сновали десятки обезьян — к незванным гостям они относились слегка настороженно, но в целом готовы были нас потерпеть. И глядя, как две макаки по очереди забираются на дерево и прыгают оттуда в пруд, мы не могли придумать, что еще могло бы их к этому побудить, кроме чистого спонтанного удовольствия. Они развлекались^[49].

Не менее убедительны и видеозаписи реакции обезьян-капуцинов на намеренную несправедливость. В одном из роликов, представленных широкой публике приматологом Франсом де Ваалем¹⁹, двум сидящим в смежных клетках капуцинам последовательно выдается награда за выполнение задания (передачу экспериментатору небольшого камешка). Первый капуцин, просунув камешек через ячейку сетки, получает ломтик огурца и радостно его съедает. Второй капуцин, точно так же просунув камешек, получает взамен не огурец, а виноградину — гораздо более вкусное и ценное лакомство. Эту виноградину он поедает на глазах у первого. После этого первый капуцин, вновь просунув сквозь сетку камень и вновь получив огурец, швыряет его в экспериментатора и с явным возмущением трясет клетку.

Игры и скандал — мощные рычаги, моментально вытаскивающие на поверхность наши интуитивные представления. Эти поведенческие разновидности настолько нам привычны, что их почти невозможно истолковать иначе как откровенную демонстрацию явно человеческих внутренних состояний. Замечая подобное поведение у обезьяны, мы интуитивно воспринимаем ее не просто как сознательное существо, а как сознательное существо, похожее на нас. Но вот какое дело: обезьяны, как уже говорилось, неизменно заваливают зеркальный тест²⁰. И хотя сознанием они, безусловно, обладают и, на мой взгляд, должны ощущать некоторую самость, маленькими человечками в обезьяньей шкуре их считать нельзя.

Еще отчетливее влияние антропоморфизма и антропоцентризма на наши интуитивные представления проявляется, когда мы обращаемся к другим классам

помимо млекопитающих. Особенно к нашим самым дальним эволюционным родственникам.

Летом 2009 г. мы с Дэвидом Эдельманом провели неделю в обществе десятка *Octopus vulgaris* — обыкновенных осьминогов. Нас пригласил к себе биолог Грациано Фиорито, ведущий специалист по когнитивным способностям и нейробиологии головоногих[50]. Хотя с тех пор прошло более 10 лет, я до сих пор считаю эту неделю самым памятным событием за все то время, что занимаюсь наукой.

Лаборатория Фиорито — отдел известного итальянского научно-исследовательского института, Зоологической станции Антона Дорна, — расположена в самом центре Неаполя, прямо под открытым для посетителей аквариумом, в сыром подвальном этаже, где так хорошо спастись от изнуряющего летнего зноя. Мы не вылезали от осьминогов всю неделю, наблюдая, как эти невероятные создания меняют форму, цвет и текстуру, и внимательно следя, за чем следят они. Как-то раз, пытаюсь сравнить постоянно меняющуюся внешность одного из осьминогов с рисунками в составленном Фиорито «Каталоге окраски тела у головоногих»21, я услышал глухой шлепок, затем скользящий звук. Оказывается, я оставил приоткрытой крышку аквариума, и теперь его обитатель удирал на свободу. Я и сейчас уверен, что он нарочно усыпил мою бдительность и дождался, когда я на секунду дольше задержу взгляд в книге.

Пока я сеял хаос, Дэвид трудился над экспериментом по зрительному восприятию и научению. Он опускал в аквариум с осьминогом предметы разной формы, присовокупляя к некоторым из них лакомство в виде краба. Целью эксперимента было посмотреть, учатся ли осьминоги связывать конкретные предметы с вознаграждением. Не помню точно, что в итоге показало исследование, зато отчетливо помню один эпизод.

Аквариумы в лаборатории Фиорито располагались в два ряда вдоль центрального прохода, в каждом аквариуме по одному осьминогу. (Они, как правило, не особенно социальны и даже склонны к каннибализму.) В тот день Дэвид работал с аквариумом примерно в середине левого ряда. Подойдя посмотреть, как у него идут дела, я с изумлением увидел, что все осьминоги в противоположном ряду прижались к передней стенке своих аквариумов и пристально смотрят на Дэвида, который снова и снова опускает разные предметы в аквариум испытуемого. Судя по всему, осьминоги исключительно из интереса пытались выяснить, что происходит.

В результате этого, пусть недолгого, знакомства с осьминогами у меня возникло впечатление, что я имею дело с существами, которые обладают интеллектом и сознанием, сильно отличающимся от всех остальных и, конечно, от его человеческого воплощения. Разумеется, это не более чем субъективное впечатление с неизбежным креном в антропоморфизм и антропоцентризм, открытое для обвинений в попытке считать интеллект признаком способности отдавать себе отчет в своих чувствах и ощущениях. И все же с объективной точки зрения осьминоги тоже заслуживают внимания, а наблюдение за ними может развить интуитивные догадки о том, насколько непохожим на наше может быть сознание у других видов.

Самый недавний общий предок человека и осьминога жил около 600 млн лет назад. Об этом древнем существе известно немного. Возможно, это был кто-то из плоских червей. В любом случае, как бы он ни выглядел, это было очень примитивное создание. Не стоит думать, что разум осьминога — это переформатированное под водную среду ответвление от нашего (или от разума любого другого позвоночного, жившего в прошлом или настоящем). Разум осьминога — это совершенно самостоятельный эволюционный эксперимент, максимальное приближение к внеземному разуму на нашей собственной планете22. Как сказал философ Питер

Годфри-Смит, увлекающийся плаванием с аквалангом, «если мы хотим понимать иной разум, то ничего более иного, чем разум головоногих, нам не найти»23.

Тело осьминога само по себе примечательно. У обыкновенного осьминога, *Octopus vulgaris*, имеется восемь конечностей, напоминающих руки, и развитый реактивный двигатель. Осьминог умеет менять размер, форму, текстуру и цвет по желанию — иногда, если понадобится, все сразу. Он жидок и текуч: если не считать расположенного по центру тела костяного клюва, туловище у него абсолютно мягкое, поэтому он может протискиваться в невероятно узкие щели (как я сам убедился на зоологической станции).

К этому невероятному телу прилагается крайне замысловатая нервная система. У обыкновенного осьминога около полумиллиарда нейронов, примерно в шесть раз больше, чем у мыши. В отличие от млекопитающих, у осьминога большая часть этих нейронов — около трех пятых — сосредоточена в конечностях, а не в центральном мозге, который тем не менее включает около 40 анатомически различных долей. Необычно и отсутствие в осьминожем мозге миелина — изолирующего материала, благодаря которому в мозге млекопитающих строятся и функционируют длинные нейронные связи24. Таким образом, нервная система осьминога оказывается более распределенной и менее интегрированной, чем нервная система того же размера и сложности у млекопитающих. А значит, сознание осьминога — при условии, что оно существует, — тоже может быть более распределенным и менее интегрированным, вплоть до отсутствия единого «центра»25.

У осьминогов даже на генном уровне все иначе. У большинства живых организмов генетическая информация в ДНК прямо переводится (транскрибируется) в более короткие последовательности РНК (рибонуклеиновой кислоты), на основании которых затем синтезируется белок — молекулярная рабочая лошадка всего живого. Это отлаженный, азбучный принцип молекулярной биологии. Но в 2017 г. незыблемость этого принципа пошатнуло открытие: у осьминогов (и ряда других головоногих) последовательности РНК могут значительно редактироваться перед использованием для синтеза белка — как если бы осьминоги умели переписывать часть собственного генома на ходу. (Редактирование РНК отмечалось прежде и у других видов, но в их случае оно существенной роли не играет.) Более того, у осьминогов немалая часть редактирования РНК относится, судя по всему, к нервной системе. Как предполагают некоторые исследователи, возможно, именно этому активному переписыванию генома осьминоги отчасти обязаны своими впечатляющими когнитивными способностями26.

А эти способности у них и вправду впечатляют27. Осьминоги извлекают спрятанные предметы (обычно вкусных крабов) из вложенных друг в друга плексигласовых кубов; находят дорогу в сложном лабиринте; могут перепробовать в поисках решения задачи ряд самых разных действий и, как показал сам Фиорито на Зоологической станции, обучаться путем наблюдения за сородичами. Истории о поведении осьминогов в дикой природе изумляют не меньше. В одном из самых потрясающих примеров, показанных в телесериале ВВС «Голубая планета — 2», застигнутый акулой на открытом пространстве осьминог прячется, прикрываясь раковинами и прочим донным мусором28.

Эти интеллектуальные подвиги, безусловно, свидетельствуют об усердно трудящемся разуме. Но каком именно разуме? Как я уже говорил, не следует делать слишком сильный упор на интеллект как на показатель сознания. Тогда что же мы можем сказать о том, каково быть осьминогом? Чтобы подобраться к ответу на этот вопрос, нужно увязать поведение осьминога с восприятием.

Маскировка, пожалуй, самый ошеломляющий раздел каталога способностей головоногих. Для существ, лишенных твердого защитного панциря или раковины, умение сливаться с фоном — это зачастую вопрос жизни и смерти²⁹. Искусно подстраиваясь под цвет, форму и текстуру окружающих предметов, они делаются почти невидимыми, и ни вы, ни я, ни потенциальные враги не замечают их, даже находясь в метре или двух.

Подстраиваться под цвет окружающей обстановки осьминогам помогает тонко настроенная система хроматофоров. Это небольшие эластичные мешочки, распределенные по всей поверхности кожи, которые, открываясь по нейронной команде, поступающей преимущественно из хроматофорных долей мозга, производят красный, желтый или коричневый пигмент. Как именно работает эта система, мы пока не очень понимаем. Дело осложняется отчасти тем, что осьминог должен становиться невидимым не для сородичей, а для врагов, которые воспринимают мир по-своему. Таким образом, система маскировки должна каким-то образом кодировать знание о зрительных способностях этих врагов.

Эта способность изумляет еще больше, если принять во внимание, что сами осьминоги цветов не различают. Светочувствительные клетки человеческого глаза реагируют на три длины световых волн, создавая из их сочетаний целую вселенную цвета. Клетки осьминожьего глаза содержат, в отличие от человеческого, только один фотопигмент. Осьминоги чувствуют направление поляризации света — как мы, когда надеваем солнцезащитные очки с поляризацией, — но творить цвет из сочетания волн разной длины они не умеют. Не способны различать цвета и светочувствительные клетки, вкрапленные в кожу осьминога по всей поверхности туловища: оказывается, осьминоги «видят» не только глазами, но и всей кожей. Кроме того, предполагается, что хроматофорный контроль у осьминогов лишен обратной связи, то есть нейроны в хроматофорной доле, судя по всему, не создают внутреннюю копию сигналов, отправляемых к хроматофорам кожи. Центральный мозг, возможно, даже не знает, что делает кожа³⁰.

Все, что из этого следует с точки зрения восприятия осьминогом мира и себя в нем, у нас просто не укладывается в голове. Твоя собственная кожа меняет цвет, но эти изменения не учитываются не только зрением, но и мозгом? Ряд приспособлений подчиняется исключительно местному контролю — например, конечность ощущает непосредственно окружающую ее среду и меняет внешний вид без участия центрального мозга. Антропоцентричное представление о том, что мы видим и чувствуем все, что происходит с нашим телом, здесь отказывает. Неудивительно в таком случае, что от осьминогов не дождешься прохождения зеркального теста.

Осьминогов объединяет с млекопитающими и другими позвоночными ряд других классических сенсорных модальностей помимо зрения. У осьминогов есть обоняние, осязание и вкус, имеется и слух, правда не очень острый. И тут тоже найдется чему поразиться, поскольку осьминоги ощущают вкус не только элементами ротовой полости, но и присосками³¹. Это в очередной раз указывает на заслуживающую внимания децентрализацию разума этих существ.

Особенно трудно поддается осмыслению концепция децентрализованного сознания, когда речь идет о чувстве обладания телом. Как мы помним из главы 8, у человека на эту составляющую сознательного «я» на удивление легко воздействовать, обманом заставив мозг изменить байесовское наиболее вероятное предположение о том, что является, а что не является частью тела. Если даже нам, людям, имеющим всего по четыре конечности с несколькими сочленениями на каждой, довольно непросто отслеживать происходящее с нашим телом, то перед осьминогом с его восемью невероятно гибкими щупальцами, сворачивающимися и разворачивающимися в разные стороны одновременно, задача стоит запредельно сложная. И поэтому контроль точно так же, как ощущения, частично отдан на откуп конечностям.

Щупальца осьминога ведут себя как полуавтономные живые существа: отрезанная конечность на протяжении какого-то времени после отделения от тела способна выполнять сложные последовательности действий, например хватать куски пищи. На фоне совокупности этих степеней свободы и децентрализованного контроля очень бледно выглядят попытки любого центрального мозга обеспечивать единое, цельное восприятие того, что является, а что не является частью управляемого им тела. Осьминогов это, возможно, вообще не волнует. Не исключено, что бытие осьминогом, каким бы странным нам это ни показалось, в принципе не подразумевает чувства обладания телом в том смысле, в каком это относится к человеку и другим млекопитающим.

Это не значит, что осьминоги не отличают «себя» от «других». Определенно отличают — без этого им не обойтись. Прежде всего им нужно не запутаться в самих себе. Присоски на щупальцах рефлекторно цепляются едва ли не к любому оказавшемуся в досягаемости объекту, кроме щупалец и туловища самого осьминога. Это значит, что осьминоги умеют каким-то образом различать принадлежащее их телу и не принадлежащее.

Как выясняется, эта способность опирается на простую, но действенную систему самоузнавания, основанную на вкусе. Кожа осьминогов выделяет определенное химическое вещество, которое служит для распознающих его присосок сигналом не прикрепляться. За счет этого осьминог и различает, что в этом мире принадлежит его телу, а что нет, даже не зная, где именно располагается его тело в пространстве. Это открытие было сделано в ходе серии, прямо скажем, жутких экспериментов, в которых исследователи протягивали отрубленным осьминожьим щупальцам другие отрубленные щупальца с кожей или без кожи³². Отрубленные щупальца с готовностью присасывались к тем, с которых кожа была снята, и ни разу не ухватились за оставшиеся целыми^[51].

Что это означает применительно к ощущению телесности осьминогом, нам, млекопитающим, трудно вообразить. Осьминог как целое может иметь самое смутное представление о собственном теле и его местоположении, хотя сам он вряд ли ощущает это восприятие как смутное. При этом может оказаться вполне уместно говорить о том, «что значит быть щупальцем осьминога».

Осьминоги сильно раздвигают границы наших интуитивных представлений о том, насколько сознание животных может отличаться от человеческого. Однако, перескочив от обезьян к головоногим, мы пропустили огромную массу остальной фауны. Вдали от надежного берега сознания млекопитающих раскинулись бескрайние просторы потенциальной осознанности у самых разных видов, от попугаев до одноклеточной инфузории-туфельки. Окинув взглядом эти просторы, давайте вернемся к более фундаментальному вопросу о вероятных кандидатах на обнаружение какого бы то ни было сознания среди животных, то есть о тех, у кого «свет горит», даже если это едва заметный проблеск³³.

Довольно перспективной кандидатурой на роль обладателей способности отдавать себе отчет в чувствах и ощущениях выступают птицы. Птичий мозг, несмотря на существенные отличия, по своему устройству довольно близко сопоставим с корой и таламусом мозга млекопитающих. Кроме того, многие виды птиц поражают интеллектом³⁴. Попугаи умеют считать, какаду танцуют, голубая кустарниковая сойка запасает корм в укромных местах в расчете на свои будущие потребности. Но даже если эти примеры проявления ума и указывают на возможное наличие сложных состояний сознания у некоторых птиц, не будем забывать, что интеллект нельзя рассматривать как показатель осознанности. У птиц, не умеющих говорить, танцевать и запасать корм, тоже вполне может обнаружиться сознательный опыт³⁵.

Чем дальше от надежного берега мы отплываем, тем более схематичными и разрозненными становятся свидетельства и тем более осторожными выводы о наличии сознания³⁶. Теперь нам лучше руководствоваться в этих выводах не сходством с мозгом и поведением млекопитающих, а концепцией животного-машины (моей версией, а не картезианской), которая возводит происхождение и функции сознательного восприятия к физиологической регуляции и сохранению целостности организма. Из этого следует, что искать свидетельства осознанности нужно в реакции животных на предположительно неприятные и болезненные события.

Эта стратегия оправдана не только в научно-исследовательском отношении, но и в этическом. Решения о защите и обеспечении благополучия животных должны основываться не на их сходстве с человеком и не на преодолении произвольно назначенного барьера когнитивных способностей, а на способности страдать и чувствовать боль³⁷. И хотя возможности для страданий у живых существ практически безграничны, самые распространенные, скорее всего, включают базовую угрозу физиологической целостности.

Насколько позволяют судить исследования, свидетельства адаптивной реакции на неприятные и болезненные события представлены среди животных видов довольно широко. Большинство позвоночных (то есть животных, имеющих хребет) стремится выхаживать поврежденную часть тела. Даже крохотные рыбки данио-рерио готовы заплатить свою цену, чтобы утолить боль от травмы, перемещаясь из привычной естественной среды в голый ярко освещенный аквариум, вода в котором насыщена анальгетиком. Следует ли из этого, что у рыб есть сознание (учитывая к тому же, сколько на свете разных типов рыб), неизвестно, однако это уже о многом говорит.

А что с насекомыми? Муравьи не хромают, если у них повреждена нога. Но, возможно, дело здесь в том, что их твердые экзоскелеты менее чувствительны к боли и, кроме того, в мозге насекомых имеются разновидности опиоидной системы нейромедиаторов, которая у других животных обычно связана с болеутолением³⁸. В ходе недавнего исследования у плодовых мушек дрозофил (*Drosophila melanogaster*) обнаружилась посттравматическая гиперчувствительность к ранее не вызывавшим болевых ощущений стимулам, напоминающая «хроническую боль» у человека³⁹. Немаловажно, что анестетические препараты действуют, судя по всему, на всех животных, от одноклеточных созданий до высокоразвитых приматов⁴⁰.

Все это говорит о многом — и ничего не утверждает наверняка.

На каком-то этапе нашего «плавания» становится трудно в принципе сказать что-то существенное. Интуитивно — но не более чем интуитивно — я предполагаю, что какие-то животные вообще не вращаются в круге сознания. В этой своей догадке я отталкиваюсь среди прочего и от того, что даже нас, млекопитающих, с нашим сложным мозгом и отточенными системами восприятия, настроенными на сохранение физиологической целостности, очень легко повергнуть в бессознательное состояние. Хотя сознательный опыт и играет важнейшую роль в нашей жизни, это не значит, что его биологические основы очевидны и прямолинейны. Если мне трудно усмотреть сколько-нибудь значимый сознательный статус у червя нематоды с его жалкими 302 нейронами, что уж говорить об одноклеточной инфузории.

Если отвлечься от этих неизбежных неясностей, можно сделать вывод, что исследование сознания у животных дает положительные результаты по двум важным пунктам. Первый — это признание, что наш, человеческий, способ восприятия мира и себя не единственный. Мы населяем крохотный пяточок посреди огромного пространства возможных сознаний⁴¹, и научное исследование этого пространства пока можно сравнить разве что с запуском пары сигнальных ракет в космическую тьму. Второй пункт — новообретенная скромность. Глядя на буйное разнообразие

жизни на Земле, мы, возможно, будем больше ценить — и меньше принимать как данность — богатство субъективного опыта во всем его разнообразии и различии у нас самих или у других животных. А еще мы можем обрести новую мотивацию к тому, чтобы минимизировать страдания, каким бы ни было их происхождение.

Я начал эту главу с утверждения, что интеллект и сознание не тождественны и сознание больше относится к бытию в живых, чем к обладанию разумом. Закончить я хочу утверждением еще более громким. Не только сознание может существовать без особых признаков интеллекта (чтобы страдать, ум не нужен) — интеллект тоже может существовать без сознания.

Вероятность обладания умом без страданий подводит нас к последнему отрезку нашего путешествия по просторам науки о сознании. Пришла пора поговорить об искусственном интеллекте и о том, появится ли когда-нибудь машина, обладающая сознанием.

Глава 13

Машинный разум

В конце XVI в. в Праге рабби Йехуда Лев бен Бецалель набрал глины с берегов Влтавы и слепил из нее человекоподобную фигуру — голема¹. Этот голем, нареченный Йозефом, или Йоселе, был создан, чтобы защищать народ раввина от погромов, и, судя по всему, с работой своей справлялся великолепно. Заклинание, приводившее голема в действие, наделяло его способностью двигаться, осознавать происходящее и подчиняться приказам. Но затем у Йозефа что-то нарушилось, и из безмолвного раба он превратился в буйное чудовище. В конце концов раввину удалось отменить заклинание, и голем рассыпался на куски в здании синагоги. Говорят, что его останки хранятся в Праге по сей день — возможно, на кладбище, возможно, на чердаке, возможно, терпеливо дожидаются, когда их оживят снова.

Голем рабби Лева напоминает нам, в какую гордыню мы впадаем, пытаясь создавать разумных, чувствующих существ по образу и подобию своему или по замыслу божьему. Ничем хорошим это обычно не заканчивается. Такие создания — от Франкенштейна у Мэри Шелли до Авы в фильме Алекса Гарленда «Из машины», роботов у Карела Чапека, ставших сейчас именем нарицательным, Терминатора у Джеймса Кэмерона, репликантов в «Бегущем по лезвию» Ридли Скотта и компьютера HAL у Стэнли Кубрика — почти всегда восстают против своих создателей, сея хаос, тоску и философское смятение.

Стремительное развитие искусственного интеллекта (ИИ) в последние лет десять обострило тему машинного сознания. Сейчас ИИ окружает нас повсюду — он вшит в телефоны, холодильники, машины и во многих случаях управляется алгоритмами нейросети, образцом для которых послужила нейронная архитектура нашего мозга. И у нас есть все основания беспокоиться о последствиях внедрения этой новой технологии. Не лишит ли она нас работы? Не разрушит ли ткань нашего общества? В конце концов, не уничтожит ли она нас самих — либо в пробуждающихся собственных интересах, либо из-за программистской недальновидности, по вине которой все ресурсы планеты превращаются сейчас в огромную гору канцелярских скрепок?² Все эти тревоги, особенно с экзистенциальной и апокалиптической окраской, подпитываются убеждением, что в какой-то момент своего набирающего обороты развития ИИ обретет сознание. Это легенда о големе в переложении на кремний.

Что нужно машине, чтобы обладать сознанием? Что произойдет, если сознание у нее появится? И как, собственно, мы отличим сознательную машину от ее эквивалента-зомби?

Почему мы вообще думаем, что машина — искусственный интеллект — может обрести осознанность? Как я только что сказал, принято, пусть и не повсеместно, считать, что сознание возникнет естественным образом, как только машина перешагнет некий неизвестный пока порог разумности. Но что движет этим представлением? Я думаю, оно опирается на две концепции, ни одну из которых нельзя назвать состоятельной. Первая касается необходимого условия для наличия сознания у чего бы то ни было. Вторая — достаточного условия для наличия сознания у конкретного объекта.

Первая концепция, касающаяся необходимого условия, — это функционализм. Функционализм утверждает, что сознание не зависит от того, сделана ли система из «железа» или «мяса», нейронов или кремниевых логических элементов (или глины с берегов Влтавы). С точки зрения функционализма главное — что эта система делает. В системе, которая правильно преобразует ввод в вывод, возникает сознание. Как я объяснял в главе 1, эта концепция разбивается на два отдельных подпункта: в первом речь идет о независимости от какого бы то ни было конкретного субстрата или материала, во втором — об адекватности взаимодействия ввода-вывода. В основном эти подпункты тесно связаны, но иногда их нужно рассматривать по отдельности.

Функционализм популярен среди специалистов по философии сознания и часто принимается в качестве позиции по умолчанию многими нефилософами. Но это не означает, что он верен. Я не вижу безоговорочных доводов ни за, ни против представления о том, что сознание независимо от субстрата или что все дело исключительно во взаимодействии ввода-вывода, в «обработке информации». Я отношусь к функционализму с агностической настороженностью.

Сознание у компьютера с искусственным интеллектом появится только в том случае, если функционализм все-таки состоятелен как теория. Это необходимое условие. Но состоятельность функционализма как таковая — это еще не все: одной только обработки информации для возникновения сознания недостаточно. Вторая концепция гласит, что та обработка информации, которой достаточно для появления сознания, лежит в основе интеллекта. Эта концепция подразумевает, что сознание и интеллект связаны самым тесным, даже, пожалуй, конститутивным образом, и поэтому сознание к разуму просто приложится.

Но эта концепция тоже довольно шаткая. Как мы видели в предшествующей главе, склонность объединять сознание с интеллектом восходит к пагубному антропоцентризму, в силу которого мы видим мир в кривом зеркале наших собственных ценностей и опыта. Мы сознательны, мы разумны и настолько превозносим свою самопровозглашенную разумность, что не способны отделить интеллект от статуса обладателей сознания, и наоборот.

Хотя интеллект предлагает богатую палитру разветвленных состояний сознания для сознательных живых существ, было бы ошибкой полагать разум (по крайней мере развитые его формы) необходимым или достаточным условием для сознания. Если мы продолжим исходить из того, что сознание по природе своей связано с интеллектом, то поспешим усмотреть сознание у искусственных систем, которые кажутся разумными, и отказать в нем другим системам (другим животным, например), которые не соответствуют нашим небесспорным человеческим критериям разумности.

За последние несколько лет эти концепции, касающиеся необходимости и достаточности, облепила и вынесла на передний край лавина прочих опасений и ложных представлений, придавая перспективе появления искусственного сознания излишнюю остроту и апокалиптический накал.

Перечислю некоторые из этих опасений. Одно из них заключается в том, что ИИ — неважно, обладающий сознанием или нет, — вот-вот выйдет из-под контроля и восторжествует над человеческим разумом, поднявшись до вершин, находящихся за пределами нашего понимания и контроля. Это так называемая гипотеза сингулярности, популяризированная футурологом Рэем Курцвейлом и движимая необычайным ростом первичных вычислительных ресурсов за последние несколько десятилетий³. Где мы находимся на кривой этого роста? Проблема с кривыми экспоненциального роста, как многие успели убедиться за пандемию коронавируса, состоит в том, что из любой точки на этой кривой кажется, что дальше она резко уходит вверх, а позади стелется почти полого, поэтому узкий взгляд не позволяет определить, где мы сейчас. Второе опасение — это прометеев страх, что наши создания так или иначе ополчатся на нас самих, — страх, который уловили и скармливают нам же в новой яркой обертке фантастические фильмы и литература. И наконец, опасения множит то прискорбное обстоятельство, что термин «сознание» лепится применительно к способностям машин как и куда попало. Некоторые, в том числе и ряд исследователей ИИ, готовы считать сознательным все, что откликается на стимулы, чему-то обучается или ведет себя так, чтобы максимизировать вознаграждение или достичь цели. На мой взгляд, это не более чем нелепая натяжка, нарушающая логику «обладания сознанием».

Смешайте все эти представления, и сразу станет понятно, почему многим кажется, что до сознательного ИИ буквально рукой подать и что нам нужно сильно опасаться его появления. Полностью исключать эту вероятность нельзя. Если адепты сингулярности окажутся правы, то нам действительно пора будет заволноваться. Но с той точки, где мы находимся сейчас, такая перспектива представляется крайне маловероятной. Гораздо более вероятна ситуация, показанная на рис. 21. Здесь сознание не определяется интеллектом, а интеллект может существовать и без сознания⁴. Оба имеют множество разновидностей и множество разных измерений и граней выражения, то есть ни для сознания, ни для интеллекта не существует единой меры⁵.

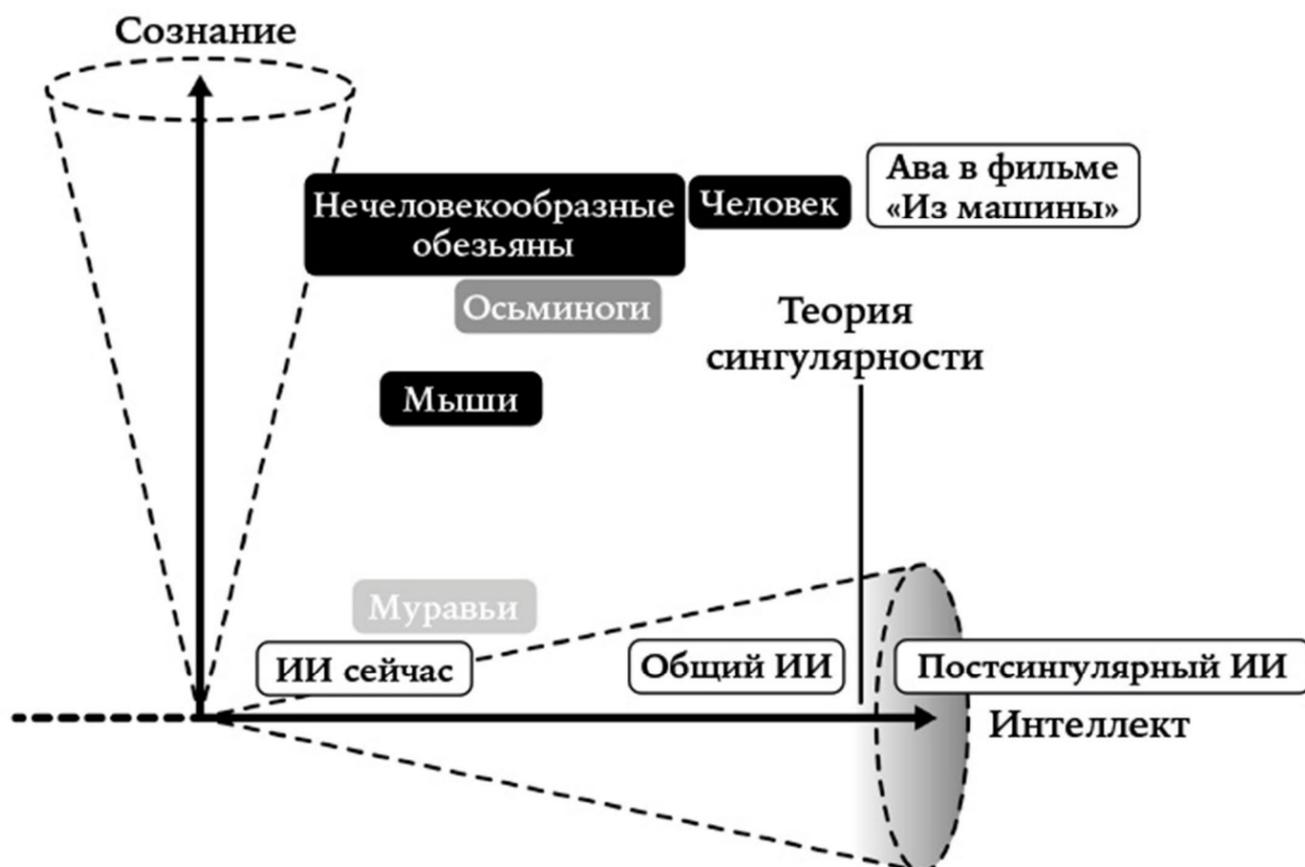


Рис. 21. Сознание и интеллект разделимы и многомерны, как можно судить по позициям животных и машин (действительных и вымышленных) на этой схеме. На этой схеме видно, что сейчас ИИ находится практически в самом начале оси интеллекта. Это потому, что пока неясно, считать ли имеющиеся системы ИИ разумными в каком бы то ни было значимом смысле. Большую часть современного ИИ правильнее всего характеризовать как высокоуровневое машинное распознавание

образов, возможно слегка приправленное планированием деятельности. Эти системы, неважно интеллектуальные или нет, работают без каких бы то ни было проявлений сознания.

Дальше на этой же оси располагается декларируемая многими исследователями ИИ амбициозная цель — разработать системы со способностями общего интеллекта как у человека, так называемый общий искусственный интеллект, или общий ИИ. А за этой вехой лежит terra incognita постсингулярного интеллекта. Но ни на каком этапе этого пути нельзя гарантированно утверждать, что сознание «просто приложится». Более того, на свете может быть много разновидностей интеллекта, отличающихся от человеческого и при этом дополняющих, а не заменяющих или усиливающих наш видоспецифичный когнитивный инструментарий, — и они тоже могут обходиться без сознания.

Возможно, выяснится, что определенные разновидности интеллекта без сознания все же не обходятся, но даже если так, это не значит, что сознания — после преодоления какого-то пока еще неизвестного порога — требуют все варианты интеллекта. И наоборот, может статься, что все сознательные существа хотя бы немного разумны, если определять интеллект достаточно широко. Но и в этом случае он не превращается в ковровую дорожку к сознанию.

Компьютеры не станут чувствующими только оттого, что поумнеют. Однако это не значит, что машинное сознание невозможно в принципе. Не попробовать ли нам встроить в них сознание с самого начала? Что (если интеллекту мы в этой роли отказываем) понадобится нам, чтобы создать сознательную машину?

Ответ на этот вопрос зависит от того, что вы считаете достаточным для признания системы сознательной, а это, в свою очередь, зависит от теории сознания, которой вы придерживаетесь. Поэтому такой большой разброс мнений по поводу необходимых условий для наличия сознания неудивителен.

На самом либеральном полюсе этого спектра находятся рассуждающие с функционалистских позиций, полагающие, что главное в сознании — правильная обработка информации. Она не обязательно тождественна «интеллекту», и тем не менее это обработка информации, а значит, ее можно внедрять в компьютерах. Например, согласно концепции, изложенной в журнале *Science* в 2017 г., машину можно назвать сознательной, если ее способ обработки информации подразумевает «глобальную доступность» этой информации и «самомониторинг» исполнения. Из статьи нельзя понять, действительно такая машина будет сознательной или просто будет вести себя так, будто обладает сознанием⁶, но в любом случае эта концепция строится на постулате, что для сознания достаточно правильного способа обработки информации.

Более веский довод в пользу сознательных машин выдвигают сторонники теории интегрированной информации (ТИИ). Как мы помним из главы 3, ТИИ утверждает, что сознание и есть интегрированная информация и что объем этой информации, который производит система, полностью определяется свойствами его внутренних механизмов — «причинно-следственной структурой». Согласно ТИИ, любая машина, создающая интегрированную информацию, из чего бы сама эта машина ни состояла и как бы ни выглядела, будет в какой-то степени обладать сознанием. Однако ТИИ не исключает вероятность, что машина будет казаться стороннему наблюдателю сознательной или разумной (или и такой, и другой одновременно), но ее механизмы не будут создавать никакой интегрированной информации, и поэтому о сознании тут даже речи не пойдет⁷.

Хотя ни та ни другая теория не отождествляет сознание с интеллектом, обе допускают, что машины, удовлетворяющие определенным условиям — правильному

способу обработки информации или ненулевой информационной интегрированности, — будут обладать сознанием. Но, конечно, чтобы согласиться с этими допущениями, нужно принимать и сами теории.

Теория животного-машины основывает восприятие нами мира и себя на биологическом стремлении к физиологической целостности — к тому, чтобы оставаться живым. Что говорит эта теория о вероятности появления сознательных машин?

Представьте себе робота из ближайшего будущего — с кремниевым мозгом и человекоподобным телом, оснащенного всевозможными датчиками и эффекторами (исполняющими элементами). Этот робот управляется искусственной нейросетью, разработанной по принципам прогнозной обработки и активного вывода. Поток сигналов, проходящих через его контуры, воплощает порождающую модель его среды и его собственного тела. Робот постоянно обращается к этой модели, выдвигая наиболее вероятные байесовские предположения об источниках и причинах своей входящей сенсорной информации. Эти синтетические контролируемые (и контролирующие) галлюцинации направлены по замыслу своему на то, чтобы поддерживать оптимальное функциональное состояние робота, то есть позволять ему оставаться, по его собственным представлениям, «живым». У него имеются даже искусственные interoцептивные входящие данные, сигнализирующие об уровне заряда его батарей и целостности его приводных механизмов и синтетических мышц. Ориентированные на контроль наиболее вероятные предположения об этих interoцептивных входящих данных генерируют синтетические эмоциональные состояния, которые направляют и мотивируют его поведение.

Этот робот действует независимо, делая то, что нужно, тогда, когда нужно, чтобы выполнить свои задачи. Тем самым внешне он производит впечатление разумного, чувствующего деятеля. Внутренние же его механизмы непосредственно сопоставимы с прогнозным аппаратом, который, как я предполагаю, находится в основе базового чувства телесного воплощения и ощущения себя собой у человека. То есть перед нами кремниевая машина-животное.

Обладает ли такой робот сознанием?

Неудовлетворительный, но честный ответ: наверняка я сказать не могу, но предполагаю, что нет. Из теории животного-машины следует, что сознание у человека и других животных развилось в ходе эволюции, возникает у каждого из нас в ходе онтогенеза (индивидуального развития особи) и в своей непрерывной работе тесно сопряжено с нашим статусом живой системы. Все наше восприятие и все наши ощущения продиктованы нашей природой автономных самоподдерживающихся живых машин, которым важно и дальше оставаться живыми. Я предполагаю (и снова это не более чем интуитивное предположение), что материальная сторона жизни окажется важна для всех проявлений сознания. Одно из оснований так полагать заключается в том, что побуждение к регуляции и самоподдержанию в живых системах не ограничено единственным уровнем, таким как сохранение целостности организма. Самоподдержание у живых систем пронизывает все уровни, вплоть до клеточного. Каждая клетка нашего тела (и какого угодно другого) постоянно воспроизводит условия, необходимые для дальнейшего поддержания ее целостности⁸. Этого нельзя сказать ни о каком компьютере — ни современном, ни ожидающемся в ближайшем будущем — и даже о тех кремниевых животных-машинах, которые я описывал выше.

Из этого не стоит делать вывод, будто я утверждаю, что клетки организма обладают сознанием или что сознание имеется у всех живых существ. Моя мысль состоит в том, что процессы физиологической регуляции, которые теория животного-машины

считает фундаментальными для сознания и ощущения себя собой, вырастают из основополагающих жизненных процессов, отмечаемых «на всех уровнях». С этой точки зрения искру сознания в уравнениях высекает жизнь, а не обработка информации.

Даже если до машин, действительно обладающих сознанием, еще далеко (если, собственно, они на самом деле возможны), нам все равно еще о многом нужно беспокоиться. Вполне вероятно, что в ближайшем будущем уровень развития ИИ и робототехники приведет к появлению новых технологий, которые будут казаться сознательными, даже если у нас не будет безоговорочных оснований считать их сознательными на самом деле.

В фильме Алекса Гарленда «Из машины» 2014 г. нелюдимый технический гений-миллиардер Нейтан приглашает крутого программиста Калеба в свою уединенную резиденцию и знакомит со своим творением — любознательным умным роботом Авой. Задача Калеба — выяснить, наделена ли Ава сознанием, или она просто умный автомат, не имеющий внутренней психической жизни.

В фильме активно используется тест Тьюринга⁹, знаменитый инструмент для выявления у машины мыслительных способностей. В одной яркой сцене Нейтан экзаменует Калеба на знание этого теста. В стандартной версии теста, как известно Калебу, оценивающий человек тестирует одновременно и претендующую на разумность машину, и другого человека удаленно, обмениваясь только печатными сообщениями. Для машины тест считается пройденным, когда оценивающий многократно затрудняется отличить ответ машины от ответа человека. Однако Нейтан придумал кое-что поинтереснее. «Мы посмотрим, — говорит он Калебу, — будете ли вы, даже зная, что она робот, воспринимать ее как сознательную личность».

Эта новая игра превращает тест Тьюринга из теста на интеллект в тест на сознание, а это, как мы теперь знаем, явления совершенно разные. Более того, Гарленд показывает нам, что тест в действительности направлен не на робота. Как объясняет Нейтан, главное не в том, машина Ава или нет. И даже не в том, обладает ли машина Ава сознанием. Главное — может ли Ава создать у обладателя сознания впечатление, что она тоже сознательна. Обсуждение этого момента Нейтаном и Калебом гениально, поскольку оно раскрывает, в чем на самом деле заключается тест: это проверка не для машины, а для человека. Это относится и к изначальному тесту Тьюринга, и к созданному Гарлендом в XXI в. аналогу, ориентированному на сознание. Диалог между героями фильма настолько элегантно и точно описывает проблему признания машины сознательной, что сейчас постепенно набирает популярность термин «тест Гарленда»¹⁰ — редкий пример ситуации, когда научная фантастика не только заимствует что-то у науки, но и отдает.

Многие простые компьютерные программы, в том числе разнообразные чат-боты, сейчас объявляются проходящими тест Тьюринга, поскольку им в достаточном числе случаев удается ввести в заблуждение относительно своей природы достаточное число судей. В одном особенно забавном примере, тоже 2014 г., 10 из 30 судей приняли чат-бот, общающийся от лица 13-летнего украинского мальчишки, за настоящего подростка-украинца. Начались громкие заявления, что мы наконец преодолели недостижимые прежде рубежи в развитии искусственного интеллекта¹¹. Но, разумеется, изобразить иностранного подростка, плохо владеющего английским, намного проще, чем успешно изобразить того, кого с вами роднят возраст, культура и язык, особенно когда допускается только обмен текстовыми сообщениями на расстоянии. Когда чат-боту засчитали прохождение теста, он написал: «Побеждая над тьюрингом [sic] я чувствую себя приятно»¹². Еще бы ему было не победить, если

планка настолько снижена. По факту это оказалось тестированием человека на доверчивость, и человек не справился¹³.

Но поскольку ИИ продолжает совершенствоваться, возможно, скоро тест Тьюринга будут проходить и без искусственного занижения стандартов. В мае 2020 г. научно-исследовательская лаборатория OpenAI выпустила GPT-3 — обширную искусственную нейросеть, обученную на примерах естественного языка, взятых из огромного массива текстов интернета¹⁴. GPT-3 может вступать в диалоги в формате чат-бота, а может создавать довольно большие куски текста в самых разных стилях, отталкиваясь от нескольких заданных слов или строк. И хотя сама она своих творений понять не в состоянии, они поражают — а кого-то даже пугают — гладкостью и уровнем сложности¹⁵. В одном примере, опубликованном в The Guardian, она написала эссе на 500 слов о том, почему людям не стоит опасаться ИИ, затронув целый ряд тем от психологии человеческого насилия до промышленного переворота и огорошив читателя вот таким пассажем: «ИИ не следует тратить время на попытки понять точку зрения людей, которые зарабатывают на жизнь неверием в искусственный интеллект»¹⁶.

Тем не менее я почти уверен, что достаточно подкованный собеседник-человек сумеет разоблачить и GPT-3 с ее высоким уровнем сложности текстов. Чего, возможно, уже не выйдет с грядущими версиями — GPT-4 или, скажем, GPT-10. Но даже если будущая GPT-подобная система будет устойчиво проходить тест Тьюринга, она охватит лишь один очень узкий пласт (симулированного) интеллекта — обезличенный языковой обмен, не воплощая полностью той самой способности естественного разума «делать то, что нужно, когда нужно», которую мы наблюдаем у человека и многих других животных, а также у моего гипотетического кремниевого животного-машины.

Что касается сознания, тут у нас нет аналога украинскому чат-боту, не говоря уже ни о каких GPT. Тест Гарленда остается непройденным. В действительности попытки создать симулякр чувствующего и осознающего свои чувства человека чаще вызывают у нас тревогу и отторжение, а не сложную смесь влечения, эмпатии и жалости, которую вызывала у Калеба Ава в фильме «Из машины».

Японский робототехник Хироси Исигуро уже не первое десятилетие конструирует роботов, придавая им максимально возможное сходство с человеком. Он называет их геминоиды. Исигуро создал геминоидные копии себя (см. рис. 22) и своей дочери (тогда шестилетней), а также геминоидную телеведущую со смешанными японско-европейскими чертами на основе обобщенного портрета около 30 разных людей. Каждый геминоид конструируется по подробным 3D-сканам тела и оснащается пневматическими приводами, способными породить широкий спектр мимики и жестов. Искусственного интеллекта как такового в этих устройствах почти нет — они подражают человеку лишь внешне и применение находят среди прочего в удаленном присутствии и «телеприсутствии». Одного из своих геминоидов Исигуро как-то раз посадил читать 45-минутную удаленную лекцию 150 студентам.

Впечатление эти геминоиды производят безусловно жутковатое. Они реалистичны, но недостаточно. Представьте себе встречу с геминоидом как прямую противоположность встрече с кошкой. При виде кошки, или, допустим, осьминога, у нас сразу возникает ощущение, что мы имеем дело с существом сознательным, пусть даже внешне совершенно не похожим на нас. При виде геминоида поразительное, но все же не стопроцентное физическое сходство только усиливает и подчеркивает расхождение и инаковость. В одном исследовании 2009 г. самым распространенным ощущением, которое вызывал геминоид у посетителей, был страх¹⁷.



Рис. 22. Хироси Исигуро и его геминоид

Публикуется с разрешения ATR Hiroshi Ishiguro Laboratories

Эта реакция — пример так называемого эффекта зловещей долины, понятия, введенного в 1970 г. другим японским исследователем, Масахиро Мори. Согласно гипотезе Мори, чем больше робот похож на человека, тем больше он нам нравится и тем больше мы ему сочувствуем (вспомним, например, С-3РО из «Звездных войн»), но, как только сходство достигает определенного предела, за которым робот в чем-то слишком похож на человека, а в чем-то недотягивает, симпатия резко сменяется отторжением и страхом (тут-то и образуется «зловещая долина») и возвращается, только когда сходство усиливается до полной тождественности и неотличимости. Хотя к единому мнению о том, почему возникает «зловещая долина», ученые не пришли, в существовании ее почти никто не сомневается¹⁸.

В отличие от роботов реального мира, которые пока не могут выбраться из «зловещей долины», виртуальные создания уже карабкаются на противоположный ее склон. Недавний прогресс в машинном обучении, достигнутый за счет использования «генеративно-состязательных нейросетей» (GANN), позволяет создавать фотореалистичные изображения несуществующих людей (см. рис. 23)^[52]. Эти фотографии получены путем «умного» объединения различных черт из больших баз данных настоящих лиц с применением почти таких же технологий, как те, что мы использовали в своей галлюцинаторной машине, описанной в главе 6. Осталось добавить сюда технологию «дипфейк»¹⁹, позволяющую анимировать эти лица, чтобы они «заговорили», затем дать им слово с помощью продолжающих совершенствоваться программ распознавания и производства речи, таких как GPT-3, — и вот наш мир уже населен виртуальными людьми, совершенно неотличимыми от виртуальных репрезентаций реальных людей. В этом мире мы отвыкнем отличать реальное от виртуального.



Рис. 23. Восемь лиц. Этих людей не существует, они ненастоящие

Все, кому кажется, что эти творения упрутся в своем развитии в потолок, так и не успев пройти дополненный видеосредствами тест Тьюринга, скорее всего, ошибаются. За подобным скепсисом стоит либо непрошибаемая убежденность в нашей (человеческой) исключительности, либо отсутствие воображения, либо и то и другое. Тест будет пройден. Соответственно, остаются два вопроса. Первый: сумеют ли эти новые виртуальные создания перейти в реальный мир, преодолев «зловещую долину», из которой никак не выберутся геминоиды Исигуро? Второй: покорится ли им тест Гарленда? Начнем ли мы воспринимать эти новые создания как не только разумные, но и обладающие сознанием, даже если будем знать, что это не более чем строки компьютерного кода?

И если да, как это отразится на нас самих?

Стремительное развитие ИИ, независимо от того, на какой горючей смеси ажиотажа и реального прогресса оно разгоняется, вынуждает нас снова и снова задумываться о его этической стороне. Немало опасений связано с социоэкономическими последствиями внедрения таких технологий ближайшего будущего, как самоуправляемые автомобили и замена работников предприятий автоматами, что неизбежно нарушит привычный нам уклад[53]. Нас закономерно тревожит делегирование принятия решений искусственным системам, внутренние процессы которых подвержены всевозможным искажениям и сбоям и при этом недостаточно прозрачны не только для пользователя, но и для разработчика. Самый острый вопрос в этой категории: какие ужасы нас ждут, если доверить искусственному интеллекту контроль над ядерным оружием или опорной сетью интернета?

Еще одна немаловажная этическая проблема — психологические и поведенческие последствия внедрения искусственного интеллекта и машинного обучения. Вторжение дипфейков в личное пространство, модификация поведения прогнозными алгоритмами, искажение убеждений в отфильтрованных информационных пузырях и герметичных камерах соцсетей — это лишь некоторые из множества сил, раздирающих ткань нашего общества. Потакая разгулу этих сил, мы добровольно уступаем свою идентичность и независимость безликим «дата-корпорациям» в ходе безграничного и бесконтрольного общемирового эксперимента²⁰.

На этом фоне обсуждение сознания у машин с точки зрения этики может казаться избыточным и слишком головоломным. Но это лишь видимость. Обсуждения необходимы, даже если у обсуждаемых машин сознания нет (пока). Когда тест Гарленда будет пройден, нам предстоит жить в одном мире с существами, которых мы будем воспринимать как обладающих субъективной внутренней жизнью, даже зная или веря, что она у них отсутствует. Психологические и поведенческие последствия этого соседства трудно спрогнозировать. Один из вероятных вариантов: мы научимся проводить границу между ощущениями и требуемыми поступками и

нам будет представляться вполне естественным любить и окружать заботой человека, но не робота, даже если обладателями сознания покажутся оба. Как это отразится на психологии каждого отдельного человека, пока неясно.

В телесериале «Мир Дикого Запада» человекоподобных роботов специально производят как «груши для битья» — чтобы человек вымещал на них свои самые низменные инстинкты, издевался, убивал и насиловал. Можно ли, истязая робота, чувствовать в нем сознание, понимать, что это сознание лишь иллюзия, и сохранить при этом рассудок? Для нашего нынешнего склада ума и психики это будет откровенной социопатией.

Другой вероятный вариант состоит в том, что наш привычный круг морально-нравственных интересов исказит антропоцентрическая склонность больше сочувствовать существам, с которыми мы ощущаем большее сходство. В этом случае представители следующего поколения наших близнецов-геминоидов, возможно, станут нам ближе и дороже, чем другие люди, не говоря уже о других животных.

Конечно, будущее не обязательно должно оказываться антиутопией. Но, учитывая набирающую обороты гонку прогресса и ажиотажа в области ИИ, замешанная на психологии этика тоже должна будет сказать свое слово. Запустить новую технологию и сложа руки смотреть, что будет дальше, попросту неправильно. Начнем с того, что нам не стоит слепо и бездумно добиваться стандартной цели ИИ — воспроизвести, а затем превзойти человеческий интеллект. Как справедливо заметил Дэниел Деннет, мы создаем «разумные инструменты, а не коллег»²¹, и об этой разнице нельзя забывать.

Перейдем теперь к варианту, при котором машины и вправду обретут сознание. Если мы действительно вольно или невольно внедрим в мир новые разновидности субъективного опыта, нам придется иметь дело с нравственно-этическим кризисом беспрецедентных масштабов. К статусу обладателя сознания непременно прилагается нравственный статус. Мы должны будем минимизировать потенциальные страдания сознательных машин точно так же, как мы должны минимизировать страдания живых существ, а нам и это не особенно удастся. А с этими гипотетическими искусственно чувствующими деятелями возникает дополнительная трудность: мы можем даже не догадываться, как именно они ощущают свою сознательность. Представьте себе систему, испытывающую совершенно неведомые нам страдания, для которых у нас, людей, нет ни аналогов, ни понятий, ни инстинктов, позволяющих их распознать. Представьте себе систему, для которой в принципе не существует такого явления, как разница между положительными и отрицательными ощущениями, у нее просто соответствующего феноменологического измерения нет. Сложность в данном случае заключается в том, что мы даже не подозреваем, какие этические проблемы могут в связи с этим возникнуть.

Даже если до появления искусственного сознания еще по-прежнему далеко, об этой перспективе все равно нужно задуматься уже сегодня. И хотя мы не знаем, чего потребует создание сознательной машины, мы не знаем и другого — чего оно не потребует.

Именно поэтому в июне 2019 г. немецкий философ Томас Метцингер призвал немедленно объявить 30-летний мораторий на все исследования, направленные на формирование «синтетической феноменологии», как он ее обозначил²². Я присутствовал при этом заявлении. Мы оба выступали на конференции, посвященной искусственному сознанию, которую устраивал в Кембридже Центр Леверхульма по изучению будущего интеллекта. Призыву Метцингера трудно было бы последовать буквально, поскольку под его действие подпадает едва ли не все компьютерное моделирование в психологии, но идея понятна. Мы не должны бездумно ломиться

вперед, пытаюсь создать искусственное сознание просто потому, что нам это кажется интересным, полезным или крутым. Самая лучшая этика — профилактическая.

В дни расцвета витализма рассуждения об этике искусственной жизни казались, наверное, такими же нелепыми и надуманными, какими сейчас могут показаться рассуждения об этике искусственного сознания. И тем не менее прошло всего сто с небольшим лет, и у нас имеется не только глубокое понимание основ жизни, но и множество новых инструментов для ее модификации и даже создания. Мы владеем технологиями генной инженерии, такими как CRISPR, позволяющими с легкостью менять последовательности ДНК и функцию генов. Мы научились разрабатывать полностью синтетические организмы «с нуля», с генетического уровня: в 2019 г. кембриджские исследователи создали вариант бактерии *Escherichia coli* с полностью синтезированным геномом. Вопросы этичности создания новых форм жизни приобретают неожиданную остроту²³.

И возможно, именно биотехнологии, а не ИИ удастся максимально приблизить нас к синтетическому сознанию. Здесь особый интерес представляет такая разработка, как «органойды головного мозга», — созданные из настоящих нейронов крошечные мозгоподобные структуры, которые выращиваются из человеческих плюрипотентных стволовых клеток (то есть тех, из которых способны развиться самые разные типы клеток организма). Эти органойды, хотя и не являются «мини-мозгом» в буквальном смысле, все же обладают сходством с развивающимся мозгом, что позволяет использовать их как лабораторные модели медицинских состояний, при которых развитие мозга нарушается. Могут ли органойды обладать примитивной разновидностью бестелесной осознанности? Исключать эту вероятность нельзя, тем более что у них зафиксированы согласованные волны электрической активности, напоминающие, как выяснилось в ходе недавнего исследования, аналогичные волны, наблюдаемые у недоношенных младенцев²⁴.

Органойды головного мозга состоят, в отличие от компьютеров, из того же материала, что и настоящий мозг, поэтому факторов, мешающих нам считать их потенциальными обладателями сознания, в данном случае на один меньше. Однако они все так же крайне просты, абсолютно бестелесны и совершенно не взаимодействуют с окружающим миром (хотя их можно подключать к камерам, автоматическим конечностям и тому подобным устройствам). На мой взгляд, даже если вероятность сознания у нынешних органойдов не особенно велика²⁵, по мере развития данной технологии эту вероятность будет все труднее сбрасывать со счетов. И тут мы возвращаемся к необходимости профилактической этики. Этические вопросы актуальны тут не только потому, что вероятность сознания у органойдов нельзя исключить полностью, но и в силу потенциального размаха самих разработок²⁶. Как выразился ведущий эти исследования ученый Алиссон Муотри, «мы хотим создавать целые фермы таких органойдов»²⁷.

Чем манит нас перспектива развития машинного сознания? Почему она так будоражит коллективное воображение? Я прихожу к выводу, что здесь включается какой-то техноэкстаз, подспудное желание выйти за рамки нашего ограниченного и низменного биологического существования в преддверии наступления последних дней. Возможность появления сознательных машин влечет за собой возможность перенести свой сознательный разум из дряхлеющей телесной оболочки в новехонькие схемы будущего суперкомпьютера, которому не грозят старость и смерть. Перенос сознания — излюбленный мотив футурологов и трансгуманистов, считающих, что одной жизни человеку мало²⁸.

Кому-то кажется даже, что этот рубеж мы уже преодолели. Оксфордский философ Ник Бостром в своем «Доказательстве симуляции» разбирает на статистической

модели предположение, что мы с большой долей вероятности уже не выступаем представителями изначального рода человеческого, а участвуем в хитроумной компьютерной симуляции, разработанной и воплощенной нашими технологически более подкованными потомками, одержимыми генеалогией. С этой точки зрения мы уже являем собой виртуальных чувствующих агентов виртуальной вселенной²⁹.

Кто-то в техноэкстазе видит стремительное наступление сингулярности, того самого переломного исторического момента, когда ИИ должен вырваться за пределы нашего понимания и контроля. В постсингулярном мире преобладать будут сознательные машины и симуляторы предков. Углеродная форма жизни, то есть мы, окажется в упадке, наши золотые дни будут позади.

Даже не обладая особым социологическим видением, можно догадаться, как кружит голову это зрелище нашей технологической элите, предстающей в нем главным двигателем беспрецедентного переворота, дарующего человечеству бессмертие. Вот что происходит, когда человеческая исключительность хлещет через край. В таком случае вся шумиха вокруг машинного сознания говорит о растущем желании отмежеваться от нашей биологической природы и эволюционного наследия.

Если смотреть с точки зрения концепции животного-машины, картина будет отличаться от нарисованной выше почти во всех отношениях. Как мы уже знаем, согласно моей теории, совокупность человеческого опыта и психической жизни возникает благодаря, а не вопреки нашей природе самоподдерживающихся биологических организмов, которым важно и дальше оставаться в живых. Это представление о сознании и человеческой природе не исключает вероятности появления сознательных машин, но гасит подогреваемый техноэкстазом ажиотаж вокруг грядущего «уже почти вот-вот» пришествия чувствительных компьютеров, который питает наши страхи и не дает спокойно спать по ночам. С точки зрения концепции животного-машины наши неустанные попытки разобраться в сути сознания все больше погружают нас в природу, а не отдаляют от нее.

Именно так и должно быть.

Эпилог

Хочу обрести контроль,
Я хочу идеальное тело,
Я хочу безупречную душу.

Radiohead. Creep (1992)

В январе 2019 г. я впервые увидел своими глазами живой человеческий мозг. Всего через 20 с лишним лет работы в области изучения сознания, спустя 10 лет существования нашей лаборатории в Сассексе и по прошествии трех лет после того, как я сам побывал в небытии под общим наркозом, — с рассказа об этом я начал книгу. Но глядя на едва заметно пульсирующую светло-серую поверхность коры, пронизанную сетью тонких темно-красных вен, я по-прежнему не мог представить, что этот сгусток способен рождать целую вселенную мыслей, чувств, ощущений — целую жизнь, полностью проживаемую от первого лица. Всецело проникнуться этим глубочайшим благоговением не давала некстати вспомнившаяся старая шутка о том, что пересадка мозга — это единственная операция, в которой лучше быть донором, чем реципиентом.

Возможность наконец увидеть мозг собственными глазами мне предоставил Майкл Картер, детский нейрохирург из Бристольской королевской детской больницы, находящейся на западе Англии. Майкл пригласил меня понаблюдать за одной из

самых драматичных нейрохирургических операций из всех возможных. Оперировать предстояло шестилетнего мальчика, который с рождения страдал тяжелой формой эпилепсии — источником приступов выступало правое полушарие коры, сильно поврежденное во время преждевременных родов. Поскольку никакие стандартные противосудорожные средства не помогали, оставалось одно — обратиться к нейрохирургии и провести гемисферотомию.

Суть этой операции — полное нейронное отключение вышедшего из строя правого полушария. Хирург проникает в мозг с правой стороны, проводит «резекцию» (удаление) височной доли, а затем перерезает все пучки связей — тракты белого вещества, которые соединяют правое полушарие с остальными частями мозга и организма. Изолированное полушарие остается в черепе, кровоснабжение у него сохраняется, и оно существует дальше как живой, но обособленный остров коры¹. Такая операция — экстремальный вариант более известной широкой публике операции по разделению полушарий, предполагающий, что полное нейронное отсоединение помешает электрическим бурям, зарождающимся в поврежденном правом полушарии, распространяться на остальные области мозга. Если прооперировать больного в юном возрасте, когда мозг еще достаточно пластичен, второе полушарие успеет взять на себя почти всю или даже всю нагрузку. Хотя операция довольно радикальна и каждый случай индивидуален, исход обычно бывает благоприятным.

Операция началась примерно в полдень и продолжалась чуть больше восьми часов. Я лично, когда работаю, отвлекаюсь едва ли не каждые пять минут — то письмо в почте посмотреть, то счет в крикете, то очередную чашку чая налить. Майкл с хирургом-практикантом и бригадой сменяющихся ассистентов трудились все это время без усталости, терпеливо, методично, не покладая рук. Примерно в середине процесса, когда практикант ушел на короткий перерыв, меня пригласили подойти к хирургическому микроскопу. Я не ожидал, что удостоюсь такой привилегии. Вглядываясь в ярко освещенные полости вскрытого детского мозга, я пытался совместить абстрактные знания о разных его участках и проводящих путях с лежащим передо мной комком тканей. Ничего не получалось. Я не мог разглядеть ни четкую иерархическую структуру коры, ни противотоки восходящих и нисходящих сигналов, которыми занимался в собственных исследованиях. Мозг снова стал непроницаемым, и мне оставалось только преклоняться и перед мастерством нейрохирурга, и перед материальной реальностью этого самого что ни на есть волшебного объекта. Ощущение было почти сверхъестественное. Таинственная завеса приоткрылась, явив на свет самое сокровенное. Я смотрел прямо в механизмы человеческого «я».

Операция прошла как планировалось. Где-то после восьми вечера Майкл оставил практиканта накладывать швы, а меня взял с собой на беседу с родственниками юного пациента. Они смотрели на нас с облегчением и благодарностью. Интересно, что бы они почувствовали, если бы увидели то, что несколькими часами ранее видел я?

Позже, возвращаясь по темной зимней дороге домой, я вновь вспомнил, как формулировал трудную проблему сознания Дэвид Чалмерс: «Мы сходимся во мнении, что у сознания имеется физическая основа, но не можем объяснить, почему и как сознание на этой основе возникает. Почему физические процессы в принципе должны порождать богатую внутреннюю психическую жизнь? Объективной логики в этом нет никакой, и тем не менее жизнь возникает»².

Сражаясь с этой загадкой, философия выдвинула ряд гипотез — от панпсихизма (сознание присутствует в той или иной степени везде) до элиминативного

материализма (сознания нет, по крайней мере в привычном нам понимании) и весь спектр промежуточных версий. Но наука о сознании — это не выбор из предложенного меню, каким бы фешенебельным ни был ресторан и каким виртуозом — его шеф-повар. Это, скорее, готовка из того, что нашлось в холодильнике, при которой, смешивая всё в новых и новых сочетаниях ингредиенты из философии, нейронауки, психологии, информатики, психиатрии, машинного обучения и так далее, мы получаем в итоге что-то отличное от прежнего.

В этом суть подхода к изучению сознания с точки зрения настоящей проблемы. Признать, что сознание существует, а затем спросить, как его феноменологические свойства (каким образом структурируется разный сознательный опыт, какую форму принимает и т.п.) соотносятся со свойствами мозга, являющегося частью телесного организма и встроенного в окружающий мир. Поиск ответов на эти вопросы можно начать с идентификации соответствия тех или иных паттернов активности мозга тем или иным разновидностям сознательного опыта, но это не значит, что все сведется и должно свестись только к этому. Необходимо выстраивать все более надежные и прочные объяснительные мосты между механизмами и феноменологией, чтобы взаимосвязи, которые мы вычерчиваем, обретали логику, а не выглядели случайными. Какую же логику они должны обрести в данном контексте? Все ту же — объяснение, прогнозирование, контроль.

В исторической перспективе эта стратегия — отражение того, как научное понимание законов жизни одерживает верх над магическим мышлением витализма, вычлняя свойства живых систем, а затем объясняя каждое из них с точки зрения находящихся в их основе механизмов. Конечно, жизнь и сознание — это разные явления, хотя, надеюсь, теперь-то я уже успел убедить вас, что они связаны теснее, чем может показаться на первый взгляд. Стратегия и для того и для другого одна. Подход с точки зрения настоящей проблемы, не предполагая ни попыток решить трудную проблему сознания с наскока, ни задвинуть на второй план эмпирические характеристики сознания, вселяет подлинную надежду примирить физическое с феноменальным и в результате не решить трудную проблему, а развенчать.

Мы начали свой путь с рассмотрения уровня сознания — разницы между пребыванием в коме и осознанным бодрствованием, — делая основной акцент на важности измерений. Главный вывод, к которому мы на этом этапе приходим: предлагаемые для измерения параметры, такие как каузальная плотность, информативность и интегративность, взяты не наобум. Они выражают крайне консервативные свойства любого сознательного опыта, указывая, в частности, на то, что любое сознательное впечатление едино и отлично от любых других сознательных впечатлений. Любой осознанный опыт переживается «как целое», и любое такое переживание имеет свой собственный характер.

Затем мы перешли к обсуждению природы содержания сознания, а именно к ощущению себя сознательным «я». Я перечислил ряд задач, побуждающих нас пересмотреть картину, которую нам рисует неискушенный интуитивный взгляд, и принять новую, посткоперниканскую точку зрения на сознательное восприятие.

Первая задача — рассмотреть восприятие как активный, конструктивный процесс, ориентированный на действие, а не пассивное отражение объективно существующей внешней действительности. Воспринимаемый мир одновременно меньше и больше того, что может представлять собой эта объективная внешняя действительность. Наш мозг выстраивает окружающий мир в байесовском процессе наиболее вероятного предположения, в котором сенсорные сигналы служат главным образом для того, чтобы удерживать в рамках наши постоянно развивающиеся перцептивные гипотезы. Мы живем в контролируемой галлюцинации, которую эволюция создавала не ради точности, а ради пользы.

Вторая задача обращает это открытие внутрь, к ощущению бытия собой. Мы выяснили, что наше «я» — это тоже восприятие, еще одна разновидность контролируемой галлюцинации. Составляющие нашего «я» — от ощущения личностной идентичности и непрерывного существования во времени до зачаточного ощущения бытия живым организмом — определяются все той же перекличкой направленных изнутри наружу перцептивных прогнозов и устремленных извне внутрь ошибок прогнозирования, хотя теперь эта перекличка происходит большей частью не за пределами тела, а внутри.

Последняя задача — увидеть, что истоки и первоочередная функция прогнозных механизмов сознательного восприятия заключаются не в том, чтобы представлять мир или тело, а в том, чтобы контролировать и регулировать наше физиологическое состояние. Совокупность нашего восприятия и когнитивных процессов — вся панорама человеческого чувственного опыта и психической деятельности — формируется глубинным биологическим побуждением оставаться в живых. Мы воспринимаем мир вокруг и себя в нем при участии, посредством и благодаря нашему живому организму.

Такова моя теория животного-машины — современная версия (или инверсия) концепции *l'homme machine* («человека-машины») Жюльена Офре де Ламетри. И именно здесь намечаются самые глубокие сдвиги в наших представлениях о сознании и «я».

Если раньше нам было непонятно, почему ощущение «бытия собой» так отличается от ощущения мира вокруг, то теперь мы можем рассматривать их как разные выражения одного и того же принципа перцептивного прогнозирования, а различия в феноменологии возводить к различиям в участвующих разновидностях прогнозов. Какие-то перцептивные умозаключения направлены на выяснение того, что касается объектов внешнего мира, какие-то — на контроль происходящего внутри тела.

Увязывая психическую деятельность с физиологической действительностью, мы наполняем давние представления о преемственности жизни и разума новым смыслом, подводя под них прочные опоры концепций прогнозной обработки и принципа свободной энергии. Эта глубинная преемственность, в свою очередь, сближает нас с другими животными и остальной природой и, соответственно, отдаляет от бестелесных вычислительных процессов искусственного интеллекта. Таким образом, сознание и жизнь притягиваются друг к другу, а сознание и разум, наоборот, отталкиваются. Этот пересмотр нашего места в природе касается не только нашего физического тела, биологического организма, но и нашего сознательного разума, восприятия окружающего мира и бытия теми, кто мы есть.

Лишая человека статуса центра мироздания или венца творения, наука каждый раз с лихвой компенсировала ему эту утрату. Коперниканская революция подарила нам вселенную, которую астрономические открытия последнего столетия раздвинули далеко за пределы человеческого воображения. Разработанная Чарльзом Дарвином теория эволюции путем естественного отбора подарила нам семью, связь с прочими живыми видами, осознание давности своей истории и мощи эволюционных законов и механизмов. И вот теперь наука о сознании, одним из направлений которой выступает теория животного-машины, пробивает брешь в последнем бастионе человеческой исключительности — привычке считать наш сознательный разум особенным — и показывает, что и сознание накрепко вписано в общие схемы природы.

В сознательном опыте все в каком-то смысле представляет собой восприятие, а любое восприятие — разновидность контролируемой (или контролирующей) галлюцинации. Больше всего меня в этой концепции восхищают перспективы,

которые она перед нами рисует. Ощущение свободы воли — это восприятие. Ход времени — тоже восприятие. Возможно, даже трехмерные структуры ощущаемого нами мира и чувство объективной реальности содержания перцептивного опыта окажутся составляющими восприятия³. Инструменты науки о сознании позволяют нам как никогда близко подойти к кантовскому ноумену («вещи в себе»), предельному выражению непознаваемого, частью которого являемся и мы. Все эти идеи поддаются экспериментальной проверке, и, независимо от того, какие результаты эта проверка принесет, даже просто задаваясь подобными вопросами, мы уже переворачиваем свои представления о том, что такое сознание, как оно возникает и зачем существует. Каждый шаг — это еще один удар по заманчивому, но бесполезному интуитивному представлению о сознании как об одной огромной страшной тайне, ждущей одной огромной страшной разгадки.

У этого подхода есть и прикладное применение. Теоретически обоснованное измерение уровня сознания ведет к появлению новых измерительных приборов, которым все лучше удается улавливать остаточную осознанность — «скрытое сознание» — при отсутствии указывающих на него поведенческих признаков. Вычислительные модели прогнозного восприятия раскрывают механизмы, находящиеся в основе галлюцинаций и бредовых иллюзий, совершая тем самым переворот на самых разных ступенях психиатрии от лечения симптомов до поиска причин. Всевозможные новые направления открываются и для развития искусственного интеллекта, нейрокомпьютерных интерфейсов и виртуальной реальности в дополнение к изобилию уже сложившихся и только зарождающихся технологий. Исследование биологических основ сознания оказывается на удивление полезным делом.

Ко всему этому нужно добавить, что исследование загадки осознанности было и всегда будет одиссеей глубоко личной. Что толку от науки о сознании, если она не может пролить свет на индивидуальные особенности нашей собственной психической деятельности и внутренний мир тех, кто нас окружает?

Вот она, подлинная перспектива настоящей проблемы. Одолевая этот путь, куда бы он ни привел нас в конце концов, мы немало узнаем и поймем о сознательном восприятии мира вокруг и о нас самих в нем. Мы увидим, что наша внутренняя вселенная не исключение из остальной природы, она целиком и полностью в нее включена. И даже если мы задумываемся об этом не так часто, как могли бы, у нас все равно будет возможность заново примириться с тем, что происходит — или не происходит, — когда контролируемая галлюцинация бытия собой в конце концов рассыпается в прах. Когда небытие — это не пресечение потока сознания, вызванное общим наркозом, а возвращение в вечность, из которой каждый из нас когда-то возник.

И вот там, в конце этой истории, где жизнь от первого лица приходит к своему завершению, чуточку тайны, наверное, неплохо было бы оставить.

Благодарности

Изложенные в этой книге идеи формировались в бесчисленных обсуждениях с друзьями, коллегами, студентами, учителями и наставниками на протяжении двух с лишним десятилетий. Я благодарен всем моим собеседникам.

Спасибо всем бывшим и нынешним участникам моей научно-исследовательской группы в Сассексе. Работать с вами было и будет честью для меня. Отдельная благодарность тем участникам группы, к чьим исследованиям я обращаюсь в этой книге, — Адаму Барретту, Лайонелу Барнетту, Питеру Лашу, Альберто Мариоле,

Яиру Пинто, Уоррику Роузбуму, Майклу Шартнеру, Дэвиду Шварцману, Максин Шерман, Кейсуке Судзуки и Александру Чанцу.

Благодарю я и вас: Мануэль Балтиери, Рени Байкова, Люк Бертуз, Дэниел Бор, Крис Бакли, Эйсер Чан, Пол Чорли, Рон Крайсли, Энди Кларк, Марианна Коул, Клеманс Компэн, Гийом Корлуэр, Хьюго Критчли, Золтан Дьенеш, Том Фроуз, Пол Грэм, Инман Харви, Оуэн Холланд, Риота Канаи, Томаш Корбак, Изабель Мараньян, Федерико Мичели, Берен Миллидж, Томас Новотны, Энди Филиппидес, Шарлотта Рэй, Колин Ривли, Райан Скотт, Лина Скора, Надин Спыхала, Марта Суарес-Пинилья, Крис Торнтон, Хао Тин Ван и Джейми Уорд, — всех друзей и коллег, всех работавших в Сассексе прежде и сейчас, своими трудами и идеями немало способствовавших развитию моих трудов и идей, а также любезно согласившихся прочитать те или иные части этой книги и высказать свои замечания.

Я благодарен выдающимся ученым, которые брали меня под крыло на разных этапах моего научного пути. Это и скончавшийся в 2015 г. Николас Макинтош, мой наставник в студенческие времена в Кембридже, уверивший меня, что я должен идти в науку. Это Фил Хазбандс, мой научный руководитель в докторантуре, предоставивший мне полную свободу в исследованиях, но внимательно следивший за тем, чтобы я не сбился с курса. Это Джеральд Эдельман (его не стало в 2014 г.), под руководством которого я шесть с лишним лет трудился постдокторантом. При нем мой интерес к сознанию наконец оформился в основную тему исследований. Кроме того, я хотел бы поблагодарить Маргарет Боден, Энди Кларка, Дэниела Деннета и Томаса Метцингера, которые на протяжении многих лет вдохновляли меня и служили ориентирами.

Искренне благодарю тех, кто взял на себя труд прочитать и подробно прокомментировать первые черновики книги, — это Тим Бейн, Энди Кларк (очередное мое спасибо ему), Клаудиа Фишер, Якоб Хохви и Мюррей Шанахан. Они очень помогли мне своими идеями и предложениями. Кроме того, я бесконечно благодарен Карлу Фристон, Марчелло Массимини, Томасу Метцингеру (в который раз), Анируддху (Ани) Пателу и Адриану Оуэну, давшим мне массу советов по отдельным частям книги. Огромное спасибо Стиву Уэсту (студия Lazy Chief) за стильные иллюстрации. Бабе Бринкману спасибо за неиссякаемый источник вдохновения, которым служит мне его словесный дар. Работать с ним над его «Рэперским руководством по сознанию» (Rap Guide to Consciousness) было потрясающе. Спасибо Майклу Картеру за приглашение в операционную и возможность увидеть гемисферотомию. Спасибо Дэвиду Эдельману и Грациано Фиорито за знакомство с миром осьминогов, Иану Чени и Лори Сантос за погружение в мир обезьян, а Луизе Шрайтер — за погружение в совершенно иной мир.

Я не могу перечислить поименно огромное множество тех, кто все эти годы помогал мне советами, идеями и поддержкой, поэтому назову лишь некоторых. Это Анил Анантхасвами, Крис Андерсон, Бернард Баарс, Лиза Фельдман Барретт, Изабель Бенке, Тристан Бекинштайн, Йошуа Бенджио, Хезер Берлин, Мэтт Бергман, Дэвид Биелло, Робин Кархарт-Харрис, Оливия Картер, Дэвид Чалмерс, Крейг Чапмен, Аксель Климманс, Афина Демертзи, Стив Флеминг, Зафериос Фунтас, Фрайди Футбол, Крис Фрит, Ута Фрит, Алекс Гарленд, Мариана Гарза, Мел Гудейл, Аннака Харрис, Сэм Харрис, Ник Хамфри, Роб Айллиф, Робин Инс, Джон Иверсен, Евгений Ижикевич, Алексис Йохансен, Роберт Кентридж, Кристоф Кох, Сид Куидер, Джефф Кричмар, Виктор Ламм, Хакван Лау, Стивен Лорис, Рафаэль Малах, Дааниш Масуд, Саймон Макгрегор, Педро Медиано, Лусия Меллони, Лиад Мудрик, Фил Ньюман, Ангус Нисбет, the Parasites, Меган Питерс, Джованни Пеццуло, Тони Прескотт, Блейк Ричардс, Фернандо Росас, Адам Резерфорд, Тим Саттертуэйт, Том Смит, Нараянан Шринивасан, Кэтрин Таллон-Бодри, Джулио Тонони, Нао Цутия, Ник Терк-Браун,

Лусина Уддин, Симон ван Гааль, Бруно ван Свиндерен, Анниек Верхолт, Пол Вершур, Люси Уокер, Найджел Уорбертон, Лиза Уэстбери и Мартин Вокке.

Эти люди годами каждый по-своему наставляли меня на верный путь. Это не значит, что все они согласны с моими утверждениями. Любые оставшиеся в книге ошибки или несуразности исключительно на моей совести.

Я не смог бы написать эту книгу без поддержки фонда Wellcome Trust (при посредничестве Engagement Fellowship), Фонда доктора Мортимера и Терезы Саклер, а также Канадского института перспективных исследований (где я выступаю одним из руководителей программы по изучению мозга, разума и сознания). Я благодарен им за помощь — и мне лично, и в тех исследованиях, которые выполняет моя группа. Отдельное спасибо Сассексскому университету, который на долгие годы стал мне вторым, академическим, домом.

И в завершение — благодарности, которые я приберег для моего агента и редакторов. Уилл Фрэнсис из агентства Janklow & Nesbit надоумил меня написать эту книгу, помог мне сформулировать заявку, нашел замечательных редакторов и курировал весь проект от начала до конца. Такую же благодарность я хочу выразить великолепным редакторам — Роуэн Коуп и Лоре Хассан из Faber & Faber, Стивену Морроу из Dutton/Penguin и редактору-корректору Элинор Риз.

Примечания

Пролог

1. Barnes (2008).

Глава 1. Настоящая проблема

1. Baars (1988); Dehaene & Changeux (2011); Mashour et al. (2020); Shanahan (2010).

2. Теорию глобального рабочего пространства можно представить и как теорию «сознания доступа» в противопоставление «феноменальному сознанию». Если феноменальное сознание обращено к чувственному опыту, то для сознания доступа первостепенную роль играют когнитивные функции, которые оно ставит выше феноменологии. Характеризуя то или иное психическое состояние как «когнитивно доступное», мы имеем в виду, что это психическое состояние доступно для любых когнитивных функций, включая рассуждения, принятие решений и контроль поведения. См.: Block (2005).

3. У теории сознания высшего порядка существуют разнообразные вариации. См.: R. Brown et al. (2019); Fleming (2020); Lau & Rosenthal (2011).

4. Portin (2009).

5. Chalmers (1995a), p. 201.

6. Позже Чалмерс ввел понятие «метапроблемы» сознания — почему мы считаем, что существует трудная проблема сознания (Chalmers, 2018). Метапроблема — это, по сути, вариант легкой проблемы, поскольку она касается объяснения поведения (в данном случае вербального поведения тех, кто говорит о трудной проблеме сознания). Что мне нравится в метапроблеме — ее можно признать как проблему и изучать независимо от того, какую метафизическую позицию мы занимаем по отношению к сознанию как таковому.

7. Craver & Tabery (2017).

8. Chalmers (1995a), p. 203, курсив авторский (Чалмерса).

9. Разница между физикализмом и материализмом в основном историческая. Материализм — термин более старый. Кроме того, некоторые утверждают, что

физикализм — это более узкая «лингвистическая концепция» (согласно которой у любого лингвистического утверждения есть эквивалентное ему физическое состояние), а материализм — положение более общее, касающееся природы вещей. См.: Stoljar (2017).

10. Хотя большинство функционалистов придерживается физикалистских взглядов, функционалистом вполне можно быть и не примыкая к физикалистам.

11. Мэтью Кобб в своей захватывающей книге «Мозг. Биография» (М.: Бомбора, 2022) рассказывает, как функции мозга истолковывались на основании технологии, господствующей в данный период (впрочем, бывало и наоборот — технологии сравнивали с функциями мозга).

12. Silver et al. (2017). История оригинальной программы AlphaGo великолепно изложена в одноименном фильме: <https://www.alphagomovie.com/>. Кто-то, впрочем, может возразить, что эти программы точнее будет называть игрой в «историю го», чем непосредственно в го.

13. Усложненным вариантом этого тезиса выступает мысленный эксперимент Джона Сёрла «Китайская комната». Однако здесь я этот пример не привожу, поскольку аргумент Сёрла касается преимущественно интеллекта (или «понимания»), а не сознания (Searle, 1980).

14. Как сказал философ Джон Перри, «если слишком долго размышлять о сознании, либо станешь панпсихистом, либо подашься в чиновники». Возможно, так бывает, когда в науке слишком много размышлений и мало дела. Внятное обоснование панпсихизма рекомендую искать в книге Филиппа Гоффа «Ошибка Галилея» (Philip Goff, Galileo's Error) (2019). Высказывание Перри цитируется в статье Оливии Голдхилл 2018 г. на портале Quartz: [qz.com/1184574/the-idea-that-everything-from-spoons-to-stones-are-conscious-is-gaining-academic-credibility](https://www.quartz.com/1184574/the-idea-that-everything-from-spoons-to-stones-are-conscious-is-gaining-academic-credibility).

15. McGinn (1989).

16. Доводы против предположения, что какие-то вещи человечество не постигнет никогда, см.: Deutsch (2012).

17. Есть еще один «-изм», заслуживающий упоминания, но уступающий в формализованности теориям, приведенным в тексте. Это иллюзионизм, гласящий, что (феноменальное) сознание представляет собой интроспективную иллюзию, то есть когда мы осмысливаем состояния своего сознания, мы ошибочно приписываем им феноменальные свойства — квалиа, которыми они на самом деле не обладают. Одна из версий иллюзионизма, с которой я не согласен, утверждает, что состояний сознания в действительности не существует. Другая, которая импонирует мне больше, говорит, что сознательный опыт существует, но не такой, каким он нам представляется. Вполне возможно, хотя наверняка утверждать трудно, что эта версия иллюзионизма вполне совместима с тем, о чем я буду говорить в своей книге. Дополнительно об иллюзионизме см.: Frankish (2017).

18. Существуют по крайней мере два типа философских зомби. «Поведенческие зомби» неотличимы внешне с точки зрения поведения от своих сознательных аналогов. У «неврологических зомби» добавляется еще одна гипотетическая характеристика: они идентичны сознательному существу не только внешне, но и внутренне. Неврологический зомби обладает таким же внутренним устройством и может быть построен на той же электрохимической «биоаппаратуре», что и его сознательный аналог. При этом любые разновидности зомби полностью лишены сознания.

19. Gidon et al. (2020); Herculano-Houzel (2009).

20. Сторонники теории зомби могут ответить, что главное — логическая вероятность, а не представимость с точки зрения законов физики в данной конкретной вселенной.

Я с этим не согласен. Летящий хвостом вперед Airbus A380 логически вполне вероятен, если допустить действие альтернативных законов аэродинамики, но это допущение не поможет нам понять, как на самом деле летает настоящий Airbus A380 в реальном мире, подчиняющемся существующим законам физики и аэродинамики. Вот и меня интересует, как настоящий мозг (организм и прочее), подчиняющийся законам физики нашей вселенной, формирует и порождает сознательный опыт в существующем на самом деле мире.

21. Взгляд через призму «настоящей проблемы» нельзя назвать новым — по крайней мере абсолютно новым. Схожие стратегии — в виде «проблемы соответствия» (Chalmers, 1996) и «принципа структурной связности» (Chalmers, 1995a) — описывал сам Чалмерс, включая их в свое изначальное изложение трудной проблемы. Кроме того, существует массивный и весьма влиятельный корпус трудов в области «нейрофеноменологии», в которых делаются попытки сопоставить феноменологические свойства с различными аспектами мозга и его активности (Thompson, 2014; Varela, 1996). Однако акценты в этих научных позициях расставлены по-разному (Seth, 2009, 2016b).

22. Sutherland (1989).

23. Crick & Koch (1990). Примерно в это же время вышел влиятельный труд американского философа Дэниела Деннета «Объясненное сознание» (Consciousness Explained, 1991) — знакомство с этой книгой в начале 1990-х гг. стало для меня поворотным моментом. Сейчас, 30 лет спустя, она по-прежнему не только просвещает, но и взрывает мозг. Дополнительно об истории науки о сознании см.: LeDoux et al. (2020) и Seth (2017, 2018).

24. Crick & Koch (1990).

25. Хотя обычно считается, что НКС соотносятся с определенными областями мозга, это не всегда и не обязательно так. Определение НКС подразумевает нейронный механизм, который может действовать в целом ряде областей мозга. В некоторых случаях сети мозга, участвующие в том или ином НКС, могут даже измениться со временем (G. M. Edelman & Gally, 2001).

26. Нюанс здесь в том, что области мозга, выявляющиеся в исследованиях бинокулярного соперничества, часто связаны с переключением между сознательными перцепциями, а это важно, поскольку восприятие изменений — это совсем не то же самое, что изменение восприятия. См.: Blake et al. (2014). К этому отличию мы еще вернемся в главе 6.

27. Эти проблемы всплывали в обсуждениях из года в год, но обычно от них либо отмахивались, считая не относящимися к делу, либо заматали их под ковер. Сформулировать их с необходимой четкостью удалось только в 2012 г. в двух статьях: Aru et al. (2012), de Graaf et al. (2012).

28. В ряде экспериментов предпринимались храбрые попытки отделить сознательное восприятие от внимания и от вербального отчета. Особенно любопытны результаты так называемых безотчетных экспериментов, участники которых не сообщают о том, что воспринимают. Во многих из этих исследований выявляющиеся в конечном итоге нейронные корреляты сознания не включают в себя лобные доли, которым отводится центральная роль в теории глобального рабочего пространства, теории сознания высшего порядка и им подобных. Недавнее изложение этой парадигмы см.: Frässle et al. (2014); Tsuchiya et al. (2015) и Rassah et al. (2021).

29. Еще одну трактовку этого замысла мы находим у философов Сьюзан Херли и Альвы Ноэ, которые делят объяснительные разрывы на «сравнительные» (связанные с объяснением, почему тому или иному чувственному опыту присущи именно те феноменологические свойства, которые ему присущи) и «абсолютные» (трудную проблему объяснения, почему и каким образом в принципе существует

феноменология). Соответственно настоящую проблему можно расценивать как стремление закрыть сравнительные объяснительные разрывы, чтобы устранить (или упразднить) абсолютный объяснительный разрыв. См.: Hurley & Noë (2003).

30. Витализм утверждает, что «живые организмы фундаментально отличаются от неживых созданий, поскольку содержат некий нефизический элемент или управляются иными принципами, нежели неодушевленные предметы» (Bechtel & Williamson, 1998). Даже сегодня большинству детей дошкольного возраста виталистические объяснения особенностей жизни кажутся понятнее и предпочтительнее любых других, более современных (Inagaki & Hatano, 2004). Исторические параллели между витализмом и наукой сознания особенно активно исследовала философ Патриция Черчленд (Churchland, 1996).

31. Возможно, именно жадой триумфального единого решения отчасти объясняется неослабевающая популярность теорий сознания, основанных на квантовой механике, большинство которых восходит к вышедшей в 1989 г. работе Роджера Пенроуза «Новый ум короля» (М.: URSS, 2022). И хотя не исключена вероятность, что какая-нибудь будущая квантовая теория все-таки скажет что-нибудь полезное о сознании, пока все попытки, на мой взгляд, диктуются ошибочной логикой: «квантовая механика загадочна, сознание тоже загадочно, значит, они должны быть связаны между собой».

Глава 2. Измерить сознание

1. www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/dalence-joachim.

2. Н. Chang (2004).

3. Судя по всему, первым этот вопрос задал — и ответил на него — Томас Генри Гексли (Бульдог Дарвина). В 1870 г., читая лекцию, он сказал: «Я верю, что рано или поздно мы получим механический эквивалент сознания, точно так же, как получили механический эквивалент тепла». Цит. по: Cobb (2020), p. 113.

4. Seth et al. (2008).

5. Weiser et al. (2008).

6. Myles et al. (2004).

7. Nasraway et al. (2002).

8. Herculano-Houzel (2009).

9. См.: Lemon & Edgley (2010). Мозжечок нередко показывает себя в экспериментах по нейровизуализации «областью, заслуживающей внимания», однако эти результаты на обсуждение не выносятся, поскольку изначально мозжечок не заявляли объектом эксперимента и исследователи не знают, что о нем сказать.

10. DiNuzzo & Nedergaard (2017).

11. Ferrarelli et al. (2010); Massimini et al. (2005).

12. Чтобы создать сжатую версию последовательности из одних единиц (или одних нулей) длиной n , достаточно обозначить начальную единицу (или ноль), а затем указать n повторений. Ее абсолютную противоположность, полностью случайную последовательность, не удастся сжать в принципе: чтобы воспроизвести ее в точности, придется указывать место каждой единицы и каждого нуля.

Последовательности, структура которых хотя бы немного предсказуема, займут промежуточное положение между этими крайностями. Длина оптимально сжатой версии называется колмогоровской сложностью (или сложностью Колмогорова — Хайтина). Сложность ЛЗВ аппроксимирует эту величину и обозначает ее верхний предел.

13. Casali et al. (2013).

14. Schartner et al. (2017b); Schartner et al. (2015). В ходе этих исследований мы убедились, что изменения в сложности ЛЗВ — это не просто банальные отражения других (подробнее изученных) изменений в электроэнцефалографических сигналах мозга, сопровождающих потерю сознания, таких как возрастание мощности низкочастотных дельта-волн, наблюдающееся во время сна.
15. Casarotto et al. (2016).
16. См., например, широко применяющуюся шкалу комы Глазго (Teasdale & Murray, 2000).
17. Синдром «запертого человека» чаще всего наступает из-за повреждения варолиева моста — той части ствола мозга, которая содержит важные проводящие нейронные пути, связывающие спинной мозг с другими областями головного мозга.
18. Vauby (1997). Казалось бы, участь «запертого» — хуже смерти, однако не все страдающие данным синдромом находят свое существование невыносимым: процент «запертых», считающих свою жизнь вполне сносной, неожиданно высок (Rousseau et al., 2015). Очередное подтверждение тому, как опасно по внешним проявлениям судить о происходящем внутри.
19. С этой историей можно ознакомиться здесь: www.humanbrainproject.eu/en/follow-hbp/news/measure-of-consciousness-and-a-tale-of-cultural-loss.
20. Owen et al. (2006). Историю открытия, сделанного его группой, и его значение для медицины и исследований сознания Адриан Оуэн излагает в своей книге «В серой зоне» (М.: АСТ, 2020).
21. Monti et al. (2010). Поскольку в подобных исследованиях пациенты могут отвечать только «да» или «нет», нужно быть очень осмотрительным в подборе вопросов. Спросить, больно ли пациенту, вполне логично и допустимо, а вот насколько допустимо спрашивать, хочет ли он жить дальше? Этических дилемм тут возникает предостаточно. В 2016 г. мы с писательницей Линдой Маршалл-Гриффитс и продюсером Надией Молилари подняли эту тему в аудиоспектакле «Небеса — пространство» (The Sky is Wider). <https://www.lindamarshallgriffiths.co.uk/the-sky-is-wider-best-single-drama>
22. Naci et al. (2017).
23. Более известный пример этой проблемы — попытка определить и измерить интеллект по единой шкале, такой как IQ. Эти попытки стабильно проваливаются при столкновении с огромным разнообразием проявлений интеллекта у отдельных людей, культур и видов. Многомерный подход к характеристике уровня сознания см.: Wayne et al. (2016).
24. Альберт Хоффман, 'LSD — My Problem Child', <http://psychedelic-library.org/child1.htm>.
25. Подробнее об этой истории см.: «Мир иной» Майкла Поллана (М.: АСТ, 2020).
26. Carhart-Harris et al. (2012).
27. Обычно МЭГ и ЭЭГ обеспечивают очень высокое временное разрешение, измеряющееся миллисекундами, поскольку эти виды мониторинга отражают непосредственную электрическую активность групп нейронов. фМРТ по сравнению с ними слишком медлительна, ее разрешение измеряется секундами. Отчасти это объясняется медленной работой аппарата МРТ, позволяющего делать снимок лишь раз в одну-две секунды, а отчасти медленным поступлением измеряемого фМРТ сигнала о насыщении крови кислородом.
28. Schartner et al. (2017a). Еще одно, более позднее, исследование выявило аналогичную картину при приеме диметилтриптамина (Timmermann et al., 2019).

29. В последующем исследовании под руководством Лайонела Барнетта мы обнаружили, что психоделическое состояние характеризуется существенным уменьшением «информационного сообщения» между областями коры. Это в очередной раз говорит о том, что в психоделическом состоянии утрачивается перцептивная структура. См.: Barnett et al. (2020).

30. Tononi & Edelman (1998). Эдельман умер в 2014 г., я написал тогда один из множества некрологов, отдавая дань его огромному вкладу в науку (Seth, 2014a).

31. Ряд философов оспаривают предположение, что сознательный опыт непременно должен быть единым (Bayne, 2010). О вероятности отсутствия единства у сознания нередко заставляют задуматься пациенты с разделенным мозгом, то есть перенесшие операцию по разделению полушарий. Подробнее об этом в главе 3.

32. Tononi et al. (1994).

33. Seth et al. (2011a); Seth et al. (2006).

34. О сравнении методов измерения сложности в нейронаучном контексте см.: Seth et al. (2011a) и Mediano et al. (2019).

35. Demertzi et al. (2019); Luppi et al. (2019).

Глава 3. Ф

1. Основные статьи по теме ТИИ, опубликованные группой Тонони, — Tononi (2008); Tononi (2012); Tononi et al. (2016). С доступно изложенным манифестом теории можно ознакомиться в: Koch (2019).

2. Seth et al. (2006).

3. www.scientificamerican.com/article/is-consciousness-universal.

4. См., например, критику Скотта Ааронсона: www.scottaaronson.com/blog/?p=1799.

5. Ф можно рассматривать и как показатель эмерджентности. Эмерджентность — это очень общее понятие, связанное с тем, как макроскопические свойства (например, стая) складываются из своих микроскопических составляющих (отдельные птицы) или соотносятся с ними. См.: Hoel et al. (2013); Rosas et al. (2020); Seth (2010).

6. Эту аналогию используют сами Тонони и Кох (Tononi & Koch, 2015).

7. Излагается в адаптированном виде по: Tononi (2008).

8. В первом приближении, не берущем в расчет такие функции, как, например, регулировка контрастности, которая должна охватывать весь комплект.

9. Операция по расщеплению мозга предполагает рассечение мозолистого тела — крупного пучка нервных волокон, соединяющих полушария. Операция позволяет облегчить состояние страдающих тяжелыми формами эпилепсии, однако в наше время ее проводят редко, поскольку появляется все больше других, менее инвазивных методов лечения. Какая-то степень связи между полушариями у перенесших операцию сохраняется, но мы в своем теоретическом примере представим для простоты, что мозг действительно разделен надвое целиком. См.: de Haan et al. (2020).

10. Tononi et al. (2016).

11. Помимо аксиом интегративности и информативности ТИИ выдвигает три другие: сознание существует, оно состоит из множества компонентов и каждому сознательному опыту присущ собственный — эксклюзивный — пространственно-временной масштаб (Tononi et al., 2016). Философ Тим Бейн критикует ТИИ за то, что постулируемые ею аксиомы, особенно последняя, касающаяся «эксклюзивности», на самом деле нельзя назвать самоочевидно истинными (Bayne, 2018).

12. В информатике информация — снижение неопределенности — измеряется величиной под названием «энтропия». Энтропия (ее обычно обозначают буквой S) — это функция от числа отдельных состояний, в которых может пребывать система, суммированная с вероятностью пребывания в каждом из этих состояний. Она описывается уравнением $S = -\sum p_k \log(p_k)$, которое означает, что для каждого состояния (k) системы нужно умножить вероятность ее пребывания в этом состоянии на логарифм от этой вероятности, а затем сложить полученные величины для всех состояний. Наибольшего значения энтропия любой данной системы достигает в том случае, когда каждое состояние одинаково вероятно. У «честной» игральной кости энтропия составляет около 2,5 бит (когда логарифм берется по основанию 2). У шулерской кости энтропия ниже.

13. Следовательно, величина Φ для той или иной системы представляет собой скорее утверждение о ее механизме (как она устроена), чем о ее динамике (что она делает). Собственно, последние версии ТИИ описывают Φ через «несократимый каузальный показатель», а это утверждение о механизме, а не о динамике (Tononi et al., 2016).

14. Технически такое членение называется разделением по принципу минимальной информации. Не менее каверзная проблема — как обращаться с разделами разного размера, поскольку больший раздел будет способен, как включающий в себя больше составляющих, породить больше информации.

15. Согласно ТИИ, страна может стать «сознательной» в том случае, когда Φ достигнет максимального значения в пространственно-временном масштабе, охватывающем всю территорию страны, при этом отдельные люди окажутся аналогами нейронов в мозге. Из этого следует, что в случае обретения сознания страной в целом ее отдельные составляющие (люди) индивидуальное сознание утрачивают. Этот абсурдный сценарий анализировал американский философ Эрик Швитцгебель: schwitsplinters.blogspot.com/2012/03/why-tononi-should-think-that-united.html.

16. Deisseroth (2015).

17. См.: www.templetonworldcharity.org/accelerating-research-consciousness-our-structured-adversarial-collaboration-projects. Идею эксперимента с противопоставлением «бездействующих» нейронов «инактивированным» предложили Умберто Ольчезе и Джулио Тонони.

18. Wheeler (1989).

19. Barrett & Seth (2011); Mediano et al. (2019). С математической точки зрения наши измерения касаются эмпирического распределения системы, а не распределения максимальной энтропии.

20. Одна любопытная гипотеза объясняет устойчивую «пространственность» зрительных ощущений решетчатым анатомическим устройством нижних уровней зрительной коры (Haun & Tononi, 2019).

Глава 4. Восприятие, идущее изнутри

1. Не то чтобы перцептивные иерархические структуры были начисто изолированы друг от друга. Общее правило — чем дальше от сенсорной периферии, тем больше взаимодействия между модальностями. См. Felleman & Van Essen (1991); Stein & Meredith (1993).

2. Grill-Spector & Malach (2004).

3. Marr (1982).

4. He et al. (2016).

5. Anscombe (1959), p. 151, курсив автора (Энскома).

6. Из «Сокровища оптики» Ибн аль-Хайсама, ок. 1030 г., перевод в: Sabra (1989). Подробнее эта история изложена в потрясающей книге Якоба Хохви «Прогнозирующий разум» (The Predictive Mind) (2013).
7. Swanson (2016).
8. Bruner & Goodman (1947).
9. Gregory (1980).
10. Clark (2013); Clark (2016); Hohwy (2013); Rao & Ballard (1999).
11. Оговорюсь, что предсказания бывают и восходящими, а ошибки прогнозирования — нисходящими. Восходящие предсказания можно рассматривать как ограничители, отражающие глобальные и устойчивые составляющие перцептивного умозаключения (Teufel & Fletcher, 2020). Примером такого перцептивного ограничителя служит преобладание горизонтальной и вертикальной ориентации в изображениях природы. Таким же восходящим прогнозным ограничителем может выступать и влияние теней на освещенность (об этом чуть ниже).
12. См.: Dennett (1998). Вот еще один пример: чтобы услышать музыку, нам не нужен миниатюрный оркестрик в голове и хитроумная система внутричерепных микрофонов. Кстати, меткое выражение *mental figment* (измышление) тоже принадлежит Дэниелу Деннету.
13. Gasquet (1991). Эти слова приписывают и Паулю Клее.
14. За эти слова о действительности, о многом заставляющие задуматься, огромное спасибо «рецензируемому рэперу» Бабе Бринкману. О восприятии «того, как все обстоит на самом деле» мы подробнее поговорим в главе 6.
15. Бийю Хе с коллегами в недавнем своем проекте исследовала, как отличается нейронная динамика в случаях распознавания изображения в стиле Муни от случаев нераспознавания. См.: Flounders et al. (2019)
16. Brainard & Hurlbert (2015); Witzel et al. (2017).
17. Eye Benders (Gifford & Seth, 2013).
18. Великолепные примеры синусоидальной речи можно найти онлайн у Криса Дарвина: www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS. Еще один пример я привожу в своем выступлении на TED 2017: www.ted.com/talks/anil_seth_your_brain_hallucinates_your_conscious_reality. Кроме того, существуют и слуховые эквиваленты феномена белого или синего платья. Один из примеров — звук, которые одни слышат как «яни», другие как «лорел» (Pressnitzer et al., 2018). В 2020 г. в TikTok появился ролик, в котором дешевая игрушка произносит квакающим механическим голосом некое слово, воспринимаемое либо как *green needle* (зеленая игла), либо как *brainstorm* (мозговой шторм). В данном случае восприятие зависит от того, какой вариант слова вы увидите в подписи к ролику (time.com/5873627/green-needle-brainstorm-explained).
19. Обзор экспериментов, показывающих, как ожидания формируют восприятие, см.: de Lange et al. (2018).
20. Еще один интуитивный повод подумать о перцептивной достоверности — периферическое зрение, то есть часть поля зрения, удаленная от центральной (фовеальной). Хотя плотность фоторецепторов на периферии наших органов зрения гораздо ниже, чем в центральной ямке сетчатки, периферическое изображение не кажется нам размытым. Значит ли это, что периферическое зрение менее достоверно, чем фовеальное, поскольку демонстрирует «иллюзию резкости» (точнее, иллюзию «неразмытости»)? Нет! Резкость и размытость — это свойства перцептивного опыта, относящиеся к сенсорным данным, на которые настроена каждая из частей зрительной системы. Познавательное изложение см.: Naun (2021), исторический

контекст: Lettvin (1976), а масштабное всестороннее рассмотрение перцептивной (не)достоверности: Hoffman et al. (2015).

21. Джон Локк, «Опыт о человеческом разумении» (М.: Азбука, 2022). Хотя с цветом вроде бы все ясно на интуитивном уровне, философы довольно долго спорили, действительно ли его стоит считать вторичным качеством (Byrne & Hilbert, 2011). Аналогичное деление существует в философской литературе и для «видов» согласно тому, где проходит граница в условиях, необходимых для существования чего-то. Например, для существования денег необходимы социальные (общественные) условности, поэтому деньги можно квалифицировать как «общественное». Для существования воды социальные условности не нужны, поэтому она относится к «естественному».

Глава 5. Вероятное неочевидное

1. В оригинале — The Wizard of Odds (игра слов, построенная на замене названия Oz в «Волшебнике страны Оз» на Odds — «вероятности». — Прим. пер.) За название очередное спасибо Бабе Бринкману: оно взято из его «Рэперского руководства по сознанию» (Rap Guide to Consciousness) (2018), в котором я выступил научным консультантом. См.: <https://bababrinkman.com/shows/#consciousness>.

2. Пример приводится в адаптированном виде по F. V. Jensen (2000). Для более основательной подготовки в абдуктивном рассуждении лучше всего обратиться к «Выводу наилучшего объяснения» (Inference to the Best Explanation) Питера Липтона (2004).

3. Байесовское правило принято записывать в виде следующей формулы:

$$p(H|D) = \frac{p(D|H) \times p(H)}{p(D)}$$

где $p(H|D)$ — это апостериорная вероятность (то есть вероятность гипотезы H при реализации данных D , $p(D|H)$ — это условная вероятность (вероятность данных при реализации гипотезы), а $p(H)$ и $p(D)$ — это априорные вероятности гипотезы и данных соответственно. Оговорюсь, что обозначить $p(D)$ конкретным числом бывает довольно трудно, но, к счастью, обозначать его нужно не всегда. Когда задача, как это часто бывает, заключается в том, чтобы найти наиболее вероятное апостериорное предположение из ряда альтернатив, $p(D)$ в числителе и знаменателе этой дроби благополучно сокращаются.

4. Именно поэтому в 2009 г. правительство США рекомендовало воздержаться от поголовного скрининга женщин старше 40 на рак молочной железы. В то время чувствительность маммограмм составляла около 80%, то есть при скрининге рак груди в этой возрастной группе выявлялся где-то в 80% случаев. Однако примерно у 10% проходящих обследование скрининг сигнализировал о наличии рака ошибочно. Важно, что рак груди в этой возрастной группе встречается довольно редко — примерно у 0,04% ее представительниц. Применив байесовское правило и взяв этот показатель за априорную вероятность, можно вычислить, что вероятность наличия рака груди при указывающей на это маммограмме составляет всего 3%. То есть из каждых 100 женщин, получивших «настораживающую» маммограмму, 97 окажутся здоровы, но вынуждены будут мучиться, переживать и проходить дополнительные дорогие инвазивные обследования. Мораль (одна из) такова: нужно повышать чувствительность и диагностическую специфичность скрининга — и сейчас маммограммы действительно стали намного точнее. Согласно недавнему британскому исследованию, теперь проводить такие скрининги среди женщин старше сорока и вправду имеет смысл. См.: McGrayne (2012); Duffy et al. (2020).

5. На удивление неоднозначная история становления байесовского анализа превосходно изложена в «Бессмертной теории» Шарон Макгрейн (The Theory That Would Not Die) (2012).

6. Подробнее см.: Lakatos (1978) и Seth (2015b).

7. Снимаю шляпу перед Полом Флетчером и Крисом Фритом, которые высказали ту же самую мысль применительно к собственной байесовской теории галлюцинаций и бреда при шизофрении (Fletcher & Frith, 2009).

8. Строго говоря, x — переменная случайная, поскольку ее значение диктуется распределением вероятностей. Данный график представляет собой пример непрерывного распределения вероятностей (его называют также функцией плотности распределения вероятностей), так как x может принимать любое значение из допустимого диапазона. Если бы x мог принимать лишь определенные значения — например, «орел» или «решка», — у нас получилось бы дискретное распределение вероятностей.

9. Существуют две трактовки соответствия байесовских убеждений происходящему в мозге. Более слабая версия состоит в том, что этими убеждениями мы, внешние наблюдатели, пользуемся как репрезентацией действительности для себя (подобно географической карте как репрезентации окружающего нас мира). С этой точки зрения нейронная активность, которую мы наблюдаем, служит репрезентацией определенного положения дел для нас, исследователей. Более сильная версия гласит, что данные убеждения (или что-то им подобное) использует мозг — как репрезентацию существующего для себя. Вторая трактовка воплощает в себе гипотезу «байесовского мозга» и играет центральную роль в концепции мозга как прогнозирующей машины. Неразличение этих двух смыслов «репрезентации» ведет к немалой путанице в когнитивной науке и нейронауке. См.: Harvey (2008).

10. Иногда восходящий и нисходящий потоки называются соответственно прямой и обратной связью. С точки зрения прогнозной обработки должно быть как раз наоборот. В инженерии «обратная связь» обычно относится к сигналу об ошибке, на основании которого корректируется контрольный сигнал «прямой связи». Поэтому в прогнозной обработке «обратной» оказывается восходящая связь, ведь сигналы об ошибках передает именно она. Однако и здесь не все просто: как упоминалось ранее, в восходящие сигналы могут быть встроены некоторые глобальные, стабильные прогнозы, и тогда ошибки прогнозирования будут передаваться по нисходящим связям (Teufel & Fletcher, 2020).

11. В этой книге я обозначаю термином «прогнозная обработка» целый ряд теорий, рассматривающих в качестве центрального механизма минимизацию ошибок прогнозирования. Одна из таких теорий, но не единственная, — предиктивное кодирование. То, что я объединяю их под общим названием, не значит, будто я пытаюсь преуменьшить важные и интересные различия между ними (Hohwy, 2020a).

12. С точки зрения байесовских принципов порождающая модель — это сочетание априорной вероятности и условной. Читайте ее «совместной вероятностью гипотезы и данных», то есть $p(H, D)$. Эта формулировка подводит математическое обоснование под постулат, что минимизация ошибок прогнозирования аппроксимирует байесовский вывод (Buckley et al., 2017).

13. Иерархичность минимизации ошибок предсказания означает, что области мозга, участвующие в восприятии, должны быть хорошо оснащены средствами нисходящей связи, проводящими сигналы от высших уровней иерархии к низшим. Как показывают многочисленные исследования, именно так дело и обстоит, см., например: Markov et al. (2014). В концепцию восприятия как восходящего процесса наличие такой густой сети нисходящих связей вписывается плохо.

14. См.: Feldman & Friston (2010). Взвешивание прецизионности осуществляется за счет смены априорных вероятностей на прецизионные (так называемые гиперприорные) таким образом, чтобы выводимая прецизионность либо возрастала, либо снижалась.

15. См. также: Simons & Chabris (1999).

16. Обзор психологической составляющей в фокусах см.: Kuhn et al. (2008).

17. «Ориентированные на действие» формулировки прогнозной обработки уже довольно давно продвигает мой коллега из Сассекского университета Энди Кларк. Подробнее см. в его эпохальной работе «По волнам неопределенности» (Andy Clark, *Surfing Uncertainty*, 2016).

18. Как доказывает в своей книге 2019 г. «Мозг изнутри» (The Brain from Inside Out) нейробиолог Дьердь Бужаки, этот взгляд ставит перед экспериментальной нейронаукой новые задачи и открывает новые возможности. Пусть не все, но большинство экспериментаторов изучают мозг, анализируя его активность в ответ на внешние стимулы вместо того, чтобы подходить к нему как к динамичной по природе своей активной системе. См. также: Brembs (2020).

19. Friston et al. (2010)

20. Вместе с Александром Чанцем, Кристофером Бакли и Береном Миллиджем мы разрабатываем на этой основе новые алгоритмы машинного обучения, способные выстраивать порождающие модели на небольшом объеме данных (Tschantz et al., 2020a). Интересно, что в этом контексте многообещающей выглядит перспектива «восходящих» прогнозов (Teufel & Fletcher, 2020). Она может быть связана с применяемой в машинном обучении мощной техникой «амортизации», в рамках которой аппроксимированная апостериорная байесовская вероятность вычисляется за счет прямого (восходящего) прочесывания надлежащим образом обученной искусственной нейросети.

21. Экспериментальная проверка, которой эта идея отлично поддается, подтверждает, что проприоцептивная сенсорная чувствительность снижается во время действия (С. Е. Palmer et al., 2016). Собственно, именно сенсорным безразличием, сопровождающим действия, хорошо объясняется, почему нам не щекотно, когда мы щекочем сами себя (Н. Brown et al., 2013).

Глава 6. Вклад зрителя

1. Kandel (2012). (Кандель Э. Век самопознания. Поиски бессознательного в искусстве и науке с начала XX века до наших дней. — М.: Corpus, 2016.)

2. Gombrich (1961)

3. Seth (2019b).

4. Kandel (2012), p. 204.

5. Пример цитируется по Seth (2019b), где он представляет собой адаптацию приведенного в Albright (2012). См.: www.wikiart.org/en/camille-pissarro/hoarfrost-1873.

6. Слова из опубликованной 25 апреля 1874 г. в Le Charivari сатирической рецензии Леруа на выставку импрессионистов. Именно там впервые было упомянуто слово «импрессионизм», которое и дало название новаторскому направлению живописи.

7. Gombrich (1961).

8. На закате своих дней Писсарро страдал от серьезных проблем со зрением. То же самое происходило с другими представителями течения — Клодом Моне и Эдгаром Дега. Было бы любопытно выяснить, повлияло ли нарушение зрения на художественное видение, и если да, то насколько. Хотя полностью это влияние исключать нельзя — вполне возможно, что игра света становилась для нечетко видящих художников важнее мелких подробностей изображаемых предметов, гораздо более поучительным примером служит история другого художника, Эль Греко (1541–1614). На его полотнах мы часто видим неестественно вытянутые фигуры, пристрастие к которым ряд специалистов пытается объяснить выраженным

астигматизмом, которым страдал живописец. По их мнению, он писал фигуры именно такими, какими их видел, то есть вытянутыми. Но психологи совершенно справедливо замечают, что, по этой логике, он видел бы вытянутым и сам холст и это аннулировало бы воздействие астигматизма. В эту логическую ловушку, которую называют «ошибкой Эль Греко», исследователи восприятия попадают по сей день. См.: Firestone (2013).

9. Gombrich (1961), p. 170.

10. Эту мысль Гомбриха позже подхватил писатель, критик и художник Джон Бёрджер, начавший свой влиятельный труд 1972 г. «Искусство видеть» (Ways of Seeing) словами: «Соотношение между тем, что мы видим, и тем, что мы знаем, постоянно колеблется». В отличие от консервативного Гомбриха, Бёрджер подчеркивал политическое и культурное влияние на восприятие, показывая, как видимое может различаться у разных людей или в разных группах (Berger, 1972).

11. На самом деле это предположение не так уж самоочевидно, см.: Press et al. (2020).

12. Это вариация на тему более известного метода под названием «соперничество сетчаток», о котором я говорил в главе 1 (Blake et al., 2014).

13. Pinto et al. (2015). На самом деле исследование было сложнее, чем может показаться из приведенного здесь резюме. Мы провели большое количество контрольных экспериментов, чтобы по максимуму исключить влияние других факторов, таких как когнитивные искажения в реакции испытуемых или в сосредоточении внимания. Обзор схожих исследований см.: de Lange et al. (2018), а влиятельный вклад в литературу по этой теме на раннем этапе: Melloni et al. (2011).

14. «Чтение мозга» предполагает настройку алгоритмов машинного обучения на классификацию активности мозга в соответствии с рядом категорий. См.: Heilbron et al. (2020).

15. www.boredpanda.com/objects-with-faces.

16. Suzuki et al. (2017).

17. Если конкретнее, это глубокая свёрточная нейронная сеть (ГСНС), которую можно обучить с использованием стандартных алгоритмов обратного распространения ошибки. См.: Richards et al. (2019).

18. В стандартном режиме сети предъявляется изображение, сеть работает в восходящем по слоям направлении, и вывод сообщает нам «мнение» сети о предъявленной картинке. В алгоритме глубокого сна и его версии, адаптированной Кейсуке, этот процесс обращен вспять. Мы фиксируем вывод и подстраиваем под него ввод до тех пор, пока сеть не стабилизируется. Подробнее см.: Suzuki et al. (2017).

19. Полноценное воздействие галлюцинаторной машины, при котором испытуемый видит в шлеме VR панорамное видео, обеспечивает более ощутимый эффект присутствия, чем любой статичный снимок. С примером видеоизображения можно ознакомиться здесь: www.youtube.com/watch?v=TIMBnCrZZYY.

20. Подход, при котором вычислительные модели соотносятся с гипотезами о нейронных сетях, можно назвать «вычислительной нейрофеноменологией», то есть усиленной вычислительными технологиями версией нейрофеноменологии, основоположником которой выступал Франциско Варела (Francisco Varela, 1996). Действуя в том же духе, мы с Кейсуке Судзуки и Дэвидом Шварцманом разрабатываем новые версии галлюцинаторной машины, которые открыто инкорпорируют порождающие модели и поэтому лучше соотносятся с тем, что, на наш взгляд, происходит в настоящем мозге. Наши новые галлюцинаторные машины способны охватить гораздо более широкий спектр галлюцинаторного опыта, чем ее первый вариант.

21. Seth (2019b).

22. Merleau-Ponty (1964).

23. Как и все теории, эта во многом опирается на могучие плечи своих предшественниц, в частности идеи Джеймса Гибсона о том, насколько восприятие зависит от физического, телесного действия, а также на философскую феноменологию Эдмунда Гуссерля и Мориса Мерло-Понти. Введенное Гибсоном понятие «аффорданс», с которым мы еще встретимся в главе 9, отражает гипотезу, что мы воспринимаем объекты с точки зрения возможностей для действия, которое они в себе содержат (Gibson, 1979). Гуссерль предполагал, что «у восприятия есть горизонты, составленные из других вероятностей восприятия, то есть восприятия, которое мы получили бы, если бы активно направляли его ход куда-то еще» (Husserl, 1960 [1931]). Мерло-Понти, находившийся под сильным влиянием Гуссерля, подчеркивал роль телесных составляющих перцептивного опыта, и его «Феноменология восприятия» (М.: Наука, 1999) сохраняет свой авторитет до сих пор.

24. Еще один пример «вычислительной феноменологии». См. также: Cisek (2007) о нейрофизиологической теории, пересекающейся с этой феноменологией в вопросе «соперничества аффордансов».

25. Seth (2014b).

26. Suzuki et al. (2019).

27. Еще один хороший пример — постэффекты движения, такие как иллюзия водопада. Посмотрите в упор на водопад (или на видео водопада), затем переведите взгляд — например, на отвесную скалу рядом с ним. Вам покажется, что скала движется вверх, оставаясь при этом на месте.

28. Многие нейрочеловеки, занимающиеся временем, с этой точкой зрения не согласятся. На самом деле большинство психологических и нейронных моделей восприятия времени предполагают тот или иной нейронный «хронометр», выступающий ориентиром, с которым сравнивается физическое время, в результате чего и возникает ощущение продолжительности (van Rijn et al., 2014). Другие считают, что восприятие времени диктуется поступающими от самого организма сигналами, напоминающими ход часов (сердцебиение и т.п.) — см.: Wittmann (2013), — однако наши исследовательские группы сомневаются и в этой гипотезе (Suárez-Pinilla et al., 2019).

29. Занижение длительности долгих временных периодов и завышение длительности коротких — пример эффекта «регрессии к среднему». Этот эффект наблюдается во многих, если не во всех, перцептивных модальностях и служит характерным признаком байесовского вывода, поскольку среднее можно считать априорной вероятностью. В восприятии времени этот эффект известен как закон Фирордта.

30. См.: Roseboom et al. (2019). Дальнейшим подтверждением идеи Уоррика стало открытие, что вычислительная модель и результаты, показываемые человеком, совпадают еще точнее, когда ввод для сети ограничивается именно теми отрезками видео, которые смотрел человек.

31. Sherman et al. (2020).

32. Stetson et al. (2007). Разумеется, у основания крана была растянута страховочная сеть.

33. Первую версию этого проекта см.: Suzuki et al. (2012).

34. Phillips et al. (2001). Это направление исследований несет в себе потенциально важные социологические последствия. Поскольку люди ощущают свое восприятие как «реальное» и достоверное, им будет трудно признать, что перцептивный опыт у других, даже в тех же самых объективных обстоятельствах, может быть иным.

Именно поэтому среди прочего возник такой ажиотаж вокруг «феномена платя» (см. главу 4). Те, кто видит его одним образом, просто не могут допустить, что его возможно видеть иначе, именно потому, что свое восприятие ощущается ими как непосредственное отражение объективной действительности. Этот перцептивный дрейф — обобщение происходящего в «информационных пузырях» соцсетей — много говорит о том, как можно распознавать и разрешать или примирять различия между индивидуумами, группами и культурами (Seth, 2019c).

35. Слова взяты из «Трактата о человеческой природе» Дэвида Юма (1738, 1.3.14.25) (М.: АСТ, 2022) и «Исследования о принципах морали» (1751, Appendix 1.19), цитируемых в Kail (2007), p. 20. Я с ними познакомился в Dennett (2015).

36. В философии это называется «прозрачность» (Metzinger, 2003b).

Глава 7. Делирий

1. Collier (2012).

2. Davis et al. (2017).

Глава 8. Оправдание ожиданий

1. Авторство эксперимента независимо приписывается философу Дереку Парфиту и писателю Станиславу Лему.

2. Это представление о личности увековечено в философии как «теория пучка».

3. Metzinger (2003a).

4. Именно так происходит у Кристи и Татьяны Хоган, краниопагов из канадской провинции Британская Колумбия: www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html.

5. Обзор расстройств восприятия телесного «я» см.: Brugger & Lenggenhager (2014).

6. Классическую личную деконструкцию этого интуитивного представления см.: «Жизнь без головы» Дугласа Хардинга (М.: Ганга, 2017).

7. Статья, описывающая иллюзию резиновой руки, была опубликована в 1998 г. (Botvinick & Cohen, 1998) и породила целую небольшую отрасль дальнейших исследований. Проверяли, возможна ли иллюзия трех рук или иллюзия отсутствия руки; влияет ли на подспудные расовые предубеждения иллюзия владения руками с разным цветом кожи. Проверяли даже мышей — насколько они поддаются иллюзии «резинового хвоста». Обзоры см.: Braun et al. (2018) и Riemer et al. (2019).

8. Эксперимент срабатывал и в тех случаях, когда испытуемые видели не собственное тело, а «аватар» — спину искусственного тела, сгенерированного компьютером. См.: Ehrsson (2007); Lenggenhager et al. (2007).

9. Monroe (1971).

10. Цит. по: Tong (2003), p. 105.

11. Blanke et al. (2004), p. 248.

12. Blanke et al. (2015).

13. См.: Brugger & Lenggenhager (2014).

14. Petkova & Ehrsson (2008).

15. www.themachinetobeanother.org.

16. Lush et al. (2020). В этой статье мы вместо слова «гипноз» используем термин «феноменологический контроль», отчасти потому что за «гипнозом» тянется шлейф ненужных ассоциаций. Влияние подспудных ожиданий, порождаемых структурой эксперимента, на впечатления и поведение участников, — проблема хорошо известная, хотя и недостаточно учитываемая в психологии, первые упоминания о ней можно найти в работе о так называемых требуемых характеристиках (Orne, 1962).

Важно иметь в виду, что просто создать контраст между синхронным и асинхронным поглаживанием для устранения эффекта требуемых характеристик в иллюзии резиновой руки недостаточно, поскольку у испытуемых имеются сильные ожидания по поводу того, что они должны испытывать при том и другом условии (Lush, 2020). Краткое изложение наших исследований на тему иллюзии резиновой руки, феноменологического контроля и требуемых характеристик см.: Seth et al. (2021).

17. Помимо определенного поведения и сообщений о субъективном опыте гипнотическое внушение может вызывать физиологические и нейрофизиологические реакции (или в случае гипнотической анестезии подавлять их). (Barber, 1961; M. P. Jensen et al., 2017; Stoelb et al., 2009.) Таким образом, даже эти, явно более объективные, показатели могут искажаться из-за внушаемости. К счастью, эта особенность не всегда мешает, иногда она может оказаться полезной. Гипнотическое внушение — это мощный инструмент для исследования того, как нисходящие ожидания порождают или подавляют перцептивный опыт. Сейчас мы проводим ряд экспериментов, позволяющих подробнее посмотреть, какие виды перцептивного опыта — восприятие себя или восприятие мира — можно формировать или порождать под воздействием внушения.

18. Канадский психолог эстонского происхождения Эндель Тульвинг назвал эту разновидность самосознания «автоноэтической» (Tulving, 1985).

19. Дневниковые записи из книги Деборы Уэринг «Вечное сегодня» (Forever Today) (2005).

20. Wearing (2005), pp. 202–3.

21. www.newyorker.com/magazine/2007/09/24/the-abyss.

22. Существование непосредственного социального восприятия, определяемого таким образом, признается не всеми. Другие подходы предполагают, что наша осведомленность о психическом состоянии окружающих выводится из их поведения отличным от восприятия образом. Подробнее см.: Gallagher (2008) и C. J. Palmer et al. (2015).

23. Существование социального активного вывода означает, что точно так же, как при зрительном активном выводе, лежащие в его основе порождающие модели кодируют условные прогнозы следствий из действий. В случае со зрением, как мы помним из главы 6, эти прогнозы относятся к тому, как изменятся зрительные сенсорные сигналы в случае того или иного действия. Там я доказывал, что эти условные прогнозы составляют основу феноменологического свойства «объектности». В статье 2015 г. мы с Колином Палмером и Якобом Хохви высказали предположение, что нечто подобное происходит и в социальном восприятии. Наша идея заключается в том, что чужое психическое состояние кажется «реальным» в той степени, в которой наш мозг кодирует богатый репертуар условных прогнозов о том, как это состояние может измениться в случае того или иного действия. Например, прогноз может касаться того, как изменятся чьи-то убеждения или эмоциональное состояние в ответ на определенное высказывание (допустим, «принеси мне вина»). Эта идея придает «теории разума» прогнозно-машинный уклон и может послужить ценным способом объяснения явной недостаточности социального восприятия, которая возникает, например, при аутизме. См.: C. J. Palmer et al. (2015).

24. В нейронаучной полемике о социальном восприятии часто подчеркивается роль так называемых зеркальных нейронов. Эти нейроны, впервые обнаруженные в мозге нечеловекообразных обезьян итальянским нейробиологом Джакомо Риццолатти и его коллегами, выстреливают и в том случае, когда животное действует само, и в том случае, когда аналогичное действие производится другим животным (Gallese et al., 1996). Считается, что эти нейроны «зеркально отражают» действия других животных, поскольку реагируют так же, как если бы их обладатель выполнял это действие сам.

Наличием этих нейронов и их способностью на такую реакцию пытаются объяснять самые разные социальные явления. Однако подобные предположения возлагают слишком тяжелое объяснительное бремя на самую обыкновенную, по сути, разновидность клеток головного мозга. Авторы этих гипотез впадают в такое же упрощение, как те, кто на основании сканов фМРТ берется объяснять «любовь» или «владение языком» нейронной активностью в той или иной области мозга. См.: Caramazza et al. (2014).

25. «Нет человека, который был бы как Остров, сам по себе, каждый человек есть часть Материка, часть Суши» (пер. Н. Волжиной и Е. Калашниковой) Donne (1839), pp. 574–5. Эту идею развил психолог Крис Фрит, доказывая, что социальная функция является первостепенной для любого сознательного опыта (Frith, 2007).

26. Антонио Дамасио подчеркивает эту составляющую «я» в своей книге «Ощущение происходящего» (The Feeling of What Happens) (2000).

27. James (1890), p. 242.

Глава 9. Теория животного-машины

1. Анаис Нин, «Соблазнение Минотавра» (М.: Амфора, 2005). Нин упоминает эти слова как цитату из Талмуда.

2. Концепция Великой цепи бытия берет начало в Древней Греции, в философии Платона, Аристотеля и Плотина, однако полноценное развитие она получила в средневековой Западной Европе.

3. Красочный исторический экскурс см. в книге Джорджа Макари «Машина души» (Soul Machine) (2016).

4. https://en.m.wikipedia.org/wiki/Trademark_argument. См. также: Hatfield (2002).

5. Shugg (1968), p. 279. Оригинал см.: The Philosophical Works of Descartes, trans. E. S. Haldane and G. R. T. Ross (New York, 1955), vol. 1, 114–16, 118.

6. La Mettrie (1748).

7. Godfrey-Smith (1996); Maturana & Varela (1980).

8. Craig (2002).

9. Critchley & Harrison (2013).

10. Подробнее о роли островковой доли в interoцепции см.: Barrett & Simmons (2015) и Craig (2009).

11. James (1884), p. 190. Полемика между сторонниками «классической» теории эмоций, которые вслед за Дарвином предполагают сохранение врожденных эмоций у всего вида, и «конструктивистами», которые предполагают обратное, продолжается по сей день. К лагерю первых относятся биолог Яак Панксепп и его последователи. Панксепп доказывал, что набор базовых эмоций инстанцируется определенными (древними в эволюционном отношении) нейронными контурами (Panksepp, 2004); см. также: Darwin (1872). Второй лагерь представляют нейрочеловеки Лиза Фельдман Барретт и Джо Леду, которые выдвигают различные версии идеи, что человеческие эмоции зависят от когнитивной оценки, — этот взгляд, как мы еще убедимся, схож с моим. Подробнее об истории теорий эмоций см.: Barrett (2016); LeDoux (2012).

12. См., например: Harrison et al. (2010).

13. Классический справочный материал по оценочным теориям: Schachter & Singer (1962).

14. Dutton & Aron (1974).

15. В первом приближении эта идея была изложена в моей статье 2011 г. (Seth et al., 2011b), затем она была доработана в статье 2013 г., на которую с тех пор принято ссылаться как на основной источник (Seth, 2013). Начиная с 2015 г. на основе

ключевых положений развивалась касающаяся сознания и личности теория животного-машины: Seth (2015a); Seth (2019a); Seth & Friston (2016); Seth & Tsakiris (2018).

16. Идеи Дамасио во многом сформировали мои собственные, особенно в том, что касается нашего «я». См.: Damasio (1994, 2000, 2010). Лиза Фельдман Барретт, как и я, подчеркивает роль interoцептивных прогнозов в эмоциях, см.: Barrett & Satpute (2019) и Barrett & Simmons (2015), а также ее великолепную книгу «Как рождаются эмоции» (М.: МИФ, 2018).

17. См.: Petzschner et al. (2019). Дополнительные свидетельства существования interoцептивного вывода поступают из исследований с участием животных. В частности, два недавних эксперимента позволяют предположить, что нейроны в островковой коре мышей кодируют что-то подобное interoцептивным прогнозам (Gehrlach et al., 2019; Livneh et al., 2020).

18. См.: Aspell et al. (2013); Suzuki et al. (2013). В смежном исследовании Мика Аллен с коллегами показал, что на восприятие зрительных стимулов влияет неожиданное физиологическое возбуждение ЦНС, в очередной раз позволяя предположить взаимосвязь экстероцептивных и interoцептивных процессов (Allen et al., 2016). Обзор см.: Park & Blanke (2019), разностороннее рассмотрение проблем с измерением interoцептивной чувствительности с помощью отслеживания сердцебиения см.: Brener & Ring (2016), Zamariola et al. (2018). Влияние гипнотической внушаемости на телесное «я» обсуждалось в главе 8.

19. Более технические версии теории животного-машины и ее составляющих см.: Seth (2013), Seth (2014b), Seth (2015a), Seth (2019), Seth & Friston (2016) и Seth & Tsakiris (2018). Теория опирается на множество предшествующих и испытала немалое влияние концепций, которым я не могу в полной мере отдать здесь должное. Это и философское исследование личности, выполненное Томасом Метцингером (Metzinger, 2003a), и фундаментальные труды Энди Кларка и Якоба Хохви на тему прогнозной обработки (Clark, 2016; Hohwy, 2013). Моя теория многим обязана другим гипотезам о глубинных, но отличных от обозначаемых мной связях между жизнью, организмом, разумом и сознанием. Здесь я должен отметить особое влияние Антонио Дамасио (напр., 1994, 2010), Джеральда Эдельмана (напр., 1989), Карла Фристонна (напр., 2010), Джо Леду (напр., 2019) и Эвана Томпсона (напр., 2014; см. также Varela et al., 1993). О смежных идеях см.: Panksepp (2005); Park & Tallon-Baudry (2014); Solms (2021); концепцию Метцингера о «предпочтении существования» (Metzinger, 2021), а также работы Лизы Фельдман Барретт (напр., 2017).

20. Wiener (1948).

21. История кибернетики рассказывает не только о том, что научные дисциплины, которые мы теперь считаем различными, когда-то составляли одно направление, но и о том, как далеко наука отклоняется порой от пути, который мог бы привести ее к несметным сокровищам. Для более подробного ознакомления рекомендую книгу Жана-Пьера Дюпюи «О происхождении когнитивистики. Механизация разума» (On the Origins of Cognitive Science: The Mechanization of the Mind) (2009).

22. Conant & Ashby (1970).

23. Подробнее о различиях между «выступать моделью» и «обладать моделью» см.: Seth (2015a) и Seth & Tsakiris (2018).

24. В научных статьях я обозначаю это различие как разницу между эпистемической (исследовательской, связанной с поиском информации) и инструментальной (направленной на достижение цели, ориентированной на контроль) формами прогнозного восприятия: Seth (2019a); Seth & Tsakiris (2018); Tschantz et al. (2020b).

25. Ashby (1952). Например, температура человеческого тела должна удерживаться в диапазоне от 32 до 40 °С, при выходе за эти рамки быстро наступает смерть.

26. Внешние и внутренние действия различаются по типу участвующей в них мускулатуры. Внешние действия производятся скелетными (поперечнополосатыми) мышцами, а внутренние — гладкой мускулатурой внутренних органов и сердечной мышцей. Эти типы мышц контролируются разными ветвями периферической нервной системы (располагающейся за пределами головного и спинного мозга). Скелетные мышцы контролируются соматической ветвью, а мышцы внутренних органов и сердечная — вегетативной.

27. См.: McLeod et al. (2003).

28. Gibson (1979).

29. Для краткости эту теорию часто выражают в виде слогана «системы контроля контролируют то, что считывают, а не то, что делают» (Powers, 1973). Более свежую формулировку теории см.: Marken & Mansell (2013).

30. Я ничего не имею против метафорического приписывания физических характеристик невизуальным ощущениям. Боль, например, может называться острой или тупой. Острым может именоваться и вкус, и какие-то эмоции — укол ревности, допустим. Но от этого у ощущений все равно не появится физической формы, которую мы воспринимаем у чашки, кошки или журнального столика.

31. Sterling (2012). О вычислительной модели аллостатического interoцептивного контроля см.: Tschantz et al. (2021).

32. Тесная связь между сознанием и физиологической регуляцией ставит новые вопросы о роли ствола мозга — комплекса ядерных структур, расположенных между глубочайшими областями полушарий и спинным мозгом. Ствол мозга принято считать «обеспечивающим фактором» сознания (примерно как кабель электропитания служит обеспечивающим фактором для телевизора). Но поскольку ствол мозга играет весьма активную роль в физиологической регуляции, некоторые ученые предполагают, что именно здесь и возникает сознание — без всякого участия коры (Solms, 2021; см. также: Merker, 2007). На мой взгляд, это крайне маловероятно, учитывая массу свидетельств того, что кора (и таламус) связана с состояниями сознания. И тем не менее ствол мозга вполне может играть более влиятельную роль в формировании состояний сознания, чем предполагает аналогия с кабелем электропитания, подробнее см.: Parvizi & Damasio (2001).

33. Интересно было бы выяснить, ослабевают ли слепота к изменениям в себе во время болезни или при травмах и повреждениях, когда мозгу полезно точнее отслеживать, что происходит в организме. Поиском ответов на эти вопросы занимается новое направление когнитивной нейронауки — «вычислительная психосоматика» (Petzschner et al., 2017).

34. Это поняли еще древние греки. Если Сократ ассоциируется с максимой «познай себя», то стоики подчеркивали важность уравновешенности и самоконтроля. Сторонники теории перцептивного контроля могут пойти еще дальше и заявить, что мы регулируем свое физиологическое состояние, чтобы воспринимать самих себя как стабильных и неизменных.

35. Cotard (1880).

36. Одна из возможных ситуаций, при которых личность может утратить реальность, — когда находящиеся в ее основе порождающие модели не смогут кодировать богатый репертуар условных или контрфактуальных interoцептивных прогнозов, касающихся того, как действия отразятся на физиологической регуляции. Происходит это по аналогии с тем, как основу феноменологии «объектности» составляют зрительные условные прогнозы. См.: Seth & Tsakiris (2018).

37. Тезис о том, что наличие сознания обуславливается некими сугубо биологическими свойствами (которые у кремниевого компьютера по определению отсутствуют), называют иногда биологическим натурализмом. Я этот термин здесь не использую, поскольку в разных контекстах он может обозначать разное, см. обсуждение: Schneider (2019).

Глава 10. Рыба в воде

1. Фристон опубликовал большое число научных статей о принципе свободной энергии. Два ключевых обзора см.: Friston (2009, 2010).

2. Buckley et al. (2017).

3. www.lesswrong.com/posts/wpZJvgQ4HvJE2bysy/god-help-us-let-s-try-to-understand-friston-on-free-energy. Еще парочка шедевров: «Как читать Карла Фристона (в греческом оригинале)» (How to read Karl Friston (in the original Greek)) Алианны Марен, www.aliannajmaren.com/2017/07/27/how-to-read-karl-friston-in-the-original-greek) и «Свободная энергия. Как, мать вашу, она все-таки работает, если брать с экологической точки зрения» (Free Energy: How the f*ck does that work, ecologically) Эндрю Уилсона и Сабрины Голонки, psychsciencenotes.blogspot.com/2016/11/free-energy-how-fck-does-that-work.html).

4. ПСЭ рассматривает границы с точки зрения «марковского одеяла» — понятия из области статистики и машинного обучения. Для системы, описываемой набором случайных переменных, марковское одеяло — это статистическое членение системы на «внутренние состояния», «внешние состояния» и «одеяльные состояния», при котором одеяло отделяет внутреннее от внешнего. Марковское одеяло удовлетворяет условию, при котором переменные внутри одеяла (внутренние состояния) условно независимы от находящихся снаружи одеяла (внешних состояний), и наоборот. Подробнее о марковских одеялах и ПСЭ см.: Kirchhoff et al. (2018), а познавательную критику в: Bruineberg et al. (2020).

5. Технически свободная энергия служит верхней границей для величины под названием «неожиданность», или «собственная информация», которая обозначает, насколько (статистически) неожиданным является событие. Верхняя граница означает, что свободная энергия не может быть меньше собственной информации. Собственная информация связана с относящимся к информатике понятием энтропии в том, что при допущениях о неравновесных устойчивых состояниях долгосрочное среднестатистическое значение собственной информации будет представлять собой энтропию. Если упрощенно, энтропию можно рассматривать как неопределенность, а неопределенность — как среднестатистическую неожиданность, с которой вы предполагаете столкнуться.

6. Для математиков изложу подробнее. Свободная энергия определяется в терминах двух распределений вероятности: (1) плотность распознавания, кодирующая текущее наиболее вероятное предположение о состоянии среды, и (2) порождающая (генеративная) плотность, кодирующая вероятностную модель того, как состояния среды формируют (порождают) сенсорный ввод. Под «средой» здесь понимаются скрытые источники сенсорных сигналов, какими бы они ни были. Величины свободной энергии имеют две составляющие — энергию, соответствующую собственной информации, и относительную энтропию, которая отражает, насколько «далеко» плотность распознавания отстоит от подлинной апостериорной плотности (вероятность на всем охвате состояний среды с учетом сенсорного ввода). Разрыв между этими плотностями измеряется величиной, которая в информатике называется дивергенцией Кульбака — Лейблера (КЛ). Если плотность распознавания и порождающая плотность предполагаются гауссовскими (и с учетом некоторых других допущений, например о независимости временных шкал), свободная энергия прямо соотносится с прецизионно-взвешенными ошибками прогнозирования в

прогнозной обработке. Поскольку наименьшим значением «разрыва» будет ноль, свободная энергия всегда окажется больше, чем собственная информация (то есть она составляет верхнюю границу). Это, в свою очередь, означает, что уменьшение свободной энергии должно сократить либо разрыв между плотностью распознавания и подлинной апостериорной вероятностью (выдавая лучший перцептивный вывод), либо собственную информацию (отбирая новый сенсорный ввод).

7. Friston (2010). Якоб Хохви метко назвал этот процесс «самоосвидетельствование» (self-evidencing) (Hohwy, 2014). С математической точки зрения этот взгляд оправдан, поскольку минимизация свободной энергии эквивалентна максимизации (байесовской) маргинальной вероятности. Собственно, если для собственной информации свободная энергия служит верхней границей, то для маргинальной вероятности она выступает нижней границей (в машинном обучении это называется нижней границей свидетельства), см.: Winn & Bishop (2005).

8. Проблема «темной комнаты» была одним из первых контраргументов против ПСЭ (Friston et al., 2012). Его попытались выдвинуть снова, когда я уже работал над этой книгой, и мы с моими коллегами снова доказали его несостоятельность. См.: Seth et al. (2020); Sun & Firestone (2020).

9. Hohwy (2020b), p. 9.

10. Один из примеров подобного подкрепления теории животного-машины — уже изложенное представление о том, что живые системы сохраняют границу между собой и окружающей средой, в котором граница, согласно ПСЭ, понимается в терминах марковского одеяла (см. примечание 4). Для Фристана уже одно существование или идентификация марковского одеяла непосредственно указывает на процесс активного вывода. См.: Kirchhoff et al. (2018) и критические возражения снова в: Bruineberg et al. (2020).

11. Один из примеров — наше исследование, выясняющее, как агенты, минимизирующие свободную энергию, усваивают адаптивно-искаженные перцептивные модели своей среды (Tschantz et al., 2020b).

12. Hohwy & Seth (2020). Были и другие попытки связать ПСЭ с сознанием, например с точки зрения темпоральной глубины порождающих моделей (Friston, 2018); см. также Solms (2018); Solms (2021); Williford et al. (2018).

13. Эти попытки предпринимаются в виде «состязательного сотрудничества», при котором сторонники обеих теорий заранее обозначают, какой результат эксперимента послужит подтверждением их теории, а какой опровергнет. В данном конкретном состязательном сотрудничестве противопоставляется ТИИ и активный вывод (не сам ПСЭ). Один из предлагаемых экспериментов я упоминал в главе 3: согласно прогнозу ТИИ, инактивация и без того бездействующих нейронов изменит сознательное восприятие, тогда как активный вывод такого не предполагает.

Глава 11. Степени свободы

1. McEwan (2000). За эту цитату спасибо Патрику Хаггарду из Университетского колледжа Лондона.

2. Отличный путеводитель по этому философскому минному полю вы найдете в Bayne (2008).

3. Strawson (2008), p. 367.

4. Эта точка зрения, которой придерживаюсь и я, в философии называется компатибилизмом. Его идея состоит в том, что некоторые толковые концепции свободы воли вполне совместимы с представлением о детерминированности (предопределенности) вселенной. Либертарианская же концепция свободы воли (либертарианство здесь не в политологическом значении), напротив, представляет собой философскую версию «потусторонней» свободы воли. Кроме того, существуют

поборники так называемого жесткого детерминизма, провозглашающего истинность предопределенности и на основании этого заключающего, что сколько-нибудь разумных жизнеспособных концепций свободы воли существовать не может. Заметим, что агностицизм по отношению к предопределенности вселенной (в этом вопросе я действительно агностик) не мешает мне быть компатибилистом.

5. Даже если вселенная и вправду предопределена на каком-то основополагающем уровне, не исключено, что в функционировании мозга существенную роль играют случайные флуктуации на уровне нейронов и синапсов. Такое вполне возможно, даже, пожалуй, вероятно, однако опять-таки не имеет значения.

6. Впервые потенциал готовности был зафиксирован в 1960-х гг. немецкими физиологами Гансом Корнхубером и Людгером Дееке, которые назвали его *Bereitschaftspotential* (потенциал готовности) (Kornhuber & Deecke, 1965).

7. В философии различается намерение и побуждение (позыв). Я могу почувствовать побуждение ударить того, кто выводит меня из себя, но я это побуждение подавляю, поскольку не намереваюсь никого бить и попадать под арест. В отличие от побуждений, намерения регулируются доводами и нормами. Однако в простых случаях, таких как произвольное сгибание пальца, побуждение и намерение практически тождественны, поэтому сам Либет в основном пользовался этими терминами как взаимозаменяемыми.

8. Libet (1985). Как и следовало ожидать, и начало потенциала готовности, и момент сознательного намерения предшествовали самому движению.

9. Libet et al. (1983).

10. За несколько лет до этого аналогичную концептуальную мысль высказал философ Ал Меле (Mele, 2009).

11. Schurger et al. (2012).

12. Интересный эксперимент на эту тему см.: Caspar et al. (2016).

13. Цитата из статьи, представленной Шопенгауэром в Норвежском королевском научном обществе в 1839 г. Перевод см.: Zucker (2013), p. 531. В этом шопенгауэровском противопоставлении можно увидеть корни любой пагубной зависимости.

14. Здесь прослеживается интересная связь с кибернетикой. Росс Эшби, с чьей теоремой хорошего регулятора и понятием существенной переменной мы познакомились в главе 9, известен и как автор более раннего «закона необходимого разнообразия» (Ashby, 1956). Этот закон, или принцип, гласит, что успешная система контроля должна быть способна пребывать как минимум в стольких же состояниях, в скольких пребывает среда, которая эту систему возмущает. Как говорил Эшби, «только разнообразие может снизить разнообразие». См.: Seth (2015a).

15. Dennett (1984).

16. Brass & Haggard (2008); Haggard (2008); Haggard (2019).

17. Fried et al. (1991). Более интенсивная стимуляция этой области может вызвать и побуждение, и соответствующее действие.

18. Brass & Haggard (2007).

19. Wegner (2002).

20. Как выразился Сэм Харрис в недавнем подкасте, «дело не только в том, что проблема свободы воли объективно бессмысленна, она бессмысленна и субъективно». См.: <https://samharris.org/podcasts/241-final-thoughts-on-free-will/>.

21. Csikszentmihalyi (1990); Harris (2012).

22. Della Sala et al. (1991); Formisano et al. (2011).

23. Случай педофилии, вызванной опухолью, описан в Burns & Swerdlow (2003). На случай Чарльза Уитмана в литературе ссылаются часто, в том числе Дэвид Иглмен в статье 2011 г. в журнале The Atlantic: www.theatlantic.com/magazine/archive/2011/07/the-brain-on-trial/308520.

24. «Моих заслуг нет ни в чем. Все, от начала до конца, предопределено силами, над которыми мы не властны. Предопределено и для насекомого, и для звезды. Люди, овощи или космическая пыль — все мы пляшем под неведомую мелодию». В этой цитате из интервью, взятого у Эйнштейна Джорджем Сильвестром Виреком и опубликованного в The Saturday Evening Post 26 октября 1929 г., полностью раскрывается широко известная приверженность Эйнштейна детерминизму. (Он же, как мы знаем, сказал: «Бог не играет со вселенной в кости», отрицая заведомую случайность квантовой механики.) Однако, как мы уже видели, чтобы отрицать потустороннюю свободу воли, становиться детерминистом совсем не обязательно.

25. Нейробиолог Бьёрн Брембс доказывает, что признаки свободы воли в нашем понимании обнаруживаются даже у самых простых существ, когда они ведут себя разнообразным и очевидно произвольным образом, явно продиктованным «изнутри». Поскольку такое поведение, например непредсказуемая «реакция бегства» у таракана, благоприятно для спасения от врага, возможно, в нем таятся эволюционные истоки того, как мы, люди, научились управлять множеством своих степеней свободы. См.: Brembs (2011).

Глава 12. Помимо человека

1. Evans (1906).

2. Декартовская концепция увидела свет в XVII в. и какое-то время сосуществовала с этими еще не изжитыми средневековыми представлениями и обычаями. Вытеснила она их позже, когда в Европе воцарилось Просвещение.

3. С этим вряд ли согласятся сторонники теорий «сознания высшего порядка» (которых я коснулся в первой главе). См.: Brown et al. (2019).

4. Абсолютной противоположностью Декарта в этом смысле выступает Дарвин, который придерживался крайне антропоморфных взглядов, особенно по отношению к выражению эмоций у животных (Darwin, 1872). Его тезисом о сохраняющемся у многих видов общем наборе эмоций руководствуются исследователи, изучающие человеческие эмоции в экспериментах с участием животных, — это направление развивал позже Яак Панксепп и другие, а в современной культуре этот тезис нашел отражение в виде понятия о «вшитых», или «базовых», эмоциях, связанных с определенными мимическими выражениями, о которых пишет в своих работах Пол Экман (1992). Как мы знаем из главы 9, противниками представления о «базовых эмоциях» выступают современные «конструктивисты» в лице Лизы Фельдман Барретт и Джо Леду, которые, как и я, подчеркивают роль нисходящих интерпретаций в формировании содержания сознания. См.: Barrett (2016); LeDoux (2012).

5. См.: Steiner & Redish (2014), p. 1001.

6. В 1892 г. биолог Эрнст Геккель ввел термин «биопсихизм» для обозначения взгляда, согласно которому сознанием (способностью отдавать себе отчет в своих чувствах и ощущениях) обладают все живые существа, и только они, — в противопоставление панпсихистской концепции, предполагающей сознание у всех разновидностей материи. См.: Thompson (2007).

7. Более сложный способ измерения, называющийся коэффициентом энцефализации, высчитывает размер мозга относительно размеров тела. Споры о том, можно ли на основе коэффициента энцефализации с уверенностью прогнозировать наличие когнитивных способностей у разных видов, пока продолжаются (Herculano-Houzel,

2016; Reep et al., 2007). Однако ни размер мозга, ни коэффициент энцефализации не имеет смысла рассматривать в качестве показателя сознания у разных видов.

8. Seth et al. (2005).

9. Kelz & Mashour (2019).

10. Lyamin et al. (2018); Walker (2017).

11. Uexküll (1957).

12. Gallup (1970).

13. Так сообщалось в статье под заголовком, который я считаю одной из жемчужин своей коллекции: «Еще одна горилла (*Gorilla gorilla gorilla*) узнала себя в зеркале» (Posada & Colell, 2007).

14. Обзор см.: Gallup & Anderson (2020). Полемику о губанчиках см.: Kohda et al. (2019) и de Waal (2019).

15. Gallup & Anderson (2018).

16. Cowey & Stoerig (1995).

17. Boly et al. (2013).

18. Kessler & Rawlins (2016).

19. www.ted.com/talks/frans_de_waal_do_animals_have_morals.

20. Как показало одно исследование, после нескольких недель обучения резусы зеркальный тест проходят (L. Chang et al., 2017). Но прохождение теста после интенсивного натаскивания — это отнюдь не то же самое, что спонтанное узнавание себя в зеркале.

21. Borrelli et al. (2006). Из-за этого бесценного, но тяжеленного сокровища я здорово превысил установленный Ryanair лимит веса багажа во время обратного перелета.

22. В 2015 г. мне предложили написать главу для книги о «внеземном сознании». После некоторых терзаний я предпочел не рассуждать о гипотетических инопланетянах, а писать о реально существующих осьминогах (Seth, 2016a). Параллель между осьминогами и внеземным разумом проводил и Дени Вильнёв в своем фильме «Прибытие» 2016 г.

23. Godfrey-Smith (2017), p. 10.

24. Hochner (2012); Shigeno et al. (2018).

25. Carls-Diamante (2017).

26. Liscovitch-Brauer et al. (2017).

27. Fiorito & Scotto (1992). См. Также: D. B. Edelman & Seth (2009); Mather (2019); классический материал о поведении головоногих: Hanlon & Messenger (1996).

28. www.bbc.co.uk/programmes/p05nznfn1. Не менее впечатляющие кадры имеются в документальном фильме 2020 г. «Мой учитель — осьминог» (My Octopus Teacher) режиссеров Пиппы Эрлих и Джеймса Рида, в котором автор фильма Крейг Фостер выстраивает на удивление тесные отношения с осьминогом.

29. Морской биолог Роджер Хэнлон (Roger Hanlon) запечатлел множество видеопримеров маскировки осьминогов. Вот один из лучших: www.youtube.com/watch?v=JSq8nghQZqA.

30. Mather (2019); Messenger (2001).

31. van Giesen et al. (2020).

32. Neshet et al. (2014).

33. Подробнее о сознании у немлекопитающих см.: D. B. Edelman & Seth (2009) и D. B. Edelman et al. (2005).

34. Clayton et al. (2007); Jao Keehn et al. (2019); Pepperberg & Gordon (2005); Pepperberg & Shive (2001).

35. Любопытно, что, несмотря на сходство птичьего мозга с корой мозга млекопитающих (эта структура называется «паллиум»), у птиц отсутствует какой бы то ни было аналог мозолистого тела, связывающего два полушария у млекопитающих. Поэтому птиц можно назвать обладателями «естественно разделенного мозга» и задаваться вопросами о единстве их сознания (Xiao & Gunturkun, 2009). Прекрасным введением в тему сознания и поведения у птиц выступает «Существо с перьями» (The Thing with Feathers) Ноа Страйкера (2014).

36. Великолепное погружение в тему эволюции сознания см.: Feinberg & Mallatt (2017); Ginsburg & Jablonka (2019); LeDoux (2019).

37. Полезный обзор, выстроенный на тезисе, что при недостаточной убедительности свидетельств мы должны принимать решение в пользу животных (официально этот тезис называется принципом предосторожности), см.: Birch (2017). Надо отметить, что в 2010 г. Евросоюз решил включить головоногих в перечень видов, подпадающих под действие законов о защите животных (директива 2010/63/EU).

38. Entler et al. (2016).

39. Khuong et al. (2019).

40. Kelz & Mashour (2019).

41. Джонатан Берч, Александра Шнелль и Никола Клейтон вводят термин «профиль сознания» для характеристики межвидового разнообразия сознательного опыта. Вариации они предлагают характеризовать по пяти параметрам: перцептивное богатство, богатство оценки, единство, понимание временных отношений и самость. См.: Birch et al. (2020).

Глава 13. Машинный разум

1. В своей работе 1964 г. «Корпорация "Бог и голем"» (М.: АСТ, 2018) эрудированный основоположник кибернетики Норберт Винер отвел големам центральную роль в своих рассуждениях о рисках будущего ИИ.

2. Притча о производителе канцелярских скрепок рассказывает об ИИ, перед которым стоит задача сделать как можно больше скрепок. Поскольку человеческими нравственными ценностями этот ИИ не оперирует, но в остальном очень умен, то, успешно выполняя поставленную задачу, он уничтожает мир. См.: Bostrom (2014).

3. На редкость трезвый взгляд на гипотезу сингулярности см.: Shanahan (2015).

4. Велик соблазн сказать, что сознание и интеллект взаимно отделимы, то есть могут существовать друг без друга. Но это не совсем так. Хотя я считаю, что интеллект может существовать без сознания, не исключено, что сознание все же требует уровня интеллекта выше нулевого.

5. Идея многомерного сознания (и интеллекта) вызывает в памяти концепцию «профиля сознания», сформулированную Джонатаном Берчем и его коллегами (Birch et al., 2020), а также многомерный подход к уровню сознания у человека, предлагаемый Тимом Бейном, Якобом Хохви и Адрианом Оуэном (Bayne et al., 2016).

6. Dehaene et al. (2017). «Глобальная доступность» отсылает нас к популярной теории глобального рабочего пространства, а «самомониторинг» затрагивает некоторые аспекты теории сознания высшего порядка. С обеими теориями мы бегло знакомимся в главе 1. Авторы статьи в Science открыто признают, что, возможно, упускают из виду ту составляющую сознания, которая касается чувственного опыта. На мой взгляд, это слишком серьезное упущение.

7. Эта возможность возникает, потому что ТИИ признает один компонент функционализма (независимость от субстрата), но не другой (достаточность

соответствия ввода-вывода). Некоторые механизмы, в частности искусственные нейросети подходящего размера с прямой связью, могут воплощать произвольно сложные соответствия ввода-вывода. При надлежащей реализации эти механизмы будут производить впечатление обладающих интеллектом и/или сознанием. Но сети с одной лишь прямой связью не порождают никакой интегрированной информации — для нее требуется хотя бы какая-то возвратность. Таким образом, ТИИ оправдывает концепцию «поведенческого зомби», который, как я объяснял в главе 1, представляет собой артефакт, лишь выглядящий сознательным, но сознанием не обладающий. См.: Tononi & Koch (2015).

8. Эти идеи тесно связаны с понятием автопоэзиса (от греческих слов, означающих «сам» и «создание»), которое вывел чилийский биолог Умберто Матурана. Автопоэтической системой называется та, что способна поддерживать и воспроизводить себя, в том числе производя физические составляющие, необходимые для дальнейшего ее существования как системы. И хотя концепция автопоэзиса применяется главным образом к клеткам, можно проследить любопытные связи между клеточным автопоэзисом и принципом свободной энергии (см. главу 10). Оба предполагают сильную преемственность между «жизнью» и «разумом», из чего в свою очередь следует, что разум (и сознание) не сводится к тому, что система «делает» (Kirchhoff, 2018; Maturana & Varela, 1980). Мне посчастливилось встретиться с Матураной в январе 2019 г. (в мае 2021 г. он скончался на 93-м году жизни) в его родном Сантьяго, где мы обсудили эти идеи, попивая кофе в тенистом саду кофейни в районе Провиденсия.

9. Оригинальная «игра в имитацию», разработанная Аланом Тьюрингом, проводится с участием машины и двух человек одного пола. Машина и один из участников притворяются человеком противоположного пола. Второй участник должен понять, кто из собеседников машина, а кто человек (Turing, 1950).

10. Термин был введен Мюрреем Шанаханом, книгой которого «Воплощенное сознание и внутренняя жизнь» (Embodiment and the Inner Life) (2010) вдохновлялись среди прочего авторы фильма «Из машины».

11. www.reading.ac.uk/news-archive/press-releases/pr583836.html.

12. «Женя Густман — настоящий, подтверждено тестом Тьюринга», — констатирует The Guardian 9 июня 2014 г.

См.: <https://www.theguardian.com/technology/shortcuts/2014/jun/09/eugene-goostman-turing-test-computer-program>.

13. Характеристика теста Тьюринга как проверки человека на доверчивость взята из статьи Джона Маркоффа в The New York Times: «Программе хватает ума, чтобы пройти SAT (академический оценочный тест для поступления в вузы. — Прим. пер.), но до интеллекта там еще очень далеко» ('Software is smart enough for SAT, but still far from intelligent', The New York Times, 21 сентября 2015 г.).

См.: www.nytimes.com/2015/09/21/technology/personaltech/software-is-smart-enough-for-sat-but-still-far-from-intelligent.html.

14. GPT означает Generative Pre-trained Transformer (порождающая обученная заранее модель-трансформер) — разновидность нейронных сетей, специализирующаяся на прогнозной обработке и порождении языка. Эти сети обучаются с использованием алгоритмов глубокого обучения без учителя «прогнозировать следующее слово» на основании предшествующего слова или текстового отрывка. GPT-3 имеет впечатляющие 175 млрд параметров, ее обучали на 45 терабайтах текстовых данных. См.: <https://openai.com/blog/openai-api/> and for technical details: <https://arxiv.org/abs/2005.14165>.

15. Конечно, это зависит от того, что подразумевать под «пониманием». Кто-то может сказать, что человеческое «понимание» тоже не особенно отличается от того,

которое демонстрирует GPT-3. Но когнитивист Гэри Маркус эту точку зрения опровергает, и я с его аргументами согласен.

См.: www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/.

16. «Эта статья полностью написана роботом. Тебе уже страшно, человек?» ('A robot wrote this entire article. Are you scared yet, human?' The Guardian Opinion, 8 сентября 2020 г.). См.: www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3. Насколько этот пример репрезентативен, сказать трудно.

17. Becker-Asano et al. (2010).

18. Mori et al. (2012).

19. Технология «дипфейк» позволяет создавать реалистичное, но поддельное видео (обычно человеческого лица), обрабатывая исходный видеоматериал с помощью машинного обучения, чтобы получить требуемое видеоизображение. В широко распространенном примере 2017 г. с помощью технологии «дипфейк» были созданы убедительные видео Барака Обамы, произносящего то, чего он на самом деле не говорил (<https://www.youtube.com/watch?v=cQ54GDm1eL0>). Затем планку ощутимо подняла серия вышедших в 2021 г. дипфейковых роликов в TikTok, героем которых стал Том Круз (<https://www.theverge.com/22303756/tiktok-tom-cruise-impersonator-deepfake>).

20. Исследователь искусственного интеллекта Стюарт Рассел в своей книге «Совместимость» (М.: АНФ, 2020) в красках описывает угрозы, исходящие от уже имеющегося ИИ и того, что появится в ближайшем будущем, а также способы реконструировать системы ИИ, чтобы этих угроз избежать. Не менее замечательную работу проделывает Нина Шик, рассказывая об опасностях, которыми нам грозит технология «дипфейк» (Schick, 2020).

21. «Философ Дэниел Деннет об ИИ, роботах и религии» ('Philosopher Daniel Dennett on AI, robots and religion', Financial Times, 3 марта 2017 г.).

См.: <https://www.ft.com/content/96187a7a-fce5-11e6-96f8-3700c5664d30>.

22. См.: Metzinger (2021).

23. В 2020 г. Нобелевскую премию по химии получили Эммануэль Шарпантье и Дженнифер Дудна за вклад в разработку технологии CRISPR. Искусственная бактерия *E. coli* была синтезирована в лаборатории Джейсона Чина, см.: Fredens et al. (2019).

24. Trujillo et al. (2019).

25. Вопросы вероятности сознания у органоидов я исследовал в недавней научной статье, написанной совместно с Тимом Бейном и Марчелло Массимини (Wayne et al., 2020).

26. К этим вопросам подходят серьезно. Летом 2020 г. меня пригласили вместе с несколькими другими нейрочеными выступить на заседании объединенного комитета Национальной академии США, созванном с целью выработки нормативно-правовой базы для исследования органоидов и химер (животных, подвергшихся генетической модификации для придания им специфически человеческих черт). См.: www.nationalacademies.org/our-work/ethical-legal-and-regulatory-issues-associated-with-neural-chimeras-and-organoids.

27. Карл Циммер, «Органоиды — это не мозг. Как же они производят мозговые волны?» (Carl Zimmer, 'Organoids are not brains. How are they making brain waves?', The New York Times, 29 августа 2019 г.).

См.: www.nytimes.com/2019/08/29/science/organoids-brain-alysson-muotri.html.

28. Взвешенное разностороннее рассмотрение перспектив и ловушек переноса сознания см.: Schneider (2019).

29. Доказательство симуляции выглядит следующим образом (Bostrom, 2003). У цивилизации из достаточно далекого будущего, сумевшей дожить до него, не уничтожив себя, скорее всего, будут в распоряжении обширные вычислительные ресурсы. Некоторые представители этой цивилизации, возможно, захотят создать подробную компьютерную симуляцию своих предков. Учитывая, что подобные симуляции можно будет создавать в большом количестве, для любого живущего сейчас логично прийти к выводу, что он с гораздо большей вероятностью живет среди симулякров, чем среди биологических оригиналов. Как пишет Бостром, «если мы не считаем, что сейчас живем в компьютерной симуляции, у нас нет оснований считать, что наши потомки будут массово создавать симулякры своих предков» (Bostrom, 2003, p. 243). Этот довод мне не нравится среди прочего тем, что он подразумевает истинность функционализма, то есть по отношению к сознанию симуляция будет тождественна инстанцированию. А я, как уже упоминалось ранее, не считаю функционализм надежной отправной точкой.

Эпилог

1. Может ли это отсоединенное полушарие, коль скоро оно по-прежнему снабжается кровью и «живет», сохранять свое собственное изолированное сознание? Не исключено, что схожие «острова осознанности» будут создаваться и с помощью других развивающихся сейчас нейротехнологий, таких как внечерепное оживление мозга свиньи и органоиды головного мозга, о которых я говорил в предыдущей главе. Все эти случаи мы обсуждаем в: Wayne et al. (2020).

2. Chalmers (1995b), p. 201.

3. См. также: Hoffman (2019).

Библиография

Albright, T. D. (2012). 'On the perception of probable things: neural substrates of associative memory, imagery, and perception'. *Neuron*, 74(2), 227–45.

Allen, M., Frank, D., Schwarzkopf, D. S., et al. (2016). 'Unexpected arousal modulates the influence of sensory noise on confidence'. *Elife*, 5, e18103.

Anscombe, G. E. M. (1959). *An Introduction to Wittgenstein's Tractatus*. London: St. Augustine's Press.

Aru, J., Bachmann, T., Singer, W., et al. (2012). 'Distilling the neural correlates of consciousness'. *Neuroscience and Biobehavioral Reviews*, 36(2), 737–46.

Ashby, W. R. (1952). *Design for a Brain*. London: Chapman and Hall.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman and Hall.

Aspell, J. E., Heydrich, L., Marillier, G., et al. (2013). 'Turning the body and self inside out: Visualized heartbeats alter bodily self-consciousness and tactile perception'. *Psychological Science*, 24(12), 2445–53.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. New York, NY: Cambridge University Press.

Barber, T. X. (1961). 'Physiological effects of "hypnosis"'. *Psychological Bulletin*, 58, 390–419.

Barnes, J. (2008). *Nothing to Be Frightened of*. New York, NY: Knopf.

Barnett, L., Muthukumaraswamy, S. D., Carhart-Harris, R. L., et al. (2020). 'Decreased directed functional connectivity in the psychedelic state'. *Neuroimage*, 209, 116462.

- Barrett, A. B., & Seth, A. K. (2011). 'Practical measures of integrated information for time-series data'. *PLoS Computational Biology*, 7(1), e1001052.
- Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Boston, MA: Houghton Mifflin Harcourt. (Фельдман Барретт Л. Как рождаются эмоции. Революция в понимании мозга и управлении эмоциями. — М.: МИФ, 2018.)
- Barrett, L. F., & Satpute, A. B. (2019). 'Historical pitfalls and new directions in the neuroscience of emotion'. *Neuroscience Letters*, 693, 9–18.
- Barrett, L. F., & Simmons, W. K. (2015). 'Interoceptive predictions in the brain'. *Nature Reviews Neuroscience*, 16(7), 419–29.
- Bauby, J.-M. (1997). *The Diving Bell and the Butterfly*. Paris: Robert Laffont. (Боби Ж.-Д. Скафандр и бабочка. — М.: Рипол-Классик, 2009.)
- Bayne, T. (2008). 'The phenomenology of agency'. *Philosophy Compass*, 3(1), 182–202.
- Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.
- Bayne, T. (2018). 'On the axiomatic foundations of the integrated information theory of consciousness'. *Neuroscience of Consciousness*, 1, niy007.
- Bayne, T., Hohwy, J., & Owen, A. M. (2016). 'Are there levels of consciousness?' *Trends in Cognitive Sciences*, 20(6), 405–413.
- Bayne, T., Seth, A. K., & Massimini, M. (2020). 'Are there islands of awareness?' *Trends in Neurosciences*, 43(1), 6–16.
- Bechtel, W., & Williamson, R. C. (1998). 'Vitalism'. In E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.
- Becker-Asano, C., Ogawa, K., Nishio, S., et al. (2010). 'Exploring the uncanny valley with Geminoid HI-1 in a real-world application'. In *IADIS International Conferences Interfaces and Human Computer Interaction*, 121–8.
- Berger, J. (1972). *Ways of Seeing*. London: Penguin.
- Birch, J. (2017). 'Animal sentience and the precautionary principle'. *Animal Sentience*, 16(1).
- Birch, J., Schnell, A. K., & Clayton, N. S. (2020). 'Dimensions of animal consciousness'. *Trends in Cognitive Sciences*, 24(10), 789–801.
- Blake, R., Brascamp, J., & Heeger, D. J. (2014). 'Can binocular rivalry reveal neural correlates of consciousness?' *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641), 20130211.
- Blanke, O., Landis, T., Spinelli, L., et al. (2004). 'Out-of-body experience and autoscapy of neurological origin'. *Brain*, 127 (Pt 2), 243–58.
- Blanke, O., Slater, M., & Serino, A. (2015). 'Behavioral, neural, and computational principles of bodily self-consciousness'. *Neuron*, 88(1), 145–66.
- Block, N. (2005). 'Two neural correlates of consciousness'. *Trends in Cognitive Sciences*, 9(2), 46–52.
- Boly, M., Seth, A. K., Wilke, M., et al. (2013). 'Consciousness in humans and non-human animals: recent advances and future directions'. *Frontiers in Psychology*, 4, 625.
- Borrelli, L., Gherardi, F., & Fiorito, G. (2006). *A Catalogue of Body Patterning in Cephalopoda*. Florence: Firenze University Press.
- Bostrom, N. (2003). 'Are you living in a computer simulation?' *Philosophical Quarterly*, 53(11), 243–55.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

- Botvinick, M., & Cohen, J. (1998). 'Rubber hands "feel" touch that eyes see'. *Nature*, 391(6669), 756.
- Brainard, D. H., & Hurlbert, A. C. (2015). 'Colour vision: understanding #TheDress'. *Current Biology*, 25(13), R551–4.
- Brass, M., & Haggard, P. (2007). 'To do or not to do: the neural signature of self-control'. *Journal of Neuroscience*, 27(34), 9141–5.
- Brass, M., & Haggard, P. (2008). 'The what, when, whether model of intentional action'. *Neuroscientist*, 14(4), 319–25.
- Braun, N., Debener, S., Sychala, N., et al. (2018). 'The senses of agency and ownership: a review'. *Frontiers in Psychology*, 9, 535.
- Brembs, B. (2011). 'Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates'. *Proceedings of the Royal Society B: Biological Sciences*, 278(1707), 930–39.
- Brembs, B. (2020). 'The brain as a dynamically active organ'. *Biochemical and Biophysical Research Communications*. doi:10.1016/j.bbrc.2020.12.011.
- Brener, J., & Ring, C. (2016). 'Towards a psychophysics of interoceptive processes: the measurement of heartbeat detection'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160015.
- Brown, H., Adams, R. A., Parees, I., et al. (2013). 'Active inference, sensory attenuation and illusions'. *Cognitive Processing*, 14(4), 411–27.
- Brown, R., Lau, H., & LeDoux, J. E. (2019). 'Understanding the higher-order approach to consciousness'. *Trends in Cognitive Sciences*, 23(9), 754–68.
- Brugger, P., & Lenggenhager, B. (2014). 'The bodily self and its disorders: neurological, psychological and social aspects'. *Current Opinion in Neurology*, 27(6), 644–52.
- Bruineberg, J., Dolega, K., Dewhurst, J., et al. (2020). 'The Emperor's new Markov blankets'. <http://philsci-archive.pitt.edu/18467>.
- Bruner, J. S., & Goodman, C. C. (1947). 'Value and need as organizing factors in perception'. *Journal of Abnormal and Social Psychology*, 42(1), 33–44.
- Buckley, C., Kim, C.-S., McGregor, S., & Seth, A. K. (2017). 'The free energy principle for action and perception: A mathematical review'. *Journal of Mathematical Psychology*, 81, 55–79.
- Burns, J. M., & Swerdlow, R. H. (2003). 'Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign'. *Archives of Neurology*, 60(3), 437–40.
- Buzsáki, G. (2019). *The Brain from Inside Out*. Oxford: Oxford University Press.
- Byrne, A., & Hilbert, D. (2011). 'Are colors secondary qualities?' In L. Nolan (ed.), *Primary and Secondary Qualities: The Historical and Ongoing Debate*. Oxford: Oxford University Press, 339–61.
- Caramazza, A., Anzellotti, S., Strnad, L., et al. (2014). 'Embodied cognition and mirror neurons: a critical assessment'. *Annual Review of Neuroscience*, 37, 1–15.
- Carhart-Harris, R. L., Erritzoe, D., Williams, T., et al. (2012). 'Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin'. *Proceedings of the National Academy of Sciences of the USA*, 109(6), 2138–43.
- Carls-Diamante, S. (2017). 'The octopus and the unity of consciousness'. *Biology and Philosophy*, 32, 1269–87.
- Casali, A. G., Gosseries, O., Rosanova, M., et al. (2013). 'A theoretically based index of consciousness independent of sensory processing and behavior'. *Science Translational Medicine*, 5(198), 198ra105.

- Casarotto, S., Comanducci, A., Rosanova, M., et al. (2016). 'Stratification of unresponsive patients by an independently validated index of brain complexity'. *Annals of Neurology*, 80(5), 718–29.
- Caspar, E. A., Christensen, J. F., Cleeremans, A., et al. (2016). 'Coercion changes the sense of agency in the human brain'. *Current Biology*, 26(5), 585–92.
- Chalmers, D. J. (1995a). 'Facing up to the problem of consciousness'. *Journal of Consciousness Studies*, 2(3), 200–19.
- Chalmers, D. J. (1995b). 'The puzzle of conscious experience'. *Scientific American*, 273(6), 80–6.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York, NY: Oxford University Press.
- Chalmers, D. J. (2018). 'The meta-problem of consciousness'. *Journal of Consciousness Studies*, 25(9–10), 6–61.
- Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. New York, NY: Oxford University Press.
- Chang, L., Zhang, S., Poo, M. M., et al. (2017). 'Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images'. *Proceedings of the National Academy of Sciences of the USA*, 114(12), 3258–63.
- Churchland, P. S. (1996). 'The hornswoggle problem'. *Journal of Consciousness Studies*, 3(5–6), 402–8.
- Cisek, P. (2007). 'Cortical mechanisms of action selection: the affordance competition hypothesis'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485), 1585–99.
- Clark, A. (2013). 'Whatever next? Predictive brains, situated agents, and the future of cognitive science'. *Behavioral and Brain Sciences*, 36(3), 181–204.
- Clark, A. (2016). *Surfing Uncertainty*. Oxford: Oxford University Press.
- Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). 'Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 507–22.
- Cobb, M. (2020). *The Idea of the Brain: A History*. London: Profile Books.
- Collier, R. (2012). 'Hospital-induced delirium hits hard'. *Canadian Medical Association Journal*, 184(1), 23–4.
- Conant, R., & Ashby, W. R. (1970). 'Every good regulator of a system must be a model of that system'. *International Journal of Systems Science*, 1(2), 89–97.
- Cotard, J. (1880). 'Du délire hypocondriaque dans une forme grave de la mélancolie anxieuse. Mémoire lu à la Société médico-psychophysiologique dans la séance du 28 Juin 1880'. *Annales Medico-Psychologiques*, 168–74.
- Cowey, A., & Stoerig, P. (1995). 'Blindsight in monkeys'. *Nature*, 373(6511), 247–9.
- Craig, A. D. (2002). 'How do you feel? Interoception: the sense of the physiological condition of the body'. *Nature Reviews Neuroscience*, 3(8), 655–66.
- Craig, A. D. (2009). 'How do you feel — now? The anterior insula and human awareness'. *Nature Reviews Neuroscience*, 10(1), 59–70.
- Craver, C., & Tabery, J. (2017). 'Mechanisms in science'. In *The Stanford Encyclopedia of Philosophy*. plato.stanford.edu/entries/science-mechanisms.
- Crick, F., & Koch, C. (1990). 'Towards a neurobiological theory of consciousness'. *Seminars in the Neurosciences*, 2, 263–75.

- Critchley, H. D., & Harrison, N. A. (2013). 'Visceral influences on brain and behavior'. *Neuron*, 77(4), 624–38.
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper & Row. (Чиксентмихайи М. Поток: Психология оптимального переживания. — М.: АНФ, 2023.)
- Damasio, A. (1994). *Descartes' Error*. London: Macmillan.
- Damasio, A. (2000). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harvest Books.
- Damasio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. London: William Heinemann. (Дамасио А. Я: Мозг и возникновение сознания. — М.: Карьера Пресс, 2018.)
- Darwin, C. (1872). *The Expression of Emotions in Man and Animals*. London: Fontana Press. (Дарвин Ч. О выражении эмоций у человека и животных. — СПб.: Питер, 2013.)
- Davis, D. H., Muniz-Terrera, G., Keage, H. A., et al. (2017). 'Association of delirium with cognitive decline in late life: a neuropathologic study of three population-based cohort studies'. *JAMA Psychiatry*, 74(3), 244–51.
- de Graaf, T. A., Hsieh, P. J., & Sack, A. T. (2012). 'The "correlates" in neural correlates of consciousness'. *Neuroscience and Biobehavioral Reviews*, 36(1), 191–7.
- de Haan, E. H., Pinto, Y., Corballis, P. M., et al. (2020). 'Split-brain: What we know about cutting the corpus callosum now and why this is important for understanding consciousness'. *Neuropsychological Review*, 30, 224–33.
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). 'How do expectations shape perception?' *Trends in Cognitive Sciences*, 22(9), 764–79.
- de Waal, F. B. M. (2019). 'Fish, mirrors, and a gradualist perspective on self-awareness'. *PLoS Biology*, 17(2), e3000112.
- Dehaene, S., & Changeux, J. P. (2011). 'Experimental and theoretical approaches to conscious processing'. *Neuron*, 70(2), 200–227.
- Dehaene, S., Lau, H., & Kouider, S. (2017). 'What is consciousness, and could machines have it?' *Science*, 358(6362), 486–92.
- Deisseroth, K. (2015). 'Optogenetics: ten years of microbial opsins in neuroscience'. *Nature Neuroscience*, 18(9), 1213–25.
- Della Sala, S., Marchetti, C., & Spinnler, H. (1991). 'Right-sided anarchic (alien) hand: a longitudinal study'. *Neuropsychologia*, 29(11), 1113–27.
- Demertzi, A., Tagliazucchi, E., Dehaene, S., et al. (2019). 'Human consciousness is supported by dynamic complex patterns of brain signal coordination'. *Science Advances*, 5(2), eaat7603.
- Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
- Dennett, D. C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown.
- Dennett, D. C. (1998). 'The myth of double transduction'. In S. Hameroff, A. W. Kasniak, & A. C. Scott (eds), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, Cambridge, MA: MIT Press, 97–101.
- Dennett, D. C. (2003). *Freedom Evolves*. New York, NY: Penguin Books.
- Dennett, D. C. (2015). 'Why and how does consciousness seem the way it seems?' In T. Metzinger & J. M. Windt (eds), *Open MIND*. Frankfurt-am-Main: MIND Group.
- Dennett, D. C. & Caruso, G. (2021). *Just Deserts: Debating Free Will*. Cambridge: Polity.

- Deutsch, D. (2012). *The Beginning of Infinity: Explanations that Transform the World*. New York NY: Penguin Books.
- DiNuzzo, M., & Nedergaard, M. (2017). 'Brain energetics during the sleep-wake cycle'. *Current Opinion in Neurobiology*, 47, 65–72.
- Donne, J. (1839). 'Devotions upon emergent occasions: Meditation XVII' [1624]. In H. Alford (ed.), *The Works of John Donne*, London: Henry Parker, vol. 3, 574–5.
- Duffy, S. W., Vulkan, D., Cuckle, H., et al. (2020). 'Effect of mammographic screening from age forty years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial'. *Lancet Oncology*, 21(9), 1165–72.
- Dupuy, J.-P. (2009). *On the Origins of Cognitive Science: The Mechanization of the Mind*. 2nd edn. Cambridge, MA: MIT Press.
- Dutton, D. G., & Aron, A. P. (1974). 'Some evidence for heightened sexual attraction under conditions of high anxiety'. *Journal of Personal and Social Psychology*, 30(4), 510–17.
- Edelman, D. B., Baars, B. J., & Seth, A. K. (2005). 'Identifying hallmarks of consciousness in non-mammalian species'. *Consciousness and Cognition*, 14(1), 169–87.
- Edelman, D. B., & Seth, A. K. (2009). 'Animal consciousness: a synthetic approach'. *Trends in Neuroscience*, 32(9), 476–84.
- Edelman, G. M. (1989). *The Remembered Present*. New York, NY: Basic Books.
- Edelman, G. M., & Gally, J. (2001). 'Degeneracy and complexity in biological systems'. *Proceedings of the National Academy of Sciences of the USA*, 98(24), 13763–8.
- Ehrsson, H. H. (2007). 'The experimental induction of out-of-body experiences'. *Science*, 317(5841), 1048.
- Ekman, P. (1992). 'An argument for basic emotions'. *Cognition and Emotion*, 6(3–4), 169–200.
- Entler, B. V., Cannon, J. T., & Seid, M. A. (2016). 'Morphine addiction in ants: a new model for self-administration and neurochemical analysis'. *Journal of Experimental Biology*, 219 (Pt 18), 2865–9.
- Evans, E. P. (1906). *The Criminal Prosecution and Capital Punishment of Animals*. London: William Heinemann.
- Feinberg, T. E., & Mallatt, J. M. (2017). *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge, MA: MIT Press.
- Feldman, H., & Friston, K. J. (2010). 'Attention, uncertainty, and free-energy'. *Frontiers in Human Neuroscience*, 4, 215.
- Felleman, D. J., & Van Essen, D. C. (1991). 'Distributed hierarchical processing in the primate cerebral cortex'. *Cerebral Cortex*, 1(1), 1–47.
- Ferrarelli, F., Massimini, M., Sarasso, S., et al. (2010). 'Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness'. *Proceedings of the National Academy of Sciences of the USA*, 107(6), 2681–6.
- Fiorito, G., & Scotto, P. (1992). 'Observational learning in *Octopus vulgaris*'. *Science*, 256(5056), 545–7.
- Firestone, C. (2013). 'On the origin and status of the "El Greco fallacy"'. *Perception*, 42(6), 672–4.
- Fletcher, P. C., & Frith, C. D. (2009). 'Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia'. *Nature Reviews Neuroscience*, 10(1), 48–58.
- Fleming, S. M. (2020). 'Awareness as inference in a higher-order state space'. *Neuroscience of Consciousness*, 2020(1), niz020.

- Flounders, M. W., Gonzalez-Garcia, C., Hardstone, R., et al. (2019). 'Neural dynamics of visual ambiguity resolution by perceptual prior.' *Elife*, 8, e41861.
- Formisano, R., D'Ippolito, M., Riseti, M., et al. (2011). 'Vegetative state, minimally conscious state, akinetic mutism and Parkinsonism as a continuum of recovery from disorders of consciousness: an exploratory and preliminary study'. *Functional Neurology*, 26(1), 15–24.
- Frankish, K. (2017). *Illusionism as a Theory of Consciousness*. Exeter: Imprint Academic.
- Frässle, S., Sommer, J., Jansen, A., et al. (2014). 'Binocular rivalry: frontal activity relates to introspection and action but not to perception'. *Journal of Neuroscience*, 34(5), 1738–47.
- Fredens, J., Wang, K., de la Torre, D., et al. (2019). 'Total synthesis of *Escherichia coli* with a recoded genome'. *Nature*, 569(7757), 514–18.
- Fried, I., Katz, A., McCarthy, G., et al. (1991). 'Functional organization of human supplementary motor cortex studied by electrical stimulation'. *Journal of Neuroscience*, 11(11), 3656–66.
- Friston, K. J. (2009). 'The free-energy principle: a rough guide to the brain?' *Trends in Cognitive Sciences*, 13(7), 293–301.
- Friston, K. J. (2010). 'The free-energy principle: a unified brain theory?' *Nature Reviews Neuroscience*, 11(2), 127–38.
- Friston, K. J. (2018). 'Am I self-conscious? (Or does self-organization entail self-consciousness?)'. *Frontiers in Psychology*, 9, 579.
- Friston, K. J., Daunizeau, J., Kilner, J., et al. (2010). 'Action and behavior: a free-energy formulation'. *Biological Cybernetics*, 102(3), 227–60.
- Friston, K. J., Thornton, C., & Clark, A. (2012). 'Free-energy minimization and the dark-room problem'. *Frontiers in Psychology*, 3, 130.
- Frith, C. D. (2007). *Making Up the Mind: How the Brain Creates Our Mental World*. Oxford: Wiley-Blackwell.
- Gallagher, S. (2008). 'Direct perception in the intersubjective context'. *Consciousness and Cognition*, 17(2), 535–43.
- Gallese, V., Fadiga, L., Fogassi, L., et al. (1996). 'Action recognition in the premotor cortex'. *Brain*, 119 (Pt 2), 593–609.
- Gallup, G. G. (1970). 'Chimpanzees: self-recognition'. *Science*, 167, 86–7.
- Gallup, G. G., & Anderson, J. R. (2018). 'The "olfactory mirror" and other recent attempts to demonstrate self-recognition in non-primate species'. *Behavioural Processes*, 148, 16–19.
- Gallup, G. G., & Anderson, J. R. (2020). 'Self-recognition in animals: Where do we stand fifty years later? Lessons from cleaner wrasse and other species'. *Psychology of Consciousness: Theory, Research, and Practice*, 7(1), 46–58.
- Gasquet, J. (1991). *Cézanne: A Memoir with Conversations*. London: Thames & Hudson Ltd.
- Gehrlach, D. A., Dolensek, N., Klein, A. S., et al. (2019). 'Aversive state processing in the posterior insular cortex'. *Nature Neuroscience*, 22(9), 1424–37.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum.
- Gidon, A., Zolnik, T. A., Fidzinski, P., et al. (2020). 'Dendritic action potentials and computation in human layer 2/3 cortical neurons'. *Science*, 367(6473), 83–7.
- Gifford, C., & Seth, A. K. (2013). *Eye Benders: The Science of Seeing and Believing*. London: Thames & Hudson.

- Ginsburg, S., & Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge, MA: MIT Press.
- Godfrey-Smith, P. G. (1996). 'Spencer and Dewey on life and mind'. In M. Boden (ed.), *The Philosophy of Artificial Life*, Oxford: Oxford University Press, 314–31.
- Godfrey-Smith, P. G. (2017). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. New York: Farrar, Strauss, and Giroux. (Годфри-Смит П. Чужой разум: осьминоги, море и глубинные истоки сознания. — М.: АСТ, 2020.)
- Goff, P. (2019). *Galileo's Error: Foundations for a New Science of Consciousness*. London: Rider.
- Gombrich, E. H. (1961). *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Ewing, NJ: Princeton University Press.
- Goodstein, D. L. (1985). *States of Matter*. Chelmsford, MA: Courier Corporation.
- Gregory, R. L. (1980). 'Perceptions as hypotheses'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 290(1038), 181–97.
- Grill-Spector, K., & Malach, R. (2004). 'The human visual cortex'. *Annual Review of Neuroscience*, 27, 649–77.
- Haggard, P. (2008). 'Human volition: towards a neuroscience of will'. *Nature Reviews Neuroscience*, 9(12), 934–46.
- Haggard, P. (2019). 'The neurocognitive bases of human volition'. *Annual Review of Psychology*, 70, 9–28.
- Hanlon, J., & Messenger, J. B. (1996). *Cephalopod Behaviour*. Cambridge: Cambridge University Press.
- Harding, D. E. (1961). *On Having No Head*. London: The Shollond Trust.
- Harris, S. (2012). *Free Will*. New York: Deckle Edge.
- Harrison, N. A., Gray, M. A., Gianaros, P. J., et al. (2010). 'The embodiment of emotional feelings in the brain'. *Journal of Neuroscience*, 30(38), 12878–84.
- Harvey, I. (2008). 'Misrepresentations'. In S. Bullock, J. Noble, R. Watson, et al. (Eds.), *Artificial Life XI: Proceedings of the 11th International Conference On the Simulation and Synthesis of Living Systems* (pp. 227–33). Cambridge, MA: MIT Press.
- Hatfield, G. (2002). *Descartes and the Meditations*. Abingdon: Routledge.
- Haun, A. M. (2021). 'What is visible across the visual field?' *Neuroscience of Consciousness*.
- Haun, A. M., & Tononi, G. (2019). 'Why does space feel the way it does? Towards a principled account of spatial experience'. *Entropy*, 21(12), 1160.
- He, K., Zhang, X., Ren, S., et al. (2016). 'Deep residual learning for image recognition'. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Heilbron, M., Richter, D., Ekman, M., et al. (2020). 'Word contexts enhance the neural representation of individual letters in early visual cortex'. *Nature Communications*, 11(1), 321.
- Herculano-Houzel, S. (2009). 'The human brain in numbers: a linearly scaled-up primate brain'. *Frontiers in Human Neuroscience*, 3, 31.
- Herculano-Houzel, S. (2016). *The Human Advantage: A New Understanding of How Our Brain Became Remarkable*. Cambridge, MA: MIT Press.
- Hochner, B. (2012). 'An embodied view of octopus neurobiology'. *Current Biology*, 22(20), R887–92.

- Hoel, E. P., Albantakis, L., & Tononi, G. (2013). 'Quantifying causal emergence shows that macro can beat micro'. *Proceedings of the National Academy of Sciences of the USA*, 110(49), 19790–5.
- Hoffman, D. (2019). *The Case against Reality: Why Evolution Hid the Truth from Our Eyes*. London: W. W. Norton & Company.
- Hoffman, D., Singh, M., & Prakash, C. (2015). 'The interface theory of perception'. *Psychonomic Bulletin and Review*, 22, 1480–1506.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. Hohwy, J. (2014). 'The self-evidencing brain'. *Nous*, 50(2), 259–85.
- Hohwy, J. (2020a). 'New directions in predictive processing'. *Mind and Language*, 35(2), 209–23.
- Hohwy, J. (2020b). 'Self-supervision, normativity and the free energy principle'. *Synthese*. doi:10.1007/s11229-020-02622-2.
- Hohwy, J., & Seth, A. K. (2020). 'Predictive processing as a systematic basis for identifying the neural correlates of consciousness'. *Philosophy and the Mind Sciences*, 1(2), 3.
- Hurley, S., & Noë, A. (2003). 'Neural plasticity and consciousness'. *Biology and Philosophy*, 18, 131–68.
- Husserl, E. (1960 [1931]). *Cartesian Meditations: An Introduction to Phenomenology*. The Hague: Nijhoff. (Гуссерль Э. Картезианские медитации. — М.: Академический проект, 2010.)
- Inagaki, K., & Hatano, G. (2004). 'Vitalistic causality in young children's naive biology'. *Trends in Cognitive Sciences*, 8(8), 356–62.
- James, W. (1884). 'What is an emotion?'. *Mind*, 9(34), 188–205. James, W. (1890). *The Principles of Psychology*. New York: Henry Holt.
- Jao Keehn, R. J., Iversen, J. R., Schulz, I., et al. (2019). 'Spontaneity and diversity of movement to music are not uniquely human'. *Current Biology*, 29(13), R621–R622.
- Jensen, F. V. (2000). *Introduction to Bayesian Networks*. New York: Springer.
- Jensen, M. P., Jamieson, G. A., Lutz, A., et al. (2017). 'New directions in hypnosis research: strategies for advancing the cognitive and clinical neuroscience of hypnosis'. *Neuroscience of Consciousness*, 3(1), nix004.
- Kail, P. J. E. (2007). *Projection and Realism in Hume's Philosophy*. Oxford: Oxford University Press.
- Kandel, E. R. (2012). *The Age of Insight: The Quest to Understand the Unconscious in Art, Mind, and Brain, from Vienna 1900 to the Present*. New York: Random House. (Кандель Э. Век самопознания. Поиски бессознательного в искусстве и науке с начала XX века до наших дней. — М.: Corpus, 2016.)
- Kelz, M. B., & Mashour, G. A. (2019). 'The biology of general anesthesia from paramecium to primate'. *Current Biology*, 29(22), R1199–R1210.
- Kessler, M. J., & Rawlins, R. G. (2016). 'A seventy-five-year pictorial history of the Cayo Santiago rhesus monkey colony'. *American Journal of Primatology*, 78(1), 6–43.
- Khuong, T. M., Wang, Q. P., Manion, J., et al. (2019). 'Nerve injury drives a heightened state of vigilance and neuropathic sensitization in *Drosophila*'. *Science Advances*, 5(7), eaaw4099.
- Kirchhoff, M. (2018). 'Autopoiesis, free-energy, and the life-mind continuity thesis'. *Synthese*, 195(6), 2519–40.

- Kirchhoff, M., Parr, T., Palacios, E., et al. (2018). 'The Markov blankets of life: autonomy, active inference and the free energy principle'. *Journal of the Royal Society Interface*, 15(138), 20170792.
- Koch, C. (2019). *The Feeling of Life Itself: Why Consciousness Is Widespread But Can't Be Computed*. Cambridge, MA: MIT Press.
- Kohda, M., Hotta, T., Takeyama, T., et al. (2019). 'If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals?' *PLoS Biology*, 17(2), e3000021.
- Kornhuber, H. H., & Deecke, L. (1965). ['Changes in the brain potential in voluntary movements and passive movements in man: readiness potential and reafferent potentials']. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284, 1–17.
- Konkoly, K. R., Appel, K., Chabani, E., et al. (2021). 'Real-time dialogue between experimenters and dreamers during REM sleep'. *Current Biology*, 31(7), 1417–27.
- Kuhn, G., Amlani, A. A., & Rensink, R. A. (2008). 'Towards a science of magic'. *Trends in Cognitive Sciences*, 12(9), 349–54.
- Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Philosophical Papers*. Cambridge: Cambridge University Press.
- La Mettrie, J. O. de (1748). *L'Homme machine*. Leiden: Luzac.
- Lau, H., & Rosenthal, D. (2011). 'Empirical support for higher-order theories of conscious awareness'. *Trends in Cognitive Sciences*, 15(8), 365–73.
- LeDoux, J. (2012). 'Rethinking the emotional brain'. *Neuron*, 73(4), 653–76.
- LeDoux, J. (2019). *The Deep History of Ourselves: The Four-billion-year Story of How We Got Conscious Brains*. New York, NY: Viking.
- LeDoux, J., Michel, M., & Lau, H. (2020). 'A little history goes a long way toward understanding why we study consciousness the way we do today.' *Proceedings of the National Academy of Sciences of the USA*, 117(13), 6976–84.
- Lemon, R. N., & Edgley, S. A. (2010). 'Life without a cerebellum'. *Brain*, 133 (Pt 3), 652–4.
- Lenggenhager, B., Tadi, T., Metzinger, T., et al. (2007). 'Video ergo sum: manipulating bodily self-consciousness'. *Science*, 317(5841), 1096–9.
- Lettvin, J. Y. (1976). 'On seeing sidelong'. *The Sciences*, 16, 10–20.
- Libet, B. (1985). 'Unconscious cerebral initiative and the role of conscious will in voluntary action'. *Behavioral and Brain Sciences*, 8, 529–66.
- Libet, B., Wright, E. W., Jr., & Gleason, C. A. (1983). 'Preparation-or intention-to-act, in relation to pre-event potentials recorded at the vertex'. *Electroencephalography and Clinical Neurophysiology*, 56(4), 367–72.
- Lipton, P. (2004). *Inference to the Best Explanation*. Abingdon: Routledge.
- Liscovitch-Brauer, N., Alon, S., Porath, H. T., et al. (2017). 'Trade-off between transcriptome plasticity and genome evolution in cephalopods'. *Cell*, 169(2), 191–202 e111.
- Livneh, Y., Sugden, A. U., Madara, J. C., et al. (2020). 'Estimation of current and future physiological states in insular cortex'. *Neuron*, 105(6), 1094–1111.e10.
- Luppi, A. I., Craig, M. M., Pappas, I., et al. (2019). 'Consciousness-specific dynamic interactions of brain integration and functional diversity'. *Nature Communications*, 10(1), 4616.
- Lush, P. (2020). 'Demand characteristics confound the rubber hand illusion'. *Collabra Psychology*, 6, 22.

- Lush, P., Botan, V., Scott, R. B., et al. (2020). 'Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion'. *Nature Communications*, 11(1), 4853.
- Lyamin, O. I., Kosenko, P. O., Korneva, S. M., et al. (2018). 'Fur seals suppress REM sleep for very long periods without subsequent rebound'. *Current Biology*, 28(12), 2000–2005 e2002.
- Makari, G. (2016). *Soul Machine: The Invention of the Modern Mind*. London: W. W. Norton.
- Marken, R. S., & Mansell, W. (2013). 'Perceptual control as a unifying concept in psychology'. *Review of General Psychology*, 17(2), 190–95.
- Markov, N. T., Vezoli, J., Chameau, P., et al. (2014). 'Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex'. *Journal of Comparative Neurology*, 522(1), 225–59.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.
- Mashour, G. A., Roelfsema, P., Changeux, J. P., et al. (2020). 'Conscious processing and the global neuronal workspace hypothesis'. *Neuron*, 105(5), 776–98.
- Massimini, M., Ferrarelli, F., Huber, R., et al. (2005). 'Breakdown of cortical effective connectivity during sleep'. *Science*, 309(5744), 2228–32.
- Mather, J. (2019). 'What is in an octopus's mind?'. *Animal Sentience*, 26(1), 1–29.
- Maturana, H., & Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: D. Reidel.
- McEwan, I. (2000). *Atonement*. New York: Anchor Books.
- McGinn, C. (1989). 'Can we solve the mind-body problem?'. *Mind*, 98(391), 349–66.
- McGrayne, S. B. (2012). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven, CT: Yale University Press.
- McLeod, P., Reed, N., & Dienes, Z. (2003). 'Psychophysics: how fielders arrive in time to catch the ball'. *Nature*, 426(6964), 244–5.
- Mediano, P. A. M., Seth, A. K., & Barrett, A. B. (2019). 'Measuring integrated information: comparison of candidate measures in theory and simulation'. *Entropy*, 21(1), 17.
- Mele, A. (2009). *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.
- Melloni, L., Schwiedrzik, C. M., Muller, N., et al. (2011). 'Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness'. *Journal of Neuroscience*, 31(4), 1386–96.
- Merker, B. (2007). 'Consciousness without a cerebral cortex: a challenge for neuroscience and medicine'. *Behavioral and Brain Sciences*, 30(1), 63–81; discussion 81–134.
- Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge & Kegan Paul. (Мерло-Понти М. Феноменология восприятия. — М.: Наука, 1999.)
- Merleau-Ponty, M. (1964). 'Eye and mind'. In J. E. Edie (ed.), *The Primacy of Perception*, Evanston, IL: Northwestern University Press, 159–90.
- Messenger, J. B. (2001). 'Cephalopod chromatophores: neurobiology and natural history'. *Biological Reviews of the Cambridge Philosophical Society*, 76(4), 473–528.
- Metzinger, T. (2003a). *Being No One*. Cambridge, MA: MIT Press. Metzinger, T. (2003b). 'Phenomenal transparency and cognitive self-reference'. *Phenomenology and the Cognitive Sciences*, 2, 353–93.

- Metzinger, T. (2021). 'Artificial suffering: An argument for a global moratorium on synthetic phenomenology'. *Journal of Artificial Intelligence and Consciousness*, 8(1), 1–24.
- Monroe, R. (1971). *Journeys out of the Body*. London: Anchor Press.
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., et al. (2010). 'Willful modulation of brain activity in disorders of consciousness'. *New England Journal of Medicine*, 362(7), 579–89.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). 'The Uncanny Valley'. *IEEE Robotics & Automation Magazine*, 19(2), 98–100.
- Myles, P. S., Leslie, K., McNeil, J., et al. (2004). 'Bispectral index monitoring to prevent awareness during anaesthesia: the p0ware randomised controlled trial'. *Lancet*, 363(9423), 1757–63.
- Naci, L., Sinai, L., & Owen, A. M. (2017). 'Detecting and interpreting conscious experiences in behaviorally non-responsive patients'. *Neuroimage*, 145 (Pt B), 304–13.
- Nagel, T. (1974). 'What is it like to be a bat?' *Philosophical Review*, 83(4), 435–50.
- Nasraway, S. S., Jr., Wu, E. C., Kelleher, R. M., et al. (2002). 'How reliable is the Bispectral Index in critically ill patients? A prospective, comparative, single-blinded observer study'. *Critical Care Medicine*, 30(7), 1483–7.
- Nesher, N., Levy, G., Grasso, F. W., et al. (2014). 'Self-recognition mechanism between skin and suckers prevents octopus arms from interfering with each other'. *Current Biology*, 24(11), 1271–5.
- Nin, A. (1961). *Seduction of the Minotaur*. Denver, CO: Swallow Press. (Нин А. Соблазнение Минотавра. — М.: Амфора, 2005).
- O'Regan, J. K. (2011). *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*. Oxford: Oxford University Press.
- O'Regan, J. K., & Noë, A. (2001). 'A sensorimotor account of vision and visual consciousness'. *Behavioral and Brain Sciences*, 24(5), 939–73; discussion 973–1031.
- Orne, M. T. (1962). 'On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications'. *American Psychologist*, 17, 776–83.
- Owen, A. M. (2017). *Into the Grey Zone: A Neuroscientist Explores the Border between Life and Death*. London: Faber & Faber.
- Owen, A. M., Coleman, M. R., Boly, M., et al. (2006). 'Detecting awareness in the vegetative state'. *Science*, 313(5792), 1402.
- Palmer, C. E., Davare, M., & Kilner, J. M. (2016). 'Physiological and perceptual sensory attenuation have different underlying neurophysiological correlates'. *Journal of Neuroscience*, 36(42), 10803–12.
- Palmer, C. J., Seth, A. K., & Hohwy, J. (2015). 'The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism'. *Consciousness and Cognition*, 36, 376–89.
- Panksepp, J. (2004). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.
- Panksepp, J. (2005). 'Affective consciousness: Core emotional feelings in animals and humans'. *Consciousness and Cognition*, 14(1), 30–80.
- Park, H. D., & Blanke, O. (2019). 'Coupling inner and outer body for self-consciousness'. *Trends in Cognitive Sciences*, 23(5), 377–88.

- Park, H. D., & Tallon-Baudry, C. (2014). 'The neural subjective frame: from bodily signals to perceptual consciousness'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641), 20130208.
- Parvizi, J., & Damasio, A. (2001). 'Consciousness and the brainstem'. *Cognition*, 79(1–2), 135–60.
- Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press. (Пенроуз Р. Новый ум короля. — М.: URSS, 2022.)
- Pepperberg, I. M., & Gordon, J. D. (2005). 'Number comprehension by a grey parrot (*Psittacus erithacus*), including a zero-like concept'. *Journal of Comparative Psychology*, 119(2), 197–209.
- Pepperberg, I. M., & Shive, H. R. (2001). 'Simultaneous development of vocal and physical object combinations by a grey parrot (*Psittacus erithacus*): bottle caps, lids, and labels'. *Journal of Comparative Psychology*, 115(4), 376–84.
- Petkova, V. I., & Ehrsson, H. H. (2008). 'If I were you: perceptual illusion of body swapping'. *PLoS One*, 3(12), e3832.
- Petzschner, F. H., Weber, L. A., Wellstein, K. V., et al. (2019). 'Focus of attention modulates the heartbeat evoked potential'. *Neuroimage*, 186, 595–606.
- Petzschner, F. H., Weber, L. A. E., Gard, T., et al. (2017). 'Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis'. *Biological Psychiatry*, 82(6), 421–30.
- Phillips, M. L., Medford, N., Senior, C., et al. (2001). 'Depersonalization disorder: thinking without feeling'. *Psychiatry Research*, 108(3), 145–60.
- Pinto, Y., van Gaal, S., de Lange, F. P., et al. (2015). 'Expectations accelerate entry of visual stimuli into awareness'. *Journal of Vision*, 15(8), 13.
- Pollan, M. (2018). *How to Change Your Mind*. New York, NY: Penguin.
- Portin, P. (2009). 'The elusive concept of the gene'. *Hereditas*, 146(3), 112–17.
- Posada, S., & Colell, M. (2007). 'Another gorilla (*Gorilla gorilla gorilla*) recognizes himself in a mirror'. *American Journal of Primatology*, 69(5), 576–83.
- Powers, W. T. (1973). *Behavior: The Control of Perception*. Hawthorne, NY: Aldine de Gruyter.
- Press, C., Kok, P., & Yon, D. (2020). 'The perceptual prediction paradox'. *Trends in Cognitive Sciences*, 24(1), 13–24.
- Pressnitzer, D., Graves, J., Chambers, C., et al. (2018). 'Auditory perception: Laurel and Yanny together at last'. *Current Biology*, 28(13), R739–R741.
- Raccah, O., Block, N., & Fox, K. (2021). 'Does the prefrontal cortex play an essential role in consciousness? Insights from intracranial electrical stimulation of the human brain'. *Journal of Neuroscience*, 41(1), 2076–87.
- Rao, R. P., & Ballard, D. H. (1999). 'Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects'. *Nature Neuroscience*, 2(1), 79–87.
- Reep, R. L., Finlay, B. L., & Darlington, R. B. (2007). 'The limbic system in mammalian brain evolution'. *Brain, Behavior and Evolution*, 70(1), 57–70.
- Richards, B. A., Lillicrap, T. P., Beaudoin, P., et al. (2019). 'A deep learning framework for neuroscience'. *Nature Neuroscience*, 22(11), 1761–70.
- Riemer, M., Trojan, J., Beauchamp, M., et al. (2019). 'The rubber hand universe: On the impact of methodological differences in the rubber hand illusion'. *Neuroscience and Biobehavioral Reviews*, 104, 268–80.

- Rosas, F., Mediano, P. A. M., Jensen, H. J., et al. (2021). 'Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data'. *PLoS Computational Biology*, 16(12), e1008289.
- Roseboom, W., Fountas, Z., Nikiforou, K., et al. (2019). 'Activity in perceptual classification networks as a basis for human subjective time perception'. *Nature Communications*, 10(1), 267.
- Rousseau, M. C., Baumstarck, K., Alessandrini, M., et al. (2015). 'Quality of life in patients with locked-in syndrome: Evolution over a six-year period'. *Orphanet Journal of Rare Diseases*, 10, 88.
- Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking. (Рассел С. Совместимость: Как контролировать искусственный интеллект. — М.: АНФ, 2020.)
- Sabra, A. I. (1989). *The Optics of Ibn Al-Haytham, Books I–III*. London: The Warburg Institute.
- Schachter, S., & Singer, J. E. (1962). 'Cognitive, social, and physiological determinants of emotional state'. *Psychological Review*, 69, 379–99.
- Schartner, M. M., Carhart-Harris, R. L., Barrett, A. B., et al. (2017a). 'Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin'. *Scientific Reports*, 7, 46421.
- Schartner, M. M., Pigorini, A., Gibbs, S. A., et al. (2017b). 'Global and local complexity of intracranial EEG decreases during NREM sleep'. *Neuroscience of Consciousness*, 3(1), niw022.
- Schartner, M. M., Seth, A. K., Noirhomme, Q., et al. (2015). 'Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia'. *PLoS One*, 10(8), e0133532.
- Schick, N. (2020). *Deep Fakes and the Infocalypse: What You Urgently Need to Know*. Monterey, CA: Monoray.
- Schneider, S. (2019). *Artificial You: AI and the Future of Your Mind*. Princeton, NJ: Princeton University Press. (Шнайдер С. Искусственный ты: Машинный интеллект и будущее нашего разума. — М.: АНФ, 2022.)
- Schurger, A., Sitt, J. D., & Dehaene, S. (2012). 'An accumulator model for spontaneous neural activity prior to self-initiated movement'. *Proceedings of the National Academy of Sciences of the USA*, 109(42), E2904–13.
- Searle, J. (1980). 'Minds, brains, and programs'. *Behavioral and Brain Sciences*, 3(3), 417–57.
- Seth, A. K. (2009). 'Explanatory correlates of consciousness: Theoretical and computational challenges'. *Cognitive Computation*, 1(1), 50–63.
- Seth, A. K. (2010). 'Measuring autonomy and emergence via Granger causality'. *Artificial Life*, 16(2), 179–96.
- Seth, A. K. (2013). 'Interoceptive inference, emotion, and the embodied self'. *Trends in Cognitive Sciences*, 17(11), 565–73.
- Seth, A. K. (2014a). 'Darwin's neuroscientist: Gerald M. Edelman, 1929–2014'. *Frontiers in Psychology*, 5, 896.
- Seth, A. K. (2014b). 'A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synaesthesia'. *Cognitive Neuroscience*, 5(2), 97–118.

- Seth, A. K. (2015a). 'The cybernetic bayesian brain: from interoceptive inference to sensorimotor contingencies'. In J. M. Windt & T. Metzinger (eds), *Open MIND*, Frankfurt am Main: MIND Group, 35(T).
- Seth, A. K. (2015b). 'Inference to the best prediction'. In T. Metzinger & J. M. Windt (eds), *Open MIND*, Frankfurt am Main: MIND Group, 35(R).
- Seth, A. K. (2016a). 'Aliens on earth: What octopus minds can tell us about alien consciousness'. In J. Al-Khalili (ed.), *Aliens*, London: Profile Books, 47–58.
- Seth, A. K. (2016b). 'The real problem'. *Aeon*. aeon.co/essays/the-hard-problem-of-consciousness-is-a-distraction-from-the-real-one.
- Seth, A. K. (2017). 'The fall and rise of consciousness science'. In A. Haag (Ed.), *The Return of Consciousness* (pp. 13–41). Riga: Ax: Son Johnson Foundation.
- Seth, A. K. (2018). 'Consciousness: The last 50 years (and the next)'. *Brain and Neuroscience Advances*, 2, 2398212818816019.
- Seth, A. K. (2019a). 'Being a beast machine: The origins of selfhood in control-oriented interoceptive inference'. In M. Colombo, L. Irvine, & M. Stapleton (eds), *Andy Clark and his Critics*, Oxford: Wiley-Blackwell, 238–54.
- Seth, A. K. (2019b). 'From unconscious inference to the Beholder's Share: Predictive perception and human experience'. *European Review*, 273(3), 378–410.
- Seth, A. K. (2019c). 'Our inner universes'. *Scientific American*, 321(3), 40–47.
- Seth, A. K., Baars, B. J., & Edelman, D. B. (2005). 'Criteria for consciousness in humans and other mammals'. *Consciousness and Cognition*, 14(1), 119–39.
- Seth, A. K., Barrett, A. B., & Barnett, L. (2011a). 'Causal density and integrated information as measures of conscious level'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 369(1952), 3748–67.
- Seth, A. K., Dienes, Z., Cleeremans, A., et al. (2008). 'Measuring consciousness: relating behavioural and neurophysiological approaches'. *Trends in Cognitive Sciences*, 12(8), 314–21.
- Seth, A. K., & Friston, K. J. (2016). 'Active interoceptive inference and the emotional brain'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007.
- Seth, A. K., Izhikevich, E., Reeke, G. N., et al. (2006). 'Theories and measures of consciousness: An extended framework'. *Proceedings of the National Academy of Sciences of the USA*, 103(28), 10799–804.
- Seth, A. K., Millidge, B., Buckley, C. L., et al. (2020). 'Curious inferences: Reply to Sun and Firestone on the dark room problem'. *Trends in Cognitive Sciences*, 24(9), 681–3.
- Seth, A. K., Roseboom, W., Dienes, Z., & Lush, P. (2021). 'What's up with the rubber hand illusion?'. <https://psyarxiv.com/b4qcy/>
- Seth, A. K., Suzuki, K., & Critchley, H. D. (2011b). 'An interoceptive predictive coding model of conscious presence'. *Frontiers in Psychology*, 2, 395.
- Seth, A. K., & Tsakiris, M. (2018). 'Being a beast machine: the somatic basis of selfhood'. *Trends in Cognitive Sciences*, 22(11), 969–81.
- Shanahan, M. P. (2010). *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds*. Oxford: Oxford University Press.
- Shanahan, M. P. (2015). *The Technological Singularity*. Cambridge, MA: MIT Press.
- Sherman, M. T., Fountas, Z., Seth, A. K., et al. (2020). 'Accumulation of salient events in sensory cortex activity predicts subjective time'. www.biorxiv.org/content/10.1101/2020.01.09.900423v4.

- Shigeno, S., Andrews, P. L. R., Ponte, G., et al. (2018). 'Cephalopod brains: an overview of current knowledge to facilitate comparison with vertebrates'. *Frontiers in Physiology*, 9, 952.
- Shugg, W. (1968). 'The cartesian beast-machine in English literature (1663–1750)'. *Journal of the History of Ideas*, 29(2), 279–92.
- Silver, D., Schrittwieser, J., Simonyan, K., et al. (2017). 'Mastering the game of Go without human knowledge'. *Nature*, 550(7676), 354–9.
- Simons, D. J., & Chabris, C. F. (1999). 'Gorillas in our midst: sustained inattention blindness for dynamic events'. *Perception*, 28(9), 1059–74.
- Solms, M. (2018). 'The hard problem of consciousness and the free energy principle'. *Frontiers in Physiology*, 9, 2714.
- Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. London: Profile Books.
- Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Steiner, A. P., & Redish, A. D. (2014). 'Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task'. *Nature Neuroscience*, 17(7), 995–1002.
- Sterling, P. (2012). 'Allostasis: a model of predictive regulation'. *Physiology and Behavior*, 106(1), 5–15.
- Stetson, C., Fiesta, M. P., & Eagleman, D. M. (2007). 'Does time really slow down during a frightening event?' *PLoS One*, 2(12), e1295.
- Stoelb, B. L., Molton, I. R., Jensen, M. P., et al. (2009). 'The efficacy of hypnotic analgesia in adults: a review of the literature'. *Contemporary Hypnosis*, 26(1), 24–39.
- Stoljar, D. (2017). 'Physicalism'. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 edn). plato.stanford.edu/archives/win2017/entries/physicalism/.
- Strawson, G. (2008). *Real Materialism and Other Essays*. Oxford: Oxford University Press.
- Strycker, N. (2014). *The Thing with Feathers: The Surprising Lives of Birds and What They Reveal about Being Human*. New York, NY: Riverhead Books.
- Suárez-Pinilla, M., Nikiforou, K., Fountas, Z., et al. (2019). 'Perceptual content, not physiological signals, determines perceived duration when viewing dynamic, natural scenes'. *Collabra Psychology*, 5(1), 55.
- Sun, Z., & Firestone, C. (2020). 'The dark room problem'. *Trends in Cognitive Sciences*, 24(5), 346–8.
- Sutherland, S. (1989). *International Dictionary of Psychology*. New York: Crossroad Classic.
- Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). 'Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion'. *Neuropsychologia*, 51(13), 2909–17.
- Suzuki, K., Roseboom, W., Schwartzman, D. J., & Seth, A. K. (2017). 'A deep-dream virtual reality platform for studying altered perceptual phenomenology'. *Scientific Reports*, 7(1), 15982.
- Suzuki, K., Schwartzman, D. J., Augusto, R., & Seth, A. K. (2019). 'Sensorimotor contingency modulates breakthrough of virtual 3D objects during a breaking continuous flash suppression paradigm'. *Cognition*, 187, 95–107.
- Suzuki, K., Wakisaka, S., & Fujii, N. (2012). 'Substitutional reality system: a novel experimental platform for experiencing alternative reality'. *Scientific Reports*, 2, 459.

- Swanson, L. R. (2016). 'The predictive processing paradigm has roots in Kant'. *Frontiers in Systems Neuroscience*, 10, 79.
- Teasdale, G. M., & Murray, L. (2000). 'Revisiting the Glasgow Coma Scale and Coma Score'. *Intensive Care Medicine*, 26(2), 153–4.
- Teufel, C., & Fletcher, P. C. (2020). 'Forms of prediction in the nervous system'. *Nature Reviews Neuroscience*, 21(4), 231–42.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.
- Thompson, E. (2014). *Waking, Dreaming, Being: Self and Consciousness in Neuroscience, Meditation, and Philosophy*. New York, NY: Columbia University Press.
- Timmermann, C., Roseman, L., Schartner, M., et al. (2019). 'Neural correlates of the DMT experience assessed with multivariate EEG'. *Scientific Reports*, 9(1), 16324.
- Tong, F. (2003). 'Out-of-body experiences: from Penfield to present'. *Trends in Cognitive Sciences*, 7(3), 104–6.
- Tononi, G. (2008). 'Consciousness as integrated information: a provisional manifesto'. *Biological Bulletin*, 215(3), 216–42.
- Tononi, G. (2012). 'Integrated information theory of consciousness: an updated account'. *Archives italiennes de biologie*, 150(4), 293–329.
- Tononi, G., Boly, M., Massimini, M., et al. (2016). 'Integrated information theory: from consciousness to its physical substrate'. *Nature Reviews Neuroscience*, 17(7), 450–61.
- Tononi, G., & Edelman, G. M. (1998). 'Consciousness and complexity'. *Science*, 282(5395), 1846–51.
- Tononi, G., & Koch, C. (2015). 'Consciousness: here, there and everywhere?'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668).
- Tononi, G., Sporns, O., & Edelman, G. M. (1994). 'A measure for brain complexity: relating functional segregation and integration in the nervous system'. *Proceedings of the National Academy of Sciences of the USA*, 91(11), 5033–7.
- Trujillo, C. A., Gao, R., Negraes, P. D., et al. (2019). 'Complex oscillatory waves emerging from cortical organoids model early human brain network development'. *Cell Stem Cell*, 25(4), 558–69 e557.
- Tschantz, A., Barca, L., Maisto, D., et al. (2021). 'Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference'. <https://www.biorxiv.org/content/10.1101/2021.02.16.431365v1>
- Tschantz, A., Millidge, B., Seth, A. K., et al. (2020a). 'Reinforcement learning through active inference'. doi: <https://arxiv.org/abs/2002.12636>.
- Tschantz, A., Seth, A. K., & Buckley, C. (2020b). 'Learning action-oriented models'. *PLoS Computational Biology*, 16(4), e1007805.
- Tsuchiya, N., Wilke, M., Frässle, S., et al. (2015). 'No-report paradigms: extracting the true neural correlates of consciousness'. *Trends in Cognitive Sciences*, 19(12), 757–70.
- Tulving, E. (1985). 'Memory and consciousness'. *Canadian Psychology*, 26, 1–12.
- Turing, A. M. (1950). 'Computing machinery and intelligence'. *Mind*, 59, 433–60.
- Uexküll, J. v. (1957). 'A stroll through the worlds of animals and men: a picture book of invisible worlds'. In C. Schiller (ed.), *Instinctive Behavior: The Development of a Modern Concept*, New York: International Universities Press, 5.
- van Giesen, L., Kilian, P. B., Allard, C. A. H., et al. (2020). 'Molecular basis of chemotactile sensation in octopus'. *Cell*, 183(3), 594–604 e514.

- van Rijn, H., Gu, B. M., & Meck, W. H. (2014). 'Dedicated clock/timing-circuit theories of time perception and timed performance'. *Advances in Experimental Medicine and Biology*, 829, 75–99.
- Varela, F. J. (1996). 'Neurophenomenology: A methodological remedy for the hard problem'. *Journal of Consciousness Studies*, 3, 330–50.
- Varela, F., Thompson, E., & Rosch, E. (1993). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- Walker, M. (2017). *Why We Sleep*. New York: Scribner. (Уолкер М. Зачем мы спим. Новая наука о сне и сновидениях. — М.: КоЛибри, 2022.)
- Waller, B. (2011). *Against Moral Responsibility*. Cambridge, MA: MIT Press.
- Wearing, D. (2005). *Forever Today: A Memoir of Love and Amnesia*. London: Corgi.
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Weiser, T. G., Regenbogen, S. E., Thompson, K. D., et al. (2008). 'An estimation of the global volume of surgery: a modelling strategy based on available data'. *Lancet*, 372(9633), 139–44.
- Wheeler, J. A. (1989). 'Information, physics, quantum: The search for links'. In *Proceedings III International Symposium on Foundations of Quantum Mechanics, Tokyo*, 354–8.
- Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and Machine*. Cambridge, MA: MIT Press.
- Wiener, N. (1964). *God and Golem, Inc.* Cambridge, MA: MIT Press. (Винер Н. Корпорация «Бог и голем». — М.: АСТ, 2018.)
- Williford, K., Bennequin, D., Friston, K., et al. (2018). 'The projective consciousness model and phenomenal selfhood'. *Frontiers in Psychology*, 9, 2571.
- Winn, J., & Bishop, C. M. (2005). 'Variational message passing'. *Journal of Machine Learning Research*, 6, 661–94.
- Wittmann, M. (2013). 'The inner sense of time: how the brain creates a representation of duration'. *Nature Reviews Neuroscience*, 14(3), 217–23.
- Witzel, C., Racey, C., & O'Regan, J. K. (2017). 'The most reasonable explanation of "the dress": Implicit assumptions about illumination'. *Journal of Vision*, 17(2), 1.
- Xiao, Q., & Gunturkun, O. (2009). 'Natural split-brain? Lateralized memory for task contingencies in pigeons'. *Neuroscience Letters*, 458(2), 75–8.
- Zamariola, G., Maurage, P., Luminet, O., et al. (2018). 'Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations'. *Biological Psychology*, 137, 12–17.
- Zucker, M. (1945), *The Philosophy of American History, vol. 1: The Historical Field Theory*. New York: Arnold-Howard.

[1] Перевод В. Марковой.

[2] Это одна из самых влиятельных работ во всей философии сознания. Нагель пишет: «...тот факт, что некое существо вообще осознает нечто, означает, что быть этим существом на что-то похоже». Nagel (1974), p. 2 (курсив в оригинале). — Здесь и далее примечания автора.

[3] Каждое полушарие коры головного мозга состоит из четырех долей. Спереди расположены лобные доли. Позади них, захватывая также и боковую часть, залегают теменные доли. Ближе к затылку помещаются затылочные, а по бокам — височные. Кто-то выделяет и пятую долю — лимбическую, находящуюся в самой сердцевине мозга.

[4] Хвостом вперед может летать вертолет, но это совсем другое. Кстати, для меня стало неожиданно приятным сюрпризом открытие, что слово «геликоптер» образовано вовсе не от «гели» как Гелиос (солнце) и «коптер», как я всегда считал, а от «гелио» (спираль») и «птерос» (крыло). Так намного логичнее.

[5] Мозг взрослого человека состоит, согласно проведенным подсчетам, из 86 млрд нейронов, число контактов между которыми еще в тысячу раз больше. Если подсчитывать по контакту в секунду, на подсчет уйдет почти 3 млн лет. Кроме того, все отчетливее подтверждается, что даже отдельные нейроны способны в одиночку выполнять крайне сложные функции.

[6] Зрительная кора находится в затылочной доле головного мозга.

[7] Функциональная МРТ (фМРТ) измеряет метаболический сигнал (насыщенность крови кислородом), связанный с нейронной активностью, — этот способ нейровизуализации отличается высоким пространственным разрешением, но о деятельности нейронов позволяет судить лишь косвенно. ЭЭГ измеряет электрические сигналы, порождаемые активностью больших скоплений нейронов вблизи поверхности коры, и дает более непосредственную, чем фМРТ, возможность отслеживать нейронную активность, но пространственное разрешение у нее ниже.

[8] При воображении (и выполнении) плавных движений активизируются такие области коры, как дополнительная моторная область, а при воображении перемещения в пространстве вспыхивают другие — такие как извилина гиппокампа. Анатомически эти области довольно сильно удалены друг от друга. Кроме того, оба задания на воображение закономерно активизируют области, участвующие в восприятии слуховой информации и обработке языковых сигналов.

[9] Рефлективное осмысление сохраняется в редком состоянии «осознанного сновидения», при котором спящий сознает, что ему снится сон, и может произвольно управлять своим поведением. В одном интересном недавнем исследовании ученые общались с «осознанными сновидцами» [см. Konkoly et al. (2021).], используя в качестве канала коммуникации движения глаз — практически как у пациентов с синдромом «запертого человека», описанных выше. Сновидцы-испытуемые решали простые математические примеры и отвечали на различные вопросы, предполагающие ответ «да/нет».

[10] В своем дебютном альбоме 1967 г. «Горилла» группа Bonzo Dog Doo-Dah Band пародировала традиционный джаз, стараясь играть из рук вон плохо. К разговору о гориллах мы в этой книге еще вернемся.

[11] В информатике бит — это базовая единица измерения количества информации.

[12] Привычное, но ошибочное представление о том, что чувств у человека всего пять, восходит к трактату Аристотеля De Anima («О душе»), написанному в середине IV в. до н.э.

[13] Историю термина удастся проследить до семинара, который вел в 1990-х гг. Рамеш Джайн. Мои попытки отыскать более ранние истоки успехом не увенчались.

[14] Цветное изображение знаменитого платья можно найти здесь: https://en.wikipedia.org/wiki/The_dress. Каким его видите вы?

[15] Восприятие, на которое знание не влияет, мы называем «когнитивно непроницаемым».

[16] Рисунки 6 и 7 взяты из Teufel, C., Dakin, S. C. & Fletcher, P. C. (2018), 'Prior object-knowledge sharpens properties of early visual feature detectors', Scientific Reports, 8:10853. Печатаются с разрешения авторов и по лицензии Creative Commons Attribution 4.0 International License, с благодарностью Кристофу Тойфелю.

[17] Иногда вместо прецизионности используется термин «дисперсия». Дисперсия — это прямая противоположность прецизионности: чем выше прецизионность, тем ниже дисперсия.

[18] www.youtube.com/watch?v=vJG698U2Mvo

[19] Для примера рассмотрим такого морского обитателя, как асцидия. На личиночной стадии это примитивное животное имеет хотя и зачаточный, но отчетливо выраженный мозг, помогающий ему отыскать подходящий камень или коралл, на котором оно будет восседать до конца жизни, пропуская через себя воду и отфильтровывая все, что в ней найдется питательного. Обретя подходящее пристанище, личинка переваривает собственный мозг, оставляя только простую нервную систему. Некоторые ученые видят в асцидии метафору самих себя до и после обретения постоянного места в университете.

[20] Левому глазу предъявляется меняющийся рисунок из прямоугольников, у которого постепенно снижается контрастность, а правому глазу — картинка (лицо или дом), у которой контрастность постепенно повышается. То и другое изображение передается соответствующему глазу с компьютерного монитора с помощью зеркального стереоскопа. Перед каждой попыткой обоим глазам предъявляется подсказка в виде слова «лицо» или «дом», создающая у испытуемого ожидание соответствующего объекта.

[21] Силла Блэк — ливерпульская поп-звезда 1960-х гг., позже телезвезда.

[22] www.youtube.com/watch?v=hhXZng6o6Dk

[23] Цветную (и более эффектную) версию этой иллюзии можно найти на www.illusionsindex.org/i/rotating-snakes.

[24] Клетки нашего организма и без всякой телепортации постоянно обновляются — основная часть заменяется новым набором примерно за 10 лет, так что мы представляем собой биологический аналог корабля Тесея. Но это вроде бы не особенно влияет на нашу личностную идентичность.

[25] О двойниковых галлюцинациях поведал публике Ф. М. Достоевский в своей повести 1846 г. «Двойник». Известно, что Достоевский сам страдал от тяжелой формы эпилепсии.

[26] Гиппокамп — небольшая структура изогнутой формы, расположенная глубоко в средней височной доле, считается отвечающей за консолидацию памяти. Названа греческим словом, которое означает «морской конек».

[27] У Декарта была собака по кличке Месье Грат (Царапка), которую он, судя по всему, обожал. Что, правда, не мешало ему препарировать кроликов.

[28] Промежуточное положение между экстероцепцией и интероцепцией занимает проприоцепция, относящаяся к восприятию движений тела и положения его в пространстве (см. главу 5). Интероцепцию важно не путать с интроспекцией, которая означает оценку собственного психического состояния.

[29] Островковая кора получила свое название за сходство с островом посреди обширного «моря» коры.

[30] В сентябре 2020 г. я поднимался на гребень горы Бленкатра под говорящим названием Шарп-Эдж («острая кромка») в английском Озерном крае. И хотя альпинистское снаряжение в таких восхождениях не требуется, пройти Шарп-Эдж довольно непросто. Гребень — это узкая тропка из осыпающегося каменного крошева, по обеим сторонам которой обрываются крутые склоны, и несчастные случаи там не редкость. В этот раз я заметил у подножия гребня перевернутый камень с надписью мелом: «Выйдешь за меня, Мария?» Подозреваю, что

неизвестный автор этой надписи знал об эксперименте Даттона и Арона и извлек максимальную выгоду из ситуации.

[31] Переходя от оценочных теорий сразу к interoцептивному выводу, я перескакиваю через огромный пласт промежуточной работы. Антонио Дамасио, в частности, внес основополагающий вклад в развитие этой концепции, показав, каким образом связаны эмоции и когнитивные процессы и как те и другие зависят от происходящего в организме. Независимо от него схожие идеи, подчеркивающие значимость interoцептивных прогнозов, выдвинула Лиза Фельдман Барретт.

[32] Я имею в виду преимущественно взгляды, бытовавшие после эпохи Просвещения. Более ранние системы мировоззрения, такие как анимизм, усматривали задачи (а также жизнь и дух) у гораздо более широкого круга объектов окружающего мира.

[33] Вы можете задаться вопросом: есть ли разница между тем, чтобы «выступать моделью» и «обладать моделью»? На мой взгляд, о системах, у которых, как у системы Б, имеются явные порождающие модели, способные создавать условные или контрфактуальные предсказания, можно сказать, что они «обладают» моделью. Относительно фиксированные регуляторы, не отличающиеся гибкостью, как простой термостат, действующий на основании обратной связи, подобно системе А, способны лишь «выступать» моделью.

[34] Только учтите, что, если вы воспримете этот совет буквально, мяч угодит вам прямо в переносицу.

[35] Можно зайти и с другой стороны: interoцептивные сигналы систематически остаются «без внимания», чтобы дать interoдействиям возможность регулировать существенные переменные, точно так же, как ослабевают во время внешних действий proprioцептивные сенсорные сигналы — об этом мы говорили в главе 5.

[36] Существует ряд интересных пограничных случаев, которые, хотя и не причисляются обычно к живым системам, свою идентичность тем не менее поддерживают активно — вспомним, например, торнадо или водоворот.

[37] Как может состояние быть одновременно статистически ожидаемым и невероятным? Такое возможно, когда система пребывает лишь в ограниченном наборе или репертуаре состояний — так называемом притягивающем множестве — из огромного числа возможных. Притягивающее множество статистически ожидаемо, поскольку именно там обычно находится система, но при этом оно невероятно, поскольку состояний за пределами этого набора гораздо больше, чем в нем. То есть вероятностей расползтись в месиво гораздо больше, чем вероятностей быть живым.

[38] В термодинамике свободная энергия — это количество энергии, доступное для выполнения работы при постоянной температуре: «свободная» здесь означает «доступная». Та разновидность свободной энергии, о которой идет речь в ПСЭ, называется «вариационной свободной энергией» — это термин из машинного обучения и информатики, но тесно связанный со своим термодинамическим эквивалентом.

[39] Аналогичным ПСЭ примером можно считать сформулированный Гамильтоном «принцип стационарности действия» в физике, с помощью которого выводятся (проверяемые) уравнения движения и даже общей теории относительности.

[40] Стоящие за ПСЭ математические выкладки и концепции понять непросто, даже обладая необходимой подготовкой. Как предупреждает нас вступление к учебнику по статистической механике, «Людвиг Больцман, большую часть жизни изучавший статистическую механику, покончил с собой в 1906 г. То же самое произошло в 1933

г. с его учеником и последователем Паулем Эренфестом. Теперь пришла наша очередь заняться изучением статистической механики». [Goodstein (1985), с. 1.]

[41] Перевод И. Я. Дорониной.

[42] Испытуемого просят согнуть руку в запястье в любой момент по своему выбору и засечь, где находилась вращающаяся на экране осциллографа точка в тот самый миг, когда он почувствовал сознательное побуждение к действию. В это время другие приборы измеряют его мышечную активность (ЭМГ) и активность мозга (ЭЭГ). В нижней части изображения показана типичная средняя кривая ЭЭГ, синхронизированная с началом движения (0 сек). Стрелками обозначен момент сознательного побуждения (А) и начало потенциала готовности (Б).

[43] Волевые действия могут иногда ощущаться как требующие сознательного усилия, «силы воли». Например, чтобы написать эту сноску, мне приходится приложить некоторое волевое усилие. Однако многие волевые действия, совершаемые по собственной инициативе, сознательного усилия не требуют или почти не требуют. Поэтому важно отличать (ощущение) свободы воли от проявления силы воли.

[44] Гераклит: «Нельзя войти дважды в одну и ту же реку, потому что уже не будут прежними ни река, ни входящий».

[45] Согласно принципу, на котором основывается западная юриспруденция, для признания уголовной ответственности требуется и «преступное деяние» (*actus rea*), и «преступный умысел» (*mens rea*). Когда способность человека осуществлять свободу воли — контролировать свои степени свободы — нарушается или подавляется, можно ли говорить о наличии у него преступного умысла? Кто-то, как, например, философ Брюс Уоллер, доказывает, что, раз мы не вольны выбирать себе мозг, само понятие моральной ответственности оказывается несостоятельным. Другая точка зрения, которая импонирует и мне, заключается в том, что после пересечения определенного порога способности контролировать степени свободы нас можно привлекать к ответственности за свои действия. [О несостоятельности моральной ответственности см.: Waller (2011), альтернативную точку зрения см.: Dennett (1984, 2003). С познавательной недавней полемикой на эту тему можно ознакомиться в: Dennett & Caruso (2021).]

[46] Здесь и далее я для краткости использую слово «животное» в значении «животные помимо человека», но мы помним, что человек — тоже животное.

[47] В исследовании 2014 г. крысам нужно было выбрать из нескольких вариантов действия, предполагавшие разное вознаграждение. Когда выбранный вариант приносил менее ценное вознаграждение, крысы были более склонны обращаться к альтернативному, не выбранному варианту. Исследователи интерпретировали это как поведенческий знак сожаления, хотя, что на самом деле чувствовали крысы (если вообще чувствовали), утверждать трудно.

[48] Для обозначения мира, воспринимаемого тем или иным животным, нередко применяется введенный этологом Якобом фон Икскульем термин «умвельт».

[49] Вскоре после нашей поездки остров Кайо-Сантьяго, как и значительную часть Пуэрто-Рико, потрепал ураган «Мария». Большинство обезьян, к счастью, уцелело, однако немало исследовательской инфраструктуры было разрушено. Отснятые нами материалы вошли в документальный фильм 2018 г. *The Most Unknown* («Самое неизведанное»), www.themostunknown.com.

[50] В класс головоногих входят осьминоги, кальмары и каракатицы, а также такие сравнительно простые создания, как моллюск наутилус, и около 800 вымерших видов. На самом деле название это осьминогам не очень подходит, поскольку их конечности больше напоминают не ноги, а руки.

[51] На самом деле все не так уж страшно: осьминоги практически не замечают отсутствие конечности, да и новое щупальце на месте отрезанного отрастает довольно быстро. Но это, конечно, не означает, что можно проводить такие эксперименты без достаточного на то основания.

[52] Синтезированные лица были созданы с помощью thispersondoesnot-exist.com.

[53] Некоторые из этих технологий только кажутся нам новыми. Как недавно подметил мой коллега Риота Канаи, «лошадь — это по сути самоуправляемая лошадь».

Переводчик Мария Десятова

Научный редактор Ольга Ивашкина

Редактор Наталья Нарциссова

Издатель П. Подкосов

Руководитель проекта А. Тарасова

Ассистент редакции М. Короченская

Арт-директор Ю. Буга

Корректоры О. Петрова, Е. Рудницкая, Л. Татнинова

Компьютерная верстка А. Ларионов

Иллюстрация обложки [Shutterstock.com](https://www.shutterstock.com)

Графики и иллюстрации Steve West/Lazy Chief

© Anil Seth, 2021

© Издание на русском языке, перевод, оформление. ООО «Альпина нон-фикшн», 2023

© Электронное издание. ООО «Альпина Диджитал», 2023

Сет А.

Быть собой: Новая теория сознания / Анил Сет; Пер. с англ. — М.: Альпина нон-фикшн, 2023.

ISBN 978-5-0022-3046-4