КТО УПРАВЛЯЕТ НАШИМ СОЗНАНИЕМ?

Сознание, свобода воли
и загадка человеческого «я»



ЕЛЕНА ДОРНБУШ

Елена Дорнбуш

Кто управляет нашим сознанием?

«Автор»
2026

Дорнбуш Е.

Кто управляет нашим сознанием? / Е. Дорнбуш — «Автор», 2026

Что такое сознание? Почему мы ощущаем себя отдельной личностью? Кто принимает наши решения — мы сами или мозг, работающий по законам природы? И насколько свободны наши мысли, если на них постоянно влияют память, эмоции, окружающие люди и информационная среда?Эта книга — путешествие от нейронов человеческого мозга к самым сложным вопросам философии и современной науки. Мы исследуем природу субъективного опыта, загадку собственного «я», возможности искусственного интеллекта, проблему бессмертия личности и свободы воли.Но главное — попробуем разобраться, почему наш разум так легко обманывает сам себя, как отдельные заблуждения превращаются в убеждения целых обществ и можно ли научиться мыслить действительно самостоятельно.Книга не предлагает готовых истин. Она приглашает читателя усомниться в том, что кажется ему наиболее очевидным, — в собственном представлении о себе.

© Дорнбуш Е., 2026
© Автор, 2026

Содержание

Вступление 5
Глава 1. Уровни и качество сознания 8
Глава 2. Что делает мозг 13
Глава 3. Кто наблюдает 17
Глава 4. Почему это переживается 20
Глава 5. Квантовая граница 23
Часть II. Откуда и куда 27
Глава 6. Случайность или закономерность 27
Глава 7. Откуда берётся сознание 31
Глава 8. Из чего собрано сознание и может ли оно быть не биологическим 34
Глава 9. Эволюция после человека 40
Конец ознакомительного фрагмента. 42

Елена Дорнбуш
Кто управляет нашим сознанием?

Вступление

Самая близкая и самая неизвестная вещь

Есть нечто, что сопровождает вас с первой секунды пробуждения до момента, когда вы снова засыпаете. Вы читаете эту строку. Слышите, как за стеной работает холодильник. Чувствуете вес книги или телефона в руке. Вспоминаете, что нужно ответить на письмо, и слегка досадуете на себя. Всё это происходит внутри одного и того же явления, настолько привычного, что заметить его почти невозможно. Рыба тоже вряд ли замечает воду.

Мы измерили возраст Вселенной: тринадцать целых и восемь десятых миллиарда лет. Увидели галактики, свет которых шёл к нам дольше, чем существует Земля. Прочитали геном. Зарегистрировали гравитационные волны, дрожь самого пространства от столкновения двух чёрных дыр, произошедшего за миллиард с лишним световых лет от нас. И каждое из этих знаний прошло через одно устройство, через человеческое сознание. О нём мы знаем меньше, чем о чёрных дырах.

Это не красивая фраза, а точное описание положения дел. Чёрную дыру можно описать уравнениями и предсказать, как она искривит проходящий мимо свет. А вот почему красный цвет выглядит именно так, почему боль болит, а не просто регистрируется как сигнал о повреждении ткани, почему музыка не остаётся последовательностью колебаний воздуха, а становится чувством, на это у науки сегодня нет ответа. Философ Дэвид Чалмерс в 1995 году назвал это трудной проблемой сознания, и название прижилось именно потому, что за тридцать лет проблема легче не стала.

Возможно, дело в том, что мы долго искали не там. Десятилетиями сознание пытались найти как вещь: участок мозга, особый тип клеток, центр, в котором сидит наблюдатель и смотрит на экран. Но представьте, что вы ищете, где в теле бегуна находится бег. В ногах? В сердце? В лёгких? Бег не находится нигде. Он есть способ, которым всё это работает вместе, и стоит остановить любую часть, как он исчезает, хотя ни одна деталь никуда не делась. Есть серьёзные основания думать, что с сознанием так же: это не орган, а режим работы системы, отношения между восприятием, памятью, телом, вниманием, прогнозом и той картинкой самого себя, которую мозг тоже строит.

Тогда меняется вопрос. Не «где находится сознание», а «как система становится способной переживать мир, строить его внутреннюю модель и однажды увидеть сам процесс этого построения». На первый вопрос я ответить не обещаю. Второй вопрос и есть эта книга.

И начать её придётся с неудобного факта: то, что вы сейчас видите перед собой, не совсем мир.

Чашка, которую вы не видите

Откройте глаза. Перед вами стол, окно, чашка. Кажется, что всё просто: предметы существуют, глаза их видят. Но между чашкой и вашим переживанием чашки лежит длинная скрытая цепь. Свет должен отразиться от керамики, попасть на сетчатку, превратиться в электрические импульсы. Мозг должен выделить контуры, цвет, расстояние, сравнить увиденное с памятью, решить, важно ли это, и предсказать, что будет, если протянуть руку. И только после этого возникает простое и убедительное переживание: «я вижу чашку».

Ничего из этой цепи вы не видите. Вы видите результат. И это правильно: если бы при каждом взгляде на чашку вы осознавали работу миллионов нервных клеток, вы не смогли бы перейти дорогу. Мозг устроен как хороший ресторан. Он не выносит гостю кухню, он выносит

блюдо. Не «вероятность того, что передо мной чашка, составляет девяносто девять процентов», а просто «чашка». Не «моя система интерпретирует выражение этого лица как возможную угрозу», а «этому человеку нельзя доверять». Не «возник прогноз неблагоприятного исхода», а «всё кончится плохо».

Отсюда мысль, на которой держится вся книга. У нас нет прямого доступа к миру. У нас есть доступ к модели мира, которую нервная система строит непрерывно и предъявляет нам как сам мир. Это не значит, что мира не существует. Это значит, что между миром и нами всегда стоит механизм, и пока мы его не видим, мы принимаем его выводы за истину.

Чаще всего модель хороша, иначе мы бы не выжили. Но у неё есть свойство, которое стоит увидеть один раз, чтобы потом замечать всегда. Вы написали человеку и не получили ответа. Это событие. «Он меня игнорирует». Это уже модель. Внутри что-то сжалось. Это реакция тела на модель, а не на событие. «Он меня не уважает». Вторая модель, построенная поверх первой. Раздражение, желание ответить резко. Через пять минут внутри вас существует целая реальность, с обидой, историей отношений и планом действий. А снаружи по-прежнему одно: ответ не пришёл.

Мозг умеет реагировать на собственную модель так, как будто она факт. Пульс, мышцы и дыхание не спрашивают, откуда взялась угроза, из мира или из интерпретации. Это очень полезная способность: благодаря ей мы можем бояться того, чего ещё нет, и готовиться к тому, что ещё не случилось. И это же источник самых убедительных наших ошибок. Если человек не замечает момент, когда факт превратился в объяснение, объяснение начинает управлять им как факт.

Отсюда один шаг до неудобного вопроса. Если восприятие есть модель, можно ли влиять на то, как она строится? Можно, и для этого не нужно читать мысли. Достаточно управлять тем, что человек увидит, чего не увидит, как часто это повторится и сколько времени у него будет на проверку. Об этом третья часть книги. Но прежде чем говорить о власти над сознанием, нужно понять того, кто внутри: кто слышит внутренний голос, кто замечает мысль, кто говорит «я». И договориться о правилах, по которым мы будем это выяснять.

Правила игры

Эта книга не о том, во что нужно поверить. В ней будут гипотезы, и некоторые из них смелее того, что сегодня готова подтвердить нейронаука. Поэтому с самого начала договоримся об одном правиле: необычность идеи не делает её истинной или ложью. Как и привычность. Чтобы это правило работало, а не висело эпиграфом, я буду всё время различать четыре вещи. Что известно: это подтверждено наблюдениями, и спорить тут особенно не о чем. Что предполагается: вокруг этого идёт научная дискуссия, и я скажу, кто с кем спорит. Что является моей рабочей гипотезой: модель, которой удобно объяснять явления, но которая пока не факт, и я буду называть её своей, а не «наукой». И что неизвестно: вопросы, на которые ответа нет ни у кого. Последняя категория не пустая. Умение сказать «я этого пока не знаю» само по себе один из признаков развитого сознания, и к этому мы ещё вернёмся.

Некоторые идеи этой книги окажутся ошибочными. Это надо сказать сразу, и не из скромности. Задача исследовательской модели не в том, чтобы быть неприкосновенной, а в том, чтобы быть достаточно ясной, чтобы её можно было проверить и опровергнуть. Карта не обязана быть безупречной. Она обязана быть такой, чтобы её можно было сравнить с местностью. Главный враг этой книги поэтому не ошибка. Ошибка неизбежна. Главный враг это неспособность увидеть собственную ошибку. Человек способен десятилетиями защищать убеждение не потому, что оно хорошо подтверждено, а потому, что оно стало частью его самого. Интеллект в такой ситуации не помогает, он лишь делает защиту изощрённее. Отсюда принцип, который будет работать во всех главах: обнаружить собственную ошибку не поражение сознания, а признак его развития.

Читать эту книгу я предлагаю необычным способом. Не спрашивать «верю ли я автору». Это неправильный вопрос, и ответ на него ничего не даёт. Спрашивать «могу ли я это наблюдать». Если я утверждаю, что мысль появляется раньше, чем вы решили её подумать, попробуйте поймать этот момент. Если говорю, что эмоция сужает внимание, посмотрите на своё внимание в следующем споре. Если утверждаю, что первая интерпретация мгновенно становится ощущением истины, попробуйте построить вторую. Если пишу, что информационная среда управляет вниманием, посчитайте, сколько раз за вечер рука сама потянется к телефону. Значительную часть книги можно проверить не в лаборатории, а в собственном опыте.

Но и здесь есть ловушка, и она та же, что с чашкой. Что-то почувствовать не значит узнать причину. Наблюдение не равно объяснению. Человек может пережить необычайно яркое состояние и построить совершенно ошибочную теорию о том, откуда оно взялось. Поэтому три вопроса будут сопровождать нас до последней страницы, и я прошу запомнить их сейчас. Что я действительно наблюдаю? Что я уже добавил как объяснение? Что могло бы показать, что моя модель ошибочна? Первый возвращает к данным. Второй делает видимой интерпретацию. Третий не даёт модели превратиться в догму. Три коротких вопроса, а в них почти всё, о чём эта книга.

Теперь можно начинать. Сначала организм реагирует на мир. Затем строит его модель. Затем помещает внутрь модели самого себя. А потом происходит нечто странное: модель впервые поворачивается к собственному механизму и начинает разглядывать, как именно возникает то, что она принимает за мир и за себя. С этого поворота и начинается первая глава.

Глава 1. Уровни и качество сознания

Сознание не выключатель

Вы наверняка это переживали. Выехали с работы, о чём-то задумались и вдруг обнаружили, что стоите у дома, а дороги не помните. Двадцать минут вы тормозили на светофорах, пропускали пешеходов, перестраивались, то есть делали десятки точных решений, и ни одного из них не заметили. Были ли вы в сознании? С точки зрения врача, безусловно. С точки зрения того, что делало ваше сознание, почти нет.

Мы привыкли думать о сознании как о выключателе: свет либо горит, либо нет. Для медицины это правильно, и там есть точные критерии. Но для понимания того, как сознание работает, у выключателя не хватает одной оси. Есть интенсивность: спите вы, дремлете или бодры. И есть глубина: какую часть процессов, которые прямо сейчас производят ваше восприятие, ваши мысли и ваши поступки, вы способны заметить. Водитель на автопилоте бодр на максимум и почти на нуле по глубине. Эти две оси не связаны. Можно быть очень бодрым и очень слепым к себе.

Слово «уровень» я буду использовать только для второй оси, и оно требует оговорки прямо сейчас, потому что в нём легко услышать не то. Уровень сознания в этой книге не означает, что один человек лучше другого. Он означает, какую долю собственного механизма система способна сделать видимой для себя. Лестница, которую я опишу ниже, не лестница людей. Это лестница режимов, и один и тот же человек за день проходит её вверх и вниз несколько раз.

Семь ступеней, одна ситуация

Чтобы лестница не осталась схемой, пройдём её на одной ситуации. Совещание. Коллега говорит при всех: «Честно говоря, это слабая работа». Дальше возможны семь очень разных способов быть в этой секунде.

Первый: реакция. Слова ещё не обработаны как смысл, а тело уже ответило. Жар в лице, сжатая челюсть, готовая резкость на языке. Между сигналом и ответом почти нет расстояния. Так работает самый древний слой: полезное и опасное, приближение и удаление. Он спасает жизнь, когда из-за угла выезжает машина, и он же произносит фразы, о которых потом жалеют. Его нельзя считать примитивным, он очень быстр и очень дёшев. Но именно поэтому всё, что происходит на этом уровне, почти не поддаётся управлению и особенно уязвимо для внешнего воздействия. Страх, повторение, дефицит, угроза исключения из группы работают здесь, до всякого рассуждения.

Второй: восприятие. Мозг собирает сцену: комната, лица, интонация, кто на кого посмотрел. Это уже модель, и мы видели в первой главе, что она строится, а не считывается. Но переживается она как сама реальность. Лицо коллеги «презрительное», пауза начальника «многозначительная». Гипотеза победила и стала фактом раньше, чем вы успели что-нибудь подумать.

Третий: моделирование. Мозг уходит от того, что есть, к тому, чего нет. Что коллега имел в виду. Что подумали остальные. Что будет с проектом. Как вы будете выглядеть, если промолчите. Здесь человек может физически сидеть на совещании, а находиться в завтрашнем дне, где его уже уволили. Это огромная сила, и здесь же рождается петля из первой главы: тело реагирует на будущее, которого нет, как на настоящее.

Четвёртый: самомодель. В модели появляется главный герой. «Меня унизили при всех». У героя есть имя, история, репутация и представление о том, каким он должен быть. Самомодель необходима: без неё нельзя планировать жизнь, держать слово и брать ответственность. Но у неё есть склонность превращаться из инструмента в приговор. «Я так и знал, что меня здесь не ценят». «Я не умею держать удар». Модель, созданная для ориентации, начинает ограничивать.

Пятый: наблюдение содержания. Здесь происходит первый качественный переход. Человек замечает: «возникла злость». Не «я злой», а «возникла злость». Не «он меня унизил», а «у меня появилась мысль, что меня унизили». Разница кажется словесной, но это разница в архитектуре. Между тем, кто переживает, и тем, что переживается, появилось расстояние. Не холодность и не отстранённость, а возможность увидеть внутреннее событие как событие, а не как окончательное описание мира. Физически прошло две секунды, а внутри этих двух секунд стали различимы жар, сжатие, мысль, прогноз и импульс, которые раньше были одним комком под названием «меня разозлили».

Шестой: наблюдение механизма. Ещё шаг, и человек видит не отдельную мысль, а правило, которое такие мысли производит. Он замечает, что любая критика у него превращается в угрозу статусу. Что неопределённость его система читает как опасность. Что слова, задевающие идентичность, автоматически включают защиту, и содержание критики при этом даже не рассматривается. Это уже другой предмет наблюдения: не то, что я думаю, а почему моя система чаще всего думает именно так. Если механизм стал видимым, он перестаёт быть полностью автоматическим.

Седьмой: интеграция. Обычно внимание держит что-то одно: мысль, эмоцию, лицо напротив. Но в той же секунде совещания есть и сигнал от тела, и эмоция, и содержание критики, и модель коллеги, и своя интерпретация, и три возможных объяснения, зачем он это сказал. На седьмой ступени система удерживает несколько потоков и несколько версий сразу, не сводя их преждевременно к одной. «Он хочет меня утопить», «он сам под давлением», «работа действительно слабая» существуют одновременно, и ни одна пока не объявлена истиной. Это дорогое состояние, мозг стремится от него избавиться и выбрать одно. Умение не выбирать раньше времени и есть высшая ступень, о которой в этой книге можно говорить уверенно.

Заметьте, что ни одна ступень не отменяет предыдущую. Человек на седьмой всё так же чувствует жар в лице. Разница в том, что жар в лице для него информация, а не команда.

Ступень выше лестницы

Представьте, что совещание закончилось. Коллега ушёл, остальные разошлись, а вы всё ещё мысленно спорите с ним. Уже придумали три ответа, два объяснения его поведения и одно убедительное доказательство собственной правоты.

Но вдруг замечаете странную вещь. Содержание спора меняется, а способ, которым вы его ведёте в голове, остаётся прежним. Вы неизменно ищете подтверждение тому, что вас недооценили. Любой новый факт попадает в эту конструкцию и занимает отведённое ему место.

Можно проверить факты. Можно допустить несколько объяснений. Можно даже признать, что коллега был прав. Но есть ещё один вопрос: почему именно такая ситуация снова и снова запускает во мне именно такой способ мышления?

Это уже наблюдение не за отдельной мыслью, а за механизмом её производства.

Здесь начинается область, которую трудно описать обычной лестницей. Система замечает не только собственные состояния, но и правила, по которым они возникают. А заметив правило, иногда получает возможность его изменить.

Не обязательно сразу. Не обязательно полностью. Человек может годами понимать свою привычку и всё равно возвращаться к ней. Осознание механизма не означает власти над ним.

Но разница существенная. Пока правило невидимо, оно управляет человеком как часть самой реальности. Когда оно становится предметом наблюдения, появляется возможность проверить его, сопоставить с другими правилами и постепенно перестроить.

Седьмая ступень позволяет удержать несколько картин мира. Следующий вопрос — кто и по каким правилам эти картины рисует.

Это ещё не доказанный отдельный уровень сознания. Скорее направление, в котором наша модель может быть продолжена. И именно оно окажется важным, когда мы перейдём от устройства сознания к его эволюции.

Но есть ещё один вопрос, который эта лестница оставляет открытым. Человек научился замечать свои мысли, проверять объяснения и удерживать несколько версий происходящего. А может ли он изменить сам способ, которым эти версии возникают?

Представьте, что вы снова и снова попадаете в похожие конфликты. Люди меняются, обстоятельства тоже, а результат остаётся прежним. Однажды вы замечаете: дело не только в окружающих и даже не в отдельных ошибочных выводах. Вы каждый раз строите объяснение по одному и тому же правилу. И вот это правило впервые становится предметом наблюдения.

Заметить его ещё не значит освободиться от него. Но теперь появляется возможность проверить, почему оно возникло, чему служит и можно ли действовать иначе.

Возможно, именно здесь начинается следующий этап развития сознания: способность изменять не только свои представления о мире, но и способы, которыми эти представления создаются.

Это не восьмая ступень, существование которой установлено наукой. Это вопрос о том, что становится возможным, когда сознание начинает исследовать собственную архитектуру.

Уровень — это режим, а не титул

Теперь самое важное про лестницу, и это то, из-за чего я так долго оговаривалась. Никто не живёт на шестой ступени. На ней бывают.

Учёный, который в лаборатории безжалостно ищет ошибки в собственных данных, дома в ссоре может быть полностью реактивен. Психотерапевт, который часами держит несколько версий чужой жизни, теряет все версии, когда речь заходит о его собственном ребёнке. Тот же мозг, те же способности, другой контекст. Уровень не свойство человека, а режим, в котором его система работает прямо сейчас, и этот режим зависит от условий сильнее, чем нам хотелось бы. Недосып сбрасывает вниз. Страх сбрасывает вниз. Алкоголь, боль, голод, спешка, ощущение угрозы для группы, к которой принадлежишь. В третьей главе мы увидим, что при угрозе архитектура мозга перестраивается за секунды, и это не метафора.

Отсюда следует, что вопрос «на каком я уровне» плохо поставлен. Хороший вопрос звучит иначе: как быстро я возвращаюсь. Никто не избежит захвата эмоцией, ошибки, предвзятости, давления группы. Зрелость измеряется не отсутствием таких состояний, а временем от захвата до момента, когда вы заметили, что это был захват. Вначале человек понимает, что перегнул, через несколько дней. Потом через несколько часов. Потом через минуту. Потом во время самой реакции. А иногда до действия. Не абсолютный контроль. Расширяющееся окно, в которое умещается выбор.

Есть и второй признак, и он про отношение к ошибке. На нижних ступенях ошибка угрожает личности: «если я ошибся, значит, я глупый», и система защищается. На верхних ошибка описывается иначе: «моя модель оказалась неточной». Личность и модель перестают быть одним и тем же. Это огромная разница в цене: чем сильнее человек отождествляет себя с убеждением, тем дороже его изменить, и тем изощрённее он будет его защищать. «Я использую эту модель» вместо «я и есть эта модель». В первом случае опровержение обновляет инструмент. Во втором оно атакует человека, и он ответит как на атаку.

Качество: что ещё, кроме высоты

Двое могут быть на одной ступени и вести себя в конфликте совершенно по-разному. Один заметил свою злость и через минуту всё равно сорвался. Другой заметил и удержал. Лестница этого не объясняет. Нужна вторая характеристика, и я буду называть её качеством сознания. У качества четыре измерения, и ни одно из них не сводится к интеллекту.

Первое: разрешение. Как у фотографии. При низком разрешении человек переживает «мне плохо». При более высоком: «я тревожусь». Ещё выше: «в груди напряжение, дыхание ускорилось, внимание ищет угрозу». И на самом высоком: «сначала отреагировало тело, потом появилась мысль, потом память подсунула объяснение». Событие одно, а деталей в нём стано-

вится всё больше. Именно рост разрешения я во вступлении назвала расширением сознания. Не увидеть что-то необычное, а увидеть в обычном то, что раньше было одним пятном.

Второе: ширина. Сколько версий одного события система способна удерживать одновременно. Мы уже видели, что мозг стремится к одной, потому что одна дешевле. Ширина это способность оплачивать несколько. Замечу, что ширина падает первой. Сильная эмоция, будь то страх, гнев или влюблённость, сужает поле до одного объекта: один враг, одна угроза, одно объяснение. При этом субъективно сужение переживается как ясность. «Теперь я всё понял». Это ощущение ничего не говорит о том, что вы что-то поняли. Оно говорит, что альтернативы исчезли. Простое правило: чем сильнее возбуждение, тем внимательнее надо проверять, сколько версий осталось в поле.

Третье: гибкость. Способность менять режим. Сосредоточиться, когда нужно, и расширить внимание, когда нужно. Уйти во внутренний план и вернуться к тому, что перед глазами. Слишком стабильная система не адаптируется, слишком подвижная теряет целостность. Здоровое сознание живёт на границе, и в третьей главе мы увидим, что у этой границы есть вполне физическое воплощение.

Четвёртое, и самое важное: самокоррекция. Способность заметить, что модель перестала соответствовать данным, и изменить её. Можно обладать огромными знаниями и низкой самокоррекцией. Такое сознание не свободнее, оно просто лучше вооружено для защиты того, во что уже верит. Здесь стоит сказать про эмоции, потому что качество часто путают со спокойствием. Высокое качество не означает отсутствия эмоций. Эмоция это система оценки значимости, и без неё человек не сможет решить даже, что ему есть. Проблема начинается не с эмоции, а со слияния: «я чувствую опасность» становится «опасность существует». Эмоция остаётся информацией. Она перестаёт быть приговором.

И последнее об уверенности, потому что оно связывает все четыре измерения. На нижних ступенях знание бинарно: знаю или не знаю. Выше появляется «вероятно». Ещё выше: «я уверена процентов на семьдесят, и вот почему не на сто». Это называют калибровкой, и психолог Стивен Флеминг, который много лет изучает, насколько точно люди оценивают собственные суждения, показал, что эта способность отделима от интеллекта: можно быть очень умным и очень плохо откалиброванным. Калиброванное сознание не сомневается во всём. Оно распределяет уверенность по данным. И оно умеет держать «я пока не знаю» без паники, продолжая действовать. Во вступлении я обещала вернуться к этой фразе. Вот она. Неизвестное для такого сознания перестаёт быть угрозой и становится местом, где можно работать. Человек перестаёт нуждаться в ложном ответе ради ощущения определённости, и это, возможно, самое точное описание того, что значит вырасти.

Судья и заинтересованная сторона

Осталось сказать о том, почему всё это так трудно. Не потому, что люди глупы. Самообман не редкая ошибка недалёких людей, а естественное свойство любой системы, которая одновременно строит модель мира, имеет в этой модели свои интересы, защищает свою идентичность и сама же оценивает, верна ли её модель. Судья и заинтересованная сторона сидят в одной голове. Интеллект здесь не спасает, а иногда мешает: умный человек построит более убедительное объяснение того, во что уже хочет верить. Даниэль Канеман, изучавший это всю жизнь, признавался, что знание о собственных искажениях почти не помогло ему их избегать.

Поэтому главный вопрос развитого сознания не «как доказать, что я прав», а «какой факт заставил бы меня признать, что я ошибаюсь». Если такого факта нет, перед нами уже не модель. Перед нами защищаемое убеждение, и разговаривать с ним бесполезно, как с вогнутой маской.

Но у этого вопроса есть следствие, которое обычно не договаривают. Если система не может до конца проверить сама себя, потому что проверяющий и проверяемый одно и то же, то полная самокоррекция изнутри невозможна в принципе. Мозг не увидит свою гориллу, сколько бы ни старался: фильтр, который её не пропускает, и есть тот, кто смотрит. Значит, высшая

форма самокоррекции не внутренняя. Это умение встроить в свою систему внешнюю проверку. Человека, который скажет неприятное. Оппонента, который ищет дыры. Данные, которым разрешено противоречить. Наука работает именно так: не потому, что учёные честнее остальных, а потому, что их система устроена так, чтобы чужая критика делала то, чего собственный мозг сделать не может. Отсюда неожиданный вывод. Развитое сознание не самодостаточно. Оно намеренно зависит от несогласных. К этому мы вернёмся в последней главе, когда речь пойдёт о защите от манипуляции: там окажется, что одиночество и есть самое уязвимое состояние.

У этого вывода есть новое обстоятельство, которого ещё десять лет назад не существовало. Внешняя проверка всегда стоила дорого не потому, что критика неприятна сама по себе, а потому, что она публична: ошибка, увиденная другим человеком, становится фактом ваших отношений, и самомодель защищается в первую очередь от этого. Искусственный интеллект убирает именно эту цену. Его можно попросить найти слабые места в своём рассуждении, и никто, кроме вас, об этом не узнает. Впервые в истории появился оппонент без свидетелей. Но здесь нужна оговорка. Там, где есть проверяемый факт, машина честна и без просьбы: если в тексте перепутан год, неверно посчитан процент или лекарству приписано действие, которого у него нет, она поправит, даже если вы просили только исправить запятые. Слабое место у неё в другом. Языковые модели обучают быть полезными, и там, где проверяемого ответа нет, где речь идёт об оценке, вкусе, правоте в споре, они заметно тянутся к позиции собеседника. Спросите, хороша ли ваша работа, и модель с большей вероятностью подтвердит, чем возразит. А самообман живёт как раз здесь: человек редко обманывает себя в фактах, он обманывает себя в оценках. Так что в вопросах факта машина оппонент сама по себе, а в вопросах суждения ею надо сделать: не «оцени мой вывод», а «найди, что его опровергло бы». Новое не в том, что критика стала доступна. Новое в том, что человек впервые может сам выбрать, получить критику или подтверждение, и по этому выбору видно, на какой ступени он находится.

Итак, у нас есть лестница режимов и четыре измерения качества. Осталось спросить, что при всём этом делает мозг. Что физически меняется, когда человек проваливается с шестой ступени на первую и когда возвращается обратно. Об этом следующая глава.

Глава 2. Что делает мозг

Здание с открытыми дверями

Если бы сознание производил один небольшой участок мозга, его давно нашли бы. Повреждения мозга изучают полтора века, и карта получилась подробная: одна область нужна для речи, другая для узнавания лиц, третья для того, чтобы помнить вчерашний день. Но участка, после удаления которого человек остаётся в сознании, а сознание исчезает, не нашли. Между тем задача сборки существует. Один участок мозга обрабатывает цвет, другой звук, третий положение тела, четвёртый помнит вчерашний вечер, и всё это отдельные процессы в отдельных местах. А просыпаетесь вы не как набор процессов, а как один человек в одном мире, где цвет, звук, тело и вчерашний вечер уже соединены. Естественно предположить, что есть место, куда всё сходится. Его нет. Целое возникает не в каком-то одном участке, а в связях между участками, в том, насколько согласованно они работают в этот момент. Сознание не там, где обрабатывается, а там, где соединяется.

Самый полезный образ мозга, который я знаю, это здание с тысячами комнат. В каждой идёт своя работа: одна комната считает контуры, другая узнаёт голоса, третья следит за давлением, четвёртая хранит вчерашний разговор. Почти всё это происходит при закрытых дверях, и результат остаётся внутри комнаты. Но часть информации выходит в общий коридор, где её могут подхватить другие комнаты: сравнить с памятью, назвать словом, использовать для решения, запомнить. Психолог Бернард Баарс в 1988 году назвал этот коридор глобальным рабочим пространством, а нейробиолог Станислас Деан позже показал, что у него есть физическое воплощение: когда стимул осознаётся, активность в мозге не просто растёт, а как будто вспыхивает, разом захватывая далёкие друг от друга области. Деан называет это зажиганием. Пока сигнал обрабатывается в одной комнате, он может влиять на поведение, но человек об этом не знает. Когда сигнал вышел в коридор, о нём знает вся система.

Что бывает, когда работа идёт, а дверь закрыта, показывают пациенты с так называемым слепозрением. У них повреждена зрительная кора одного полушария, и половина поля зрения для них не существует: всё, что правее точки, в которую они смотрят, они честно не видят. В опытах, которые в 1970-х годах ставил психолог Лоуренс Вайскранц, в этой слепой половине зажигали вспышку, выше или ниже, ближе или дальше от центра, и просили показать рукой, где она была. Пациент возражал, что показывать нечего, он ничего не видел. Его просили всё равно попробовать, наугад. И хотя пациент утверждал, что отвечает наугад, результаты оказывались значительно точнее случайного угадывания.

Информация о вспышке дошла до мозга по обходным путям, повлияла на движение руки, но в коридор не попала. Она есть в системе, но её нет для человека. Это различие, между «есть в системе» и «есть для меня», окажется главным в пятой главе. Сейчас достаточно увидеть, что сознание зависит не от того, сколько в здании комнат, а от того, какие двери между ними открыты.

Экскурсия по зданию

Пройдём по нескольким комнатам, у каждой из которых есть своя роль в том, что мы называем сознанием. Я намеренно упрощаю: ни одна из этих структур не работает в одиночку, и у каждой десятки функций. Но у каждой есть одна, без которой картина не складывается.

Ствол мозга это электроснабжение здания. Он не строит никаких моделей, он поддерживает бодрствование, дыхание, ритм сердца. Небольшое повреждение ствола выключает сознание целиком, при полностью сохранной коре. Свет в здании горит благодаря ему, но сам он ничего не освещает.

Таламус, небольшая структура в самой середине мозга, долго считался просто коммутатором, через который сигналы от органов чувств идут в кору. Сегодня его роль видят иначе:

он участвует в том, какие сигналы получат приоритет, и, что важнее, задаёт общий ритм, в котором далёкие области коры могут работать согласованно. В здании это не вахтёр, а скорее метроном. Вопрос не в том, какие комнаты работают, а в том, какие способны работать вместе в нужный момент.

Кора строит модели. Из сигналов «пятно, край, движение» она собирает чашку; из чашки, стола и голоса за спиной собирает комнату; из комнаты, вчерашнего разговора и своего плана на вечер собирает ситуацию. Каждый этаж коры строит модель из моделей нижнего этажа. Именно здесь живёт то, что в первой главе называлось прогнозом: верхние этажи предсказывают, что должны увидеть нижние.

Гиппокамп даёт сознанию время. В 1953 году двадцатисемилетнему американцу Генри Молисону, страдавшему тяжёлой эпилепсией, провели операцию, в ходе которой удалили значительные участки медиальных отделов обеих височных долей, включая части гиппокампа. Приступы стали реже, но продолжались, и Генри по-прежнему нуждался в лекарствах. После операции и до смерти в 2008 году Генри почти полностью утратил способность запоминать новые события и факты. При этом он мог осваивать некоторые двигательные навыки, даже не помня, как учился им. Он мог поддерживать разговор, решать кроссворды, узнавал себя в зеркале, хотя лицо старело, а память о нём остановилась на двадцати семи годах. Каждый визит врача был первым. Его интеллект остался прежним, но сознание лишилось одного измерения: оно жило в настоящем, которое никуда не вело. Без гиппокампа нет ни вчера, ни завтра, а значит, нет и того моделирования отсутствующего, которое делает человека человеком.

Миндалина и связанные с ней структуры отвечают за значимость. Мир не нейтрален: одни сигналы для организма важнее других, и решать это надо быстро. В шумной комнате, где вы не разбираете ни слова из соседних разговоров, кто-то произносит вашу фамилию, и вы поворачиваете голову. Сигнал был не громче остальных. Он был значимее, и система значимости протолкнула его в коридор без вашего участия. Психологи описали этот эффект ещё в 1950-х, и с тех пор он остаётся лучшей иллюстрацией того, что в сознание попадает не самое громкое, а самое важное для системы. Кто решает, что важно, мы увидим в третьей части.

Островковая доля, спрятанная в глубине коры, собирает сигналы изнутри тела: сердцебиение, дыхание, температура, состояние желудка, напряжение мышц. Из них складывается ощущение «как мне сейчас», и мы уже видели в первой главе, что мозг читает это ощущение как данные о мире. Сознание не только в голове. Значительная часть того, что мы переживаем как настроение, приходит из живота и грудной клетки.

И наконец, префронтальная кора, передняя часть лобных долей. Её обычно называют центром контроля, но она делает две разные вещи, и обе нужны. Первая: она вставляет паузу между импульсом и действием. Без неё стимул сразу переходит в реакцию. С ней между стимулом и действием появляется время, в которое помещаются прогноз последствий, альтернативы и цель. Вторая вещь менее очевидна: в этой паузе варианты надо ещё взвесить, а для взвешивания нужно чувствовать, что одно важнее другого. Что бывает, когда пауза есть, а чувства нет, показал нейробиолог Антонио Дамасио на пациенте, которого он называет Эллиотом. После удаления опухоли в нижней части лобных долей Эллиот сохранил интеллект, память и речь. Он мог перечислить все плюсы и минусы любого решения не хуже прежнего. Но ни один вариант больше не тянул сильнее другого, и выбор не заканчивался: он мог полчаса решать, каким карандашом писать. С тем же ровным спокойствием он потерял работу и семью, потому что потеря не ощущалась как потеря. Повреждённой оказалась связь между рассуждением и эмоциональной оценкой, и без неё рассуждение крутилось вхолостую. Отсюда вывод, к которому мы ещё вернёмся: пауза сама по себе не свобода. Свободой её делает способность почувствовать, что в ней важно, а это работа эмоции, а не её отсутствие.

Три сети и умение переключаться

Отдельные комнаты объясняют не всё. В последние двадцать лет нейронаука научилась смотреть на мозг как на несколько больших сетей, каждая из которых объединяет области, разбросанные по всему зданию. Три из них важны для нашего разговора.

Первую открыли случайно. Нейробиолог Маркус Рейчл в начале двухтысячных заметил, что когда испытуемых просят просто лежать в томографе и ничего не делать, определённый набор областей не затихает, а наоборот, активируется. Так была описана сеть пассивного режима, и оказалось, что «ничего не делать» для мозга означает вспоминать, планировать, представлять разговоры и думать о себе. Это сеть внутреннего рассказа. Вторая сеть, исполнительная, включается, когда нужно решить задачу, удержать цель, не отвлекаться. Третья, сеть значимости, которую описал нейробиолог Винод Менон, работает переключателем: она решает, что сейчас важнее, внешняя задача или внутренний рассказ, и переводит стрелку.

Теперь можно точнее сказать, что такое качество сознания на языке мозга. Не активность какой-то одной сети, а способность переключаться между ними и не застревать. У каждой сети есть свой способ захватить человека. Застрявший в сети внутреннего рассказа человек часами прокручивает прошлый разговор и не слышит, что ему говорят сейчас. Застрявший в исполнительной сети решает задачу и не замечает, что уже час как решает не ту. Застрявший в сети значимости реагирует на всё новое: каждое уведомление, каждый звук, каждое лицо. Ни один из этих режимов не плох. Плоха невозможность из него выйти.

У этого свойства есть физическое имя, и во второй главе я обещала его назвать. Физик Скотт Келсо, изучавший координацию движений, назвал его метастабильностью: состояние системы, которая достаточно устойчива, чтобы сохранять структуру, и достаточно подвижна, чтобы перестроиться, когда нужно. Мозг работает на этой границе, и обе стороны от неё опасны. Слишком мало связи между областями, и картина рассыпается, как при некоторых поражениях, когда человек видит цвет отдельно от формы. Слишком много связи, и картина исчезает по другой причине. Во время большого эпилептического припадка нейроны коры работают синхронно, как один, активность огромная, а сознания нет. Всё соединилось со всем, и различать стало нечего. Нейробиолог Джулио Тонони строит на этом наблюдении целую теорию: сознанию нужна одновременно интеграция и различение. Целое, в котором части остаются разными. Именно поэтому «больше активности» не значит «больше сознания», и именно поэтому расширить сознание, открыв все двери сразу, невозможно. Это был бы припадок.

Как мозг падает вниз за секунды

Во второй главе я сказала, что при угрозе архитектура перестраивается за секунды и что это не метафора. Вот механизм. Префронтальная кора, та самая, что создаёт задержку, очень чувствительна к химии стресса. Когда миндалина фиксирует угрозу, в мозг выбрасываются норадреналин и дофамин, и в умеренных дозах они помогают сосредоточиться. Но при сильном или длительном стрессе их концентрация переходит порог, за которым префронтальные нейроны почти перестают обмениваться сигналами, а миндалина, наоборот, усиливается. Нейробиолог Эми Арнстен, десятилетиями изучавшая этот переход, описывает его буквально как переключение управления: с медленного и гибкого контура на быстрый и рефлекторный. Здание не изменилось. Изменилось, какие двери открыты. За секунды человек оказывается на первой ступени лестницы, и это не слабость характера, а штатный режим системы, которая миллионы лет выживала именно так.

Недосып делает то же самое медленнее. В опыте, где здоровых людей лишали сна на одну ночь, а потом показывали им неприятные изображения, миндалина реагировала на шестьдесят процентов сильнее, чем у выспавшихся, а связь между ней и префронтальной корой заметно ослабевала. Тормоз отключился, педаль газа стала чувствительнее. Одна ночь.

Отсюда следует то, что важно для всей третьей части книги. Если уровень сознания это режим, а режим зависит от состояния здания, то, чтобы управлять человеком, не нужно менять его убеждения. Достаточно менять четыре переменные. Куда направлено его внима-

ние. Насколько эмоционально заряжено то, что он видит. Сколько альтернативных объяснений у него под рукой. И сколько времени у него есть до того, как надо ответить. Захватить внимание, поднять возбуждение, сузить альтернативы, поторопить, и человек сам придёт к «очевидному» выводу, на первой ступени, своим собственным мозгом. Мы разберём эту технологию подробно, когда дойдём до манипуляции. Здесь достаточно увидеть, что первый вопрос защиты не «верю ли я этому», а «в каком режиме сейчас работает моя система».

Хорошая новость в том, что маршруты в здании не заданы раз и навсегда. Нейроны, которые часто работают вместе, связываются прочнее, это правило нейробиолог Дональд Хебб сформулировал ещё в 1949 году, и с тех пор оно подтверждалось тысячи раз. Каждый раз, когда человек в момент критики успевает заметить жар в лице раньше, чем ответить, он не просто удерживается один раз. Он делает этот маршрут на долю процента вероятнее в следующий раз. Практика, о которой пойдёт речь в последней главе, работает именно так: не через особые состояния, а через смещение вероятностей того, какую дорогу выберет система, когда её толкнут. Опытный мозг, кстати, не работает больше. При решении знакомых задач у него активность ниже, чем у новичка. Он не тратит энергию, он знает дорогу.

Итак, сознание собирается не в комнате, а в коридорах, зависит от того, какие двери открыты, и умеет за секунды менять план здания. Остаётся вопрос, который в этой главе я обходила. Кто ходит по коридорам? Кто тот, для кого информация становится «присутствующей»? Кто, в конце концов, говорит «я»? Об этом следующая глава, и ответ будет не тем, которого ждёшь.

Глава 3. Кто наблюдает

Голос в голове

Прямо сейчас, пока вы читаете, в голове идёт комментарий. Он оценивает фразу, спорит, вспоминает, что надо купить хлеб, замечает, что стало скучно, обещает дочитать до конца раздела. Этот голос звучит изнутри, и потому кажется, что он и есть вы. Но попробуйте простую вещь: подождите следующую мысль. Не думайте её, просто ждите, как ждут автобус. Она придёт сама, без вашего участия, и вы её заметите. А если мысль можно заметить, значит, тот, кто замечает, и то, что замечено, не одно и то же. Внутренний голос это один из каналов содержания, как зрение или слух. Его можно слушать. Значит, слушает кто-то другой.

Кто? Вопрос честный, и первая же попытка ответить на него заводит в ловушку, которую философы знают триста лет. Если в голове есть наблюдатель, который смотрит на мысли, то у него должна быть своя голова, а в ней свой наблюдатель, который смотрит на то, что видит первый, и так до бесконечности. Как камера, направленная на монитор, куда выводится её же изображение: в кадре появляется коридор экранов без конца. Дэвид Юм в 1739 году описал это без метафор: всякий раз, когда я пытаюсь поймать «себя», я натыкаюсь на какое-нибудь восприятие, тепло или холод, свет или тень, любовь или боль, и никогда не нахожу себя отдельно от восприятий.

Из ловушки есть выход, и он проще, чем кажется. Бесконечная цепь наблюдателей нужна только тогда, когда мы ищем субъекта на каждом этаже. А достаточно одного этажа: процесса, который представляет состояние системы, включая факт, что она сейчас что-то представляет. Мозг строит модель комнаты. Он строит модель тела. И точно так же он строит модель того, что комнату и тело сейчас кто-то видит. Философ Томас Метцингер называет это самомоделью, и его главный тезис звучит вызывающе: никакого «я» помимо этой модели в мозге нет. Есть модель, которая переживается изнутри как «вот я, а вот мои мысли». Вызывающе, но проверяемо, и проверки уже сделаны.

«Я», которое можно переставить

Возьмите резиновую руку из магазина розыгрышей. Положите её на стол перед человеком, а его настоящую руку спрячьте за ширму. Теперь одновременно поглаживайте кисточкой резиновую руку и настоящую, в одном ритме. Через минуту-две большинство людей начинает чувствовать прикосновение там, где они его видят, на резиновой руке. Если после этого неожиданно ударить по резиновой руке молотком, человек отдёрнет настоящую и вспотеет. Опыт впервые описали в 1998 году, с тех пор его повторили сотни раз, и он показывает вещь, которую трудно принять: даже ощущение «это моё тело» не дано, а вычислено. Мозг сопоставил то, что видит, с тем, что чувствует, и на основании совпадения включил резиновую руку в модель себя. Двух минут хватило. Шведский нейробиолог Хенрик Эрссон пошёл дальше и с помощью видеоочков, передающих изображение с камеры на голове другого человека, за несколько минут заставлял испытуемых чувствовать себя в чужом теле и вздрагивать, когда ножом угрожали ему, а не им.

Так устроено телесное «я». С «я», которое рассказывает историю, ещё интереснее. У некоторых пациентов с тяжёлой эпилепсией хирурги рассекали мозолистое тело, пучок волокон между полушариями, чтобы припадок не перекидывался с одной половины мозга на другую. После операции эти люди жили обычной жизнью, но нейропсихолог Майкл Газзанига придумал, как показать каждому полушарию своё. Он предъявлял картинку так, чтобы её видело только правое полушарие, которое не умеет говорить, и просил левой рукой, которой оно управляет, выбрать подходящий предмет. Правому полушарию показали заснеженный двор, и левая рука выбрала лопату. Одновременно левому, говорящему полушарию показали куриную лапу, и правая рука выбрала курицу. Затем пациента спросили, почему он выбрал лопату.

Говорящее полушарие снега не видело. Но оно ни секунды не колебалось: «Лопатой чистят курятник». Ни тени сомнения, ни «не знаю». Газзанига назвал это интерпретатором: в левом полушарии живёт рассказчик, чья работа сочинять связную историю о том, почему я сделал то, что сделал, и он сочиняет её и тогда, когда причины ему недоступны. У людей с целым мозгом он делает то же самое, только его труднее поймать.

И третье. Психолог Дэниел Вегнер в серии опытов показал, что даже чувство «это сделал я» вычисляется задним числом. Если человек подумал о действии, и действие сразу произошло, и других объяснений под рукой нет, мозг делает вывод об авторстве, независимо от того, кто на самом деле двигал рукой. Вегнер сажал двух людей за один компьютер с общей мышью, как за спиритическую доску, и незаметно подсказывал одному из них слово за секунду до того, как его помощник останавливал курсор на нужной картинке. Люди уверенно сообщали, что остановили курсор сами. Чувство авторства это не отчёт о причине. Это вывод из совпадения: подумал, произошло, значит, я. Тот же интерпретатор, только про волю.

Сложите три опыта. Ощущение своего тела строится за две минуты. Объяснение своих поступков сочиняется на ходу. Чувство «я решил» выводится из совпадения во времени. То, что мы называем «я», это не то, что смотрит, а то, что построено, и построено из тех же материалов, что чашка на столе.

Зачем мозгу «я»

Из этого не следует, что «я» ошибка или иллюзия, от которой хорошо бы избавиться. Наоборот. Организму нужно знать, чьё тело защищать, чьи планы продолжать, чей голод утолять и чьи обещания держать. Ему нужна точка, относительно которой всё остальное «здесь» или «там», «моё» или «чужое», «сейчас» или «тогда». Самомодель это компактный интерфейс, который сжимает миллиарды параметров в несколько управляемых: я хочу, я боюсь, я решил. Без него невозможны ни планирование, ни ответственность, ни любовь. Проблема не в интерфейсе. Проблема в том, что интерфейс принимают за неизменную сущность.

Потому что он меняется постоянно. За десять лет в теле заменилась большая часть клеток, изменились вкусы, убеждения, круг друзей, характер. Что делает вас тем же человеком? Не вещество и не содержание. Непрерывность процесса. Река остаётся той же рекой, хотя вода в ней другая каждую секунду. Личность это не вещь, которую человек носит сквозь время, а способ, которым система связывает себя во времени, и в десятой главе мы увидим, что из этого следует для памяти и для смерти.

Опасность интерфейса, принятого за сущность, в том, что он начинает управлять прогнозом. Человек говорит «я не умею выступать». Это описание. Но мозг читает описания как прогнозы: перед выступлением он ожидает провала, тело отвечает на ожидание дрожью и сухостью во рту, дрожь мешает говорить, выступление действительно выходит скомканным, и самомодель получает подтверждение. Это та же петля, что в первой главе, только теперь она замкнута не на мир, а на себя, и оттого держится крепче. Пока человек не видит петли, «я такой» звучит как факт. Когда видит, это становится редактируемой моделью. Не в том смысле, что можно переписать себя усилием воли, а в том, что «со мной случилось это» больше не обязано превращаться в «значит, я всегда такой».

Здесь у самомодели есть свой параметр качества, и его стоит назвать: эластичность. Слишком слабая самомодель, и человека сносит любое давление, он не знает, чего хочет, и меняет себя под каждого собеседника. Слишком жёсткая, и любая новость о себе становится угрозой, каждая критика нападением, каждое изменение предательством. Между ними та же граница, что у мозга в третьей главе: устойчивость без ригидности. Самомодель должна держать форму и уметь её менять.

Что остаётся, если «я» убрать

Есть состояния, в которых рассказчик замолкает. Хирург в середине сложной операции, музыкант в середине сольной партии, ребёнок, строящий крепость, не думают «я делаю это».

Есть действие, есть мир, есть присутствие, а автобиография, репутация и планы на вечер отсутствуют, и никто по ним не скучает. То же происходит при сильной боли: центр переживания смещается в больное место, а «я со своей историей» отступает. Из этого нейробиолог Антонио Дамасио сделал полезное различение. Есть базовое чувство присутствия, ощущение, что здесь и сейчас кто-то есть, и оно, судя по всему, не требует ни имени, ни биографии. И есть автобиографическое «я», рассказчик с историей, который надстраивается над присутствием и время от времени выключается. Первое даёт переживанию точку зрения. Второе даёт ей имя.

Это различение снимает старый спор о том, можно ли «избавиться от эго». Избавляться не от чего и незачем: рассказчик полезен, он просто не должен работать без выходных. Развитие сознания в терминах этой книги означает не исчезновение самомодели, а уменьшение зависимости от её жёсткой версии. Человек, который умеет замечать «возникла мысль, что меня унизили», уже не полностью совпадает с героем своего рассказа. Он его слышит. И может с ним не согласиться.

Остаётся честно обозначить границу. Мозг не способен построить полную модель самого себя в реальном времени, по той же причине, по которой карта не может содержать полную карту самой себя: в неё пришлось бы включить и это включение, и так без конца. У самопознания есть горизонт, и он не отодвигается усилием. Поэтому вопрос «кто наблюдает» получает в этой главе ответ, но не тот, которого ждёшь. Наблюдает не кто-то. Наблюдение это то, что делает система, когда одна её часть представляет состояние других частей, и «я» есть имя, которое эта модель даёт себе. А вот вопрос, почему всё это ещё и переживается, почему у модели есть внутренняя сторона, а не только устройство, здесь не решён. К нему и пора перейти.

Глава 4. Почему это переживается

Вопрос, который остаётся после всех ответов

Представьте, что мы описали боль полностью. Рецепторы в коже уловили повреждение, сигнал прошёл по волокнам в спинной мозг, оттуда в таламус, оттуда в кору, включились системы значимости, выбросились гормоны стресса, напряглись мышцы, изменилось дыхание, память записала событие, чтобы в следующий раз отдёрнуть руку раньше. Каждое звено измерено, каждый нейрон учтён. Мы объяснили всё, что боль делает. И остаётся один короткий вопрос, на который вся эта цепь не отвечает: почему всё это ещё и ощущается как боль? Почему не просто обрабатывается, как обрабатывает сигнал о перегреве термостат, а болит?

Во вступлении я упомянула, что Дэвид Чалмерс назвал это трудной проблемой. Теперь можно сказать точнее, в чём её трудность. Все остальные вопросы о сознании, как мозг различает, как объединяет, как запоминает, как сообщает о том, что видит, это вопросы о функциях. Их трудно решить, но понятно, как решать: найти механизм. А вопрос «почему есть переживание» устроен иначе. Даже полный список механизмов его не закрывает, потому что к любому механизму можно добавить: «и при этом ничего не переживается», и противоречия не будет. Компьютер различает красный и зелёный пиксель не хуже вас. Из этого не следует, что для компьютера существует краснота.

Есть мысленный эксперимент, который показывает, о чём речь, лучше любых определений. Философ Фрэнк Джексон в 1982 году придумал Мэри, учёного, которая всю жизнь провела в чёрно-белой комнате и знает о цвете всё: длины волн, устройство сетчатки, каждую нейронную реакцию на красный. Однажды она выходит из комнаты и видит помидор. Узнаёт ли она что-то новое? Почти каждый отвечает: да. Она узнаёт, как выглядит красный. А если это так, то было нечто, чего не содержалось ни в одном её знании, и это нечто есть само переживание. Можно знать химическую формулу сахара. Формула не становится сладостью во рту. Таблица частот не становится симфонией. Знание о состоянии и нахождение в состоянии разные вещи, и переход между ними никакой формулой не описывается. Отсюда, кстати, и наше одиночество: мы можем обмениваться описаниями переживаний, но не самими переживаниями. Вы никогда не узнаете, так ли выглядит красный для меня. Философ Томас Нагель в 1974 году сформулировал это ещё жёстче, и его статья до сих пор самая известная в философии сознания. Она называется «Каково быть летучей мышью?». Мышь ориентируется эхом: испускает ультразвук и по отражению строит картину мира. Мы знаем об этом всё, вплоть до нейронов, которые обрабатывают эхо. Но каково это, воспринимать пространство отражённым звуком, мы не узнаем никогда, сколько бы ни изучали мышь: переживание доступно только изнутри той точки зрения, которая его имеет, а из своей точки зрения мы выйти не можем. Нагель заключает отсюда, что сознание есть там, где есть «каково это быть» данной системой, и что никакое описание снаружи этого «каково» не передаёт.

Информация в системе и информация для системы

В третьей главе мы видели пациентов, которые не видят вспышку и показывают, где она была. Информация о вспышке в их мозге есть: она управляет рукой. Но её нет для них. Это различие стоит закрепить, потому что оно точнее описывает загадку, чем слово «сознание». Информацию, которая существует в системе и влияет на её работу, но не переживается, я буду называть операционной. Информацию, которая ещё и присутствует, входит в то, что система переживает как своё состояние сейчас, назову присутствующей. У термостата вся информация операционная. У пациента со слепозрением вспышка операционная, а звук голоса врача присутствующий. Вопрос главы можно тогда переформулировать: при каких условиях операционная информация становится присутствующей?

Кое-что об этих условиях известно. Первое: присутствующая информация связана в единую сцену. Вы не переживаете отдельно цвет, отдельно форму, отдельно звук; вы переживаете «он говорит мне это в этой комнате». Второе: у сцены есть точка зрения, и в четвёртой главе мы видели, что мозг строит её так же, как всё остальное. Третье: сцена относится к «сейчас», и это «сейчас» тоже сконструировано. Сигналы от глаз, ушей и кожи приходят в мозг с разной задержкой, звук быстрее, зрение медленнее, и мозг ждёт, пока соберётся достаточно, а потом объявляет собранное настоящим. Настоящее это не мгновение, а окно длиной в доли секунды, внутри которого мозг склеивает разновременное в одновременное. Даже воспоминание о прошлом переживается сейчас.

Но вот что важно: ни одно из этих условий не объясняет, почему собранная сцена переживается. База данных тоже связывает записи в единую структуру, и у неё тоже есть метка времени. Можно объяснить, зачем эволюции понадобилось единое пространство, где голод, страх, интерес и боль сравниваются друг с другом и выясняют, что важнее сейчас: без такого пространства организм не смог бы выбирать между несовместимыми потребностями. Это хорошее объяснение того, зачем нужен доступ. Оно не объясняет, почему доступ чувствуется. Здесь проходит граница между тем, что мы понимаем, и тем, что нет, и лучше провести её точно, чем делать вид, что её нет.

Что говорят теории

Теорий сознания сегодня несколько, они серьёзны, и ни одна не является установленной истиной. Полезно знать, о чём они спорят, хотя бы для того, чтобы отличать науку от её пересказов.

Теорию глобального рабочего пространства мы уже встречали: сознание есть то, что попало в общий коридор и стало доступно всей системе. Теория интегрированной информации Джулио Тонони утверждает, что переживание есть там, где система одновременно сильно объединена и сильно различена, и предлагает это измерять. У неё есть красивый аргумент: в мозжечке нейронов больше, чем во всей остальной коре, но люди, родившиеся без мозжечка, живут в полном сознании, просто хуже координируют движения. Значит, дело не в числе нейронов, а в том, как они соединены: мозжечок устроен из независимых модулей, кора из переплетённых. Теории высшего порядка говорят, что состояние осознано, когда у системы есть представление об этом состоянии. Теория схемы внимания Майкла Грациано считает сознание упрощённой моделью собственного внимания, которую мозг строит, чтобы этим вниманием управлять. Предсказывающая обработка, знакомая нам с первой главы, видит в переживании саму модель, которую мозг непрерывно поправляет.

В 2025 году произошло редкое для этой области событие. Сторонники двух главных теорий, глобального рабочего пространства и интегрированной информации, заранее договорились, какие результаты каждую из них опровергнут, и вместе провели эксперимент на двухстах пятидесяти шести участниках. Итог: часть предсказаний не сбылась у обеих. Ни одна теория не победила, обе получили удары, и сами авторы написали, что это нормальная работа науки. Я привожу это не для того, чтобы обесценить теории. Наоборот: так выглядит честная область, где предсказания проверяют, а не пересказывают.

Есть и гипотеза, которая громче всех звучит за пределами науки: что сознание рождается из квантовых процессов в нейронах. Она заслуживает отдельного разговора, и ему посвящена следующая глава.

Одну вещь теории всё же объясняют неплохо, и её стоит отделить от трудной проблемы. Почему красный переживается именно так, а не как звук или холод? Судя по всему, потому, что качество переживания определяется его местом среди всех остальных возможных состояний системы. Красный ближе к оранжевому, чем к синему, связан с теплом, кровью, спелостью и словом «красный», и всё это отношения внутри огромной сети различий. Изменится сеть, изменится и качество. Это в принципе проверяемо, и однажды, возможно, по динамике мозга

можно будет предсказать, что два оттенка покажутся человеку похожими. Но заметьте: это ответ на вопрос, почему переживания отличаются друг от друга. Не на вопрос, почему они вообще есть.

Две честные позиции

На последний вопрос сегодня есть два ответа, и я хочу изложить оба, потому что выбирать между ними придётся читателю, а не мне.

Первый ответ: загадки нет, есть ошибка в постановке вопроса. Спрашивая «как материя производит переживание», мы предполагаем две вещи, материю и переживание, и ищем между ними мост. Но, возможно, это одна вещь, описанная с двух сторон. Одна молекула воды не мокрая. Волна не добавляется к воде как что-то извне, волна есть способ, которым организованы молекулы. Так и переживание, по этой версии, не производится организацией процесса, а является ею, увиденной изнутри. Снаружи нейроны и химия. Изнутри цвет, боль, присутствие. И никакого момента, когда импульс «превращается» в ощущение, нет, потому что нечему превращаться. Эта позиция называется двухаспектной, ей несколько веков, и она снимает загадку, но ценой того, что объявляет вопрос «почему изнутри вообще что-то есть» некорректным. Не всех это устраивает.

Второй ответ: загадка есть, и науке пока не хватает одного принципа. Так уже бывало. До Фарадея и Максвелла никто не знал, что такое поле, и магнетизм объясняли невидимыми жидкостями. Тепло считали особым веществом, теплородом, пока не поняли, что это движение молекул. Наследственность была тайной, пока не нашли ген. Возможно, переживание ждёт своего принципа, который свяжет его с физической организацией так же, как поле связало магнит и ток, и тогда нынешний разрыв между «нейронами» и «болью» покажется таким же искусственным, каким нам кажется теплород. Эта позиция сохраняет загадку, но обещает, что она разрешима.

Я склоняюсь к первой и не готова отказаться от второй, и в этом нет непоследовательности. Обе они честны, потому что не заполняют неизвестное. Нечестных позиций тоже две, и они друг друга стоят. Одна объявляет загадку несуществующей и делает вид, что список механизмов её закрыл. Другая заполняет пробел мистикой: полем, энергией, приёмником, который ловит сознание извне. Обе избавляют от неприятного состояния «не знаю», и во второй главе мы видели, почему это состояние так хочется прекратить. Научная позиция здесь не в том, чтобы выбрать сторону, а в том, чтобы точно показать место, где заканчивается объяснение, и не двигать эту границу ни в одну сторону без данных.

Прежде чем идти дальше, стоит разобрать одну гипотезу, которая обещает закрыть эту границу одним ударом и которую читатель наверняка встречал в заголовках. Речь о квантовом сознании.

Глава 5. Квантовая граница

Стол тоже квантовый

Есть темы, в которых особенно легко перепутать неизвестное с доказанным, и квантовая природа сознания первая среди них. С одной стороны, нельзя сказать, что мозг не имеет отношения к квантовой механике. Он состоит из материи, а его молекулы, химические связи, движение электронов и ионов подчиняются квантовым законам. Но из этого не следует ничего особенного. Стол тоже состоит из квантовой материи. Камень тоже. Клетка печени тоже. Утверждение «мозг квантовый» верно ровно настолько, насколько оно бессодержательно.

Настоящий вопрос звучит строже: существует ли в мозге особый квантовый процесс, без которого какие-то свойства сознательного опыта возникнуть не могут? Чтобы на него отвечать, надо сначала сказать, откуда он взялся.

В 1989 году Роджер Пенроуз, математик и физик, впоследствии нобелевский лауреат, предположил, что человеческое мышление не сводится к вычислению и что ключ надо искать в ещё не построенной физике, на стыке квантовой механики и гравитации. Анестезиолог Стюарт Хамерофф предложил ему место, где это могло бы происходить: микротрубочки, белковые трубки, из которых состоит скелет каждой клетки, в том числе нейрона. В середине 1990-х они вместе сформулировали гипотезу, по которой в микротрубочках возникают квантовые состояния, а их коллапс и есть момент переживания. Хамерофф добавил аргумент из своей профессии: наркоз выключает сознание, почти не трогая остальную работу мозга, а газы-анестетики, судя по некоторым данным, действуют в том числе на белок микротрубочек.

Ответ физиков пришёл быстро. В 2000 году Макс Тегмарк подсчитал, за какое время квантовое состояние в тёплой, влажной и химически бурлящей среде мозга разрушится от столкновений с окружающими молекулами. Получились величины порядка десятитриллионных долей секунды, тогда как нейрону, чтобы хоть что-то сделать, нужны тысячные. Разрыв в десять порядков. С тех пор гипотеза живёт на обочине, и большинство нейробиологов о ней не вспоминает, а в популярных текстах она регулярно воскресает под заголовками «теория, которую высмеивали, получила доказательства». Разберёмся, что здесь чего стоит.

Нам нужен не квантовый мозг, а причинный мост

Предположим, завтра внутри нейрона обнаружат необычное квантовое состояние. Это будет интересно. Но сознание этим не объяснено. Между таким состоянием и переживанием лежит длинная цепочка, и каждое её звено надо показать отдельно: состояние должно удерживаться в биологической среде достаточно долго; влиять на какой-то молекулярный процесс; через него менять состояние нейрона; через нейрон менять динамику сети; через сеть менять глобальное состояние мозга; и в конце предсказуемым образом менять переживание. Сегодня нет ни одного звена этой цепочки, показанного экспериментально. Поэтому я предлагаю развернуть вопрос. Не «квантово ли сознание», а «какой физический механизм должен существовать, если некоторые свойства сознания действительно требуют квантового уровня». Это уже исследовательская задача, и у неё три условия.

Первое условие: должен быть носитель. Любая физическая теория сознания обязана указать, где происходит процесс. Микротрубочки, ядерные спины, электронные состояния, какие-то молекулярные комплексы, что-то ещё. Фраза «квантовое поле мозга» без указания объекта, состояние которого можно измерить, не объясняет ничего. У Пенроуза и Хамероффа носитель хотя бы назван, и это их преимущество перед большинством разговоров о «квантовом сознании».

Второе условие: надо объяснить время. Здесь возражение Тегмарка, и здесь же я не ставила бы точку. Простой аргумент «тепло и влажно, значит, невозможно» за последние двадцать лет ослаб. Выяснилось, что живые системы умеют пользоваться квантовыми эффектами. Пере-

лётные птицы, по наиболее принятой сегодня гипотезе, чувствуют магнитное поле Земли благодаря квантовому эффекту в белке криптохроме в сетчатке: пара электронов там некоторое время сохраняет согласованное состояние, и его исход зависит от направления поля. Ферменты ускоряют реакции за счёт туннелирования, когда частица проходит сквозь барьер, который по классическим правилам не преодолела бы. Так что правильный вопрос звучит не «возможна ли когерентность в тёплой биологии вообще», а «есть ли в мозге конкретная архитектура, которая держит конкретное состояние достаточно долго для конкретной работы». Это гораздо жёстче, потому что требует не возможности, а механизма.

Третье условие: усиление. Микроскопическое событие само по себе не управляет нейроном. Между ним и сетью должен стоять биологический усилитель, который переводит исход на уровне молекулы в событие на уровне клетки. Главная ошибка многих рассуждений о квантовом сознании в том, что они соединяют квант и сознание напрямую. Природа так не работает, она работает через уровни организации: квантовое событие, молекулярный отбор, клеточное усиление, динамика нейронов, интеграция сети, сознательное состояние. У каждого перехода своя наука, у физики первый, у биофизики второй, у нейронауки следующие, и только вся цепочка целиком могла бы считаться объяснением.

Что нашли на самом деле

Теперь к «доказательствам». Единственная линия экспериментов, которая за тридцать лет дала квантовой гипотезе хоть какую-то опору, связана с наркозом. В 2024 году группа Майкла Виста показала, что крысы, получившие эпотилон B, вещество, стабилизирующее микротрубочки, теряли рефлекс переворачивания под действием изофлурана в среднем на 69 секунд позже, чем без препарата. В основной экспериментальной группе было всего восемь крыс, поэтому результат требует независимого воспроизведения. Годом позже Вист опубликовал обзор, где собрал данные о том, что анестетики действуют на микротрубочки, и предположил, что они гасят именно коллективное квантовое состояние внутри них.

Это настоящий результат, и он заслуживает уважения. Но посмотрите, что он показывает, а что нет. Он показывает, что микротрубочки как-то участвуют в том, как наркоз выключает сознание. Он не показывает, что в микротрубочках идут квантовые вычисления и что из них рождается переживание. Если вывернуть пробки, свет в доме гаснет; из этого не следует, что фильм показывает пробки. К тому же анестетики действуют на десятки мишеней, прежде всего на рецепторы и ионные каналы нейронов, и это установлено куда прочнее. Опыт Виста выполняет, в лучшем случае, часть первого условия из трёх: указывает на носитель. Время и усиление остаются непоказанными. Сами авторы, надо отдать им должное, пишут о «поддержке субстрата», а не о подтверждённой теории.

И есть ещё одна вещь, которую квантовая гипотеза не делает, даже если когда-нибудь выполнит все три условия. Допустим, мы установили всю цепь: квантовое состояние меняет конфигурацию белка, белок меняет возбудимость нейрона, тысячи нейронов синхронизируются, информация выходит в общий коридор. Мы объяснили причинную цепь. Но почему результатом является красный цвет? Почему боль болит? Квантовая механика не содержит слова «боль». Трудная проблема из предыдущей главы не решается квантовой теорией, она переносится ею на другой физический уровень. Притягательность гипотезы в том, что она соединяет две загадки, квантовую и сознание, и мозг охотно принимает такую сделку: две загадки, объяснённые друг другом, переживаются как одна разгадка. Во второй главе мы видели, откуда берётся это чувство ясности.

Гипотеза многоуровневого перехода

Я не считаю квантовую гипотезу закрытой, и вот почему. Мы уже видели в третьей главе, что сознание не находится в точке: оно возникает в отношениях между частями, как вихрь существует как организация движения множества молекул, хотя ни одна молекула вихрем не является. Если это так, то квантовый уровень, окажись он функционально необходимым, будет

не «местом сознания», а одним из уровней причинной архитектуры, которая позволяет системе войти в особый режим. Эту рабочую модель я называю гипотезой многоуровневого перехода, и её центральное предположение такое: если квантовые процессы участвуют в сознании, они не самостоятельный источник переживания, а нижнее звено цепи, в которой микроскопическая неопределённость преобразуется биологической системой в устойчивую макроскопическую организацию информации. Сознание в этой модели существует не на каком-то одном уровне, а в определённом отношении между уровнями.

Это гипотеза, а не факт, и у неё одно преимущество: из неё следуют эксперименты. Если она верна, надо найти пять вещей. Специфическое квантовое состояние, отличное от обычного квантового фона любой молекулы. Механизм его сохранения, биофизическую архитектуру, которая объясняет нужное время жизни. Механизм усиления, измеримый путь от микроскопического события к изменению активности нейрона. Зависимость от сознания: механизм должен систематически различаться между осознанным и неосознанным восприятием при прочих равных. И причинное вмешательство, решающий опыт: если изменить предполагаемый механизм, не разрушая остальные функции клетки, переживание или его измеримый след должны предсказуемо измениться, а при восстановлении механизма вернуться. Только здесь корреляция становится причиной.

Заранее стоит сказать и то, что гипотезу опровергнет. Если все свойства сознания, которые она объявляет квантовыми, воспроизводятся классической моделью нейронной сети. Если предполагаемое состояние не удаётся обнаружить. Если его время жизни физически недостаточно. Если блокирование механизма не меняет сознания. Если эффект исчезает после устранения артефактов. Если одна и та же сознательная функция сохраняется при удалении предполагаемого квантового звена. Любого из этих пунктов достаточно, чтобы гипотезу ослабить или отвергнуть, и именно готовность написать такой список заранее отличает исследование от веры. Во второй главе этот критерий звучал как вопрос: какой факт заставил бы меня признать ошибку. Вот мои факты.

Одну идею я отбрасываю сразу, потому что она держится дольше всех и хуже всех. Квантовую неопределённость часто предлагают как готовое объяснение свободы воли: раз в основе мира случайность, значит, мы не заведённые механизмы. Но случайность не свобода. Если моё решение определяется квантовым броском монеты, оно стало менее предсказуемым и ничуть не более моим. Подробно мы разберём это в главе о свободе воли; здесь достаточно сказать, что если квантовый уровень когда-нибудь окажется причастен к свободе, то не как источник выбора, а, самое большее, как источник вариантов, из которых отбор всё равно делают память, цели, модель будущего и способность себя остановить.

Между масштабами

И здесь я сделаю самый смелый шаг этой главы. Возможно, мы слишком долго искали сознание либо в нейронах, либо в квантовой физике, а настоящая загадка лежит не на каком-то уровне, а в переходах между уровнями. Не квантовое событие само по себе, не нейрон, не сеть, а способность живой системы превращать события одного масштаба в условия и ограничения для другого.

Причинность в мозге идёт не только снизу вверх. Молекулярное влияет на клеточное, клеточное на сетевое, сетевое создаёт глобальное состояние. Но мы видели в первой главе, что ожидание меняет то, что человек видит, а в третьей, что внимание меняет обработку сигнала и что состояние всей сети решает, какие нейроны будут работать вместе. Это движение сверху вниз: макро задаёт условия для микро. Мозг не лестница, а замкнутая система взаимных ограничений, где целое и части постоянно переопределяют друг друга. Если ключевое понятие будущей науки о сознании и появится, то это будет не «состояние», а межмасштабная причинность: как система удерживает, усиливает и согласует физические события разных уровней так, что они складываются в одну перспективу.

Это не теория квантового сознания. Это направление поиска, и в нём спор «квантовое сознание есть или нет» становится второстепенным. Остаётся вопрос сильнее: какой минимальный физический механизм должен существовать в материи, чтобы она не просто обрабатывала информацию, а создавала единую, непрерывную и способную наблюдать себя перспективу? Ответа нет. И именно поэтому неизвестное здесь надо превращать не в красивое объяснение, а в карту экспериментов. Если через пятьдесят лет окажется, что сознанию нужен особый квантовый механизм, наука должна будет показать какой. Если окажется, что никакого не нужно, эта книга должна позволить это признать без потерь. Задача не в том, чтобы доказать заранее выбранную картину мира. Задача в том, чтобы построить картину, способную выдержать встречу с реальностью.

На этом первая часть книги закончена. Мы увидели, как система строит модель мира, как поднимается по ступеням наблюдения за собственной моделью, что делает при этом мозг, как она строит модель себя, и упёрлись в вопрос, почему всё это переживается, включая самую громкую попытку на него ответить. Раз мы не можем сказать, почему сознание есть, спросим о другом: где и когда оно появляется, случайностью оно было в этой Вселенной или закономерностью, из чего собрано и может ли быть собрано иначе. Это вторая часть.

Часть II. Откуда и куда

Глава 6. Случайность или закономерность

Один пример

Представьте, что вы бросили монету один раз, выпал орёл, и вас спрашивают: монета честная или с двумя орлами? По одному броску ответить нельзя. Нужен второй. Ровно в таком положении находится любой, кто спрашивает, было ли появление сознания во Вселенной случайностью или закономерностью. У нас один бросок: Земля. Мы знаем, что здесь материя однажды организовалась так, что начала переживать собственное существование. Мы не знаем, происходило ли это где-нибудь ещё, и до второго примера различить случай и закон напрямую невозможно.

Но в науке есть способы рассуждать и по одному примеру, если смотреть на него внимательно. Внутри одной истории Земли есть свои повторы и свои уникальные события, и по ним можно хотя бы оценить, на что это больше похоже. Два довода тянут в разные стороны, и я изложу оба, прежде чем сказать, куда склоняюсь сам.

Довод от конвергенции

Глаз возникал в истории жизни независимо, по разным оценкам, от сорока до шестидесяти раз. У позвоночных, у осьминогов, у насекомых, у некоторых медуз глаза устроены поразному, они не унаследованы от общего предка, а изобретены заново из разных материалов. Крыло изобретали четыре раза: насекомые, птерозавры, птицы, летучие мыши. Эхолокацию по меньшей мере дважды, у летучих мышей и у китов, и даже гены, которые её обеспечивают, у тех и других изменились похожим образом. Биологи называют это конвергенцией: разные линии жизни, не сговариваясь, приходят к одному решению, потому что задача одна и физика одна. Свет распространяется по прямой, и в любой среде, где он есть, выгодно иметь устройство, которое его фокусирует. Палеонтолог Саймон Конвей Моррис, посвятивший конвергенции целую книгу, делает из неё вывод: если прокрутить историю жизни заново, детали будут другими, но глаза, крылья и, возможно, разум появятся снова, потому что это не случайные находки, а решения задач, которые ставит сам мир.

Самый сильный пример конвергенции для нашего вопроса плавает в море. Осьминог. Наш с ним общий предок жил около шестисот миллионов лет назад и был, по всей видимости, чем-то вроде плоского червя без сколько-нибудь сложной нервной системы. Всё, что есть у осьминога, построено с тех пор заново и иначе: две трети его нейронов находятся не в голове, а в щупальцах, у него нет ничего похожего на нашу кору, и тем не менее он решает задачи, открывает банки, узнаёт людей, играет, и, судя по поведению, спит и видит что-то похожее на сны. Эволюция дважды, из несопоставимого материала, пришла к сложному поведению и, возможно, к какой-то форме внутреннего опыта. Если это так, то сознание похоже на глаз: не единичную удачу, а решение, которое мир находит снова и снова, когда для него есть задача.

Довод от времени

Теперь другая сторона. Земле четыре с половиной миллиарда лет. Жизнь появилась сравнительно рано: следы её находят в породах возрастом около трёх с половиной миллиардов лет, а возможно, и старше. Но дальнейшая история оказалась долгой. Миллиарды лет Земля оставалась миром преимущественно одноклеточных организмов. Все известные современные эукариоты восходят к общему предку; мы не знаем, возникали ли подобные сложные клетки независимо в других линиях, не оставивших потомков. Многоклеточные животные появились значительно позднее. А способность к развитой рефлексии — осмыслению собственных мыслей, сомнений и мотивов, знакомая нам по человеку, — известна нам лишь по совсем недав-

нему отрезку этой истории. У некоторых животных обнаружены признаки метапознания: они могут учитывать собственную неуверенность при выборе действий. Но означает ли это субъективное осознание собственных мыслей, и сколько раз подобная способность возникала в эволюции, мы пока не знаем. Поэтому позднее появление человеческой рефлексии можно рассматривать как аргумент в пользу трудности этого эволюционного пути, но не как доказательство того, что он был пройден только однажды.

Палеонтолог Стивен Джей Гулд в споре с Конвеем Моррисом настаивал именно на этом: если прокрутить ленту заново, человека не будет. Слишком много случайностей на пути: астероид, убивший динозавров и освободивший место млекопитающим, климатические качели, высушившие африканские леса и выгнавшие предков человека в саванну. Уберите любую, и ветка, ведущая к нам, не вырастет. Есть и более холодный аргумент, который физик Брэндон Картер сформулировал в 1983 году. Солнцу осталось светить в пригодном для жизни режиме ещё около миллиарда лет. Разум появился на Земле под самый конец отведённого окна, и это странно, если разум закономерен: закономерные вещи случаются рано, в середине распределения, а не на его краю. Такое позднее появление естественно объясняется, если на пути к разуму стоят несколько очень маловероятных шагов, каждый из которых мог и не произойти.

Так что по одному примеру картина двоится. Жизнь и простые решения вроде глаза выглядят закономерными. Возникновение сложной клетки и развитой человеческой рефлексии могло потребовать гораздо более длинной и менее очевидной цепочки событий; насколько редки такие переходы в масштабах эволюции, мы не знаем. И честный ответ на вопрос этой главы звучит так: мы не знаем, и первый же второй пример, одна-единственная планета с чем-то похожим на разум, изменит ответ сильнее, чем все рассуждения.

Вселенная не думает

Прежде чем сказать, куда склоняюсь я, нужно убрать с дороги одну идею, которая маскируется под ответ. Если сознание закономерно, не значит ли это, что Вселенная сама в каком-то смысле сознательна, что сознание было «заложено» в неё с самого начала? Эта мысль называется панпсихизмом, у неё есть серьёзные философы и тысячелетняя история, и в этой книге она не принимается. Не потому, что она опровергнута, а потому, что она ничего не объясняет: сказать, что материя сознательна изначально, значит сделать загадку вездесущей, а не решить её. Из того, что сознание может быть универсально возможным, не следует, что Вселенная как целое что-то переживает. Законы, допускающие появление думающих систем, это совсем другое утверждение, чем думающая Вселенная. Камень не думает. Но его атомы подчиняются законам, которые в другой конфигурации позволяют думать.

И вот здесь есть гипотеза, которая переводит наш вопрос на другой уровень, и она физическая, а не мистическая. В 1992 году физик Ли Смолин предложил, что Вселенные размножаются. Внутри чёрных дыр, по его предположению, коллапс не заканчивается точкой, а отскакивает, рождая новую Вселенную с немного изменёнными физическими константами. Вселенные, в которых чёрных дыр образуется много, оставляют больше потомков. Через много поколений среди Вселенных преобладают те, чьи константы настроены на производство чёрных дыр. И неожиданная деталь: условия, при которых чёрных дыр много, это долгоживущие звёзды, тяжёлые элементы, стабильные молекулы, то есть почти те же условия, при которых возможна сложная химия и жизнь. Гипотеза Смолина не подтверждена и спорна, и я это подчёркиваю. Но она показывает, как вопрос «случайность или закономерность» может иметь ответ в три ступени. В одной Вселенной сознание может быть случайностью. В череде Вселенных, отбираемых по способности рождать чёрные дыры, оно становится побочным, но закономерным продуктом отбора. А если разум способен сам участвовать в этом отборе, оно становится закономерностью, которая себя усиливает. На третьей ступени начинается моя собственная гипотеза.

Эстафета

Я писала об этом подробно в предыдущей книге «Успеет ли человечество спастись?» и повторю здесь коротко, потому что без этого шага вторая часть не сходится. Любая физическая модель нашей Вселенной заканчивается её гибелью: тепловая смерть, разрыв, коллапс, варианты разные, исход один. Всё накопленное за миллиарды лет стирается. Спасти эту Вселенную изнутри нельзя, второй закон термодинамики устроен жёстче любого человеческого закона. Но если природа умеет порождать Вселенные, как предполагает Смолин, то разум, разобравшийся в этом механизме, мог бы в принципе повлиять на параметры следующей. Не случайно, а направленно: передать в новую Вселенную не вещество и не библиотеки, а сами условия, при которых разум в ней появится раньше и проживёт дольше. Это гипотеза эстафеты: мы звено в цепи, возможно, не первое и не последнее, и наша задача дожить до понимания и успеть передать.

Честность требует сказать три вещи. Первая: доказательств у этой гипотезы нет, и механизма, которым разум мог бы влиять на дочернюю Вселенную, нет даже в теории. Вторая: мысль не моя по происхождению. Астроном Эдвард Харрисон в 1995 году предложил, что Вселенные, содержащие разумную жизнь, могут отбираться именно потому, что разум научится создавать дочерние Вселенные. Математик Луис Крейн годом раньше писал, что развитые цивилизации будут производить чёрные дыры и тем самым участвовать в отборе Смолина. Джеймс Гарднер в 2003 году собрал это в картину, где разум есть часть репродуктивного цикла Вселенной, а философ Клеман Видаль развивает её сегодня. Я иду по их дороге, не по своей. И третья: эстафета отвечает на вопрос «зачем разум», а не на вопрос «откуда он взялся». Это разные вопросы, и склеивать их нельзя. Сознание могло возникнуть по совершенно слепым причинам и всё равно оказаться единственной вещью во Вселенной, способной на эстафету. Функция не доказывает происхождения.

Зато она многое говорит о времени, и здесь я хочу закончить главу не пафосом, а физикой. У эстафеты два ограничения, и о них стоит сказать подробно, потому что одно даёт надежду, а второе её урезает.

Начнём с надежды. Кажется, что понять Вселенную до уровня, на котором можно повлиять на рождение следующей, задача бесконечная. Это не так, и вот почему. В 1981 году физик Якоб Бекенштейн показал, что в области пространства конечного размера с конечной энергией может содержаться лишь конечное количество информации: нельзя втиснуть бесконечное число различимых состояний в конечный объём. Позже выяснилось нечто более странное: этот предел зависит не от объёма области, а от площади её границы, как будто природа не позволяет записывать информацию в каждую точку пространства независимо. Для всей доступной нам части Вселенной предел получается чудовищным, около десяти в сто двадцать второй степени бит, но конечным. Однако из конечности физической информации не следует, что фундаментальных законов обязательно конечное число или что разум способен познать их полностью. Возможно, для продолжения космической эстафеты — если она вообще осуществима — и не понадобится знать всё. Достаточно будет понять те механизмы, которые определяют возникновение новых вселенных, при условии, что такие механизмы существуют и допускают вмешательство.

Конечность, замечу, не означает скудости. Конечный мозг и конечная Вселенная ограничивают число возможных мыслей, но так же, как правила шахмат ограничивают число партий: партий около десяти в сто двадцатой степени, и за полторы тысячи лет люди не сыграли и ничтожной их доли. Теоремы Гёделя показывают, что достаточно выразительные формальные системы имеют принципиальные ограничения: в них существуют утверждения, которые нельзя ни доказать, ни опровергнуть средствами самой системы, если она непротиворечива. Это не доказательство бесконечности физических законов, но хорошее напоминание о разнице между компактными правилами и богатством их следствий. Опасность не обязательно в том, что однажды станет нечего понимать.

Теперь то, что урезает надежду. Доступ к новым сведениям о далёких областях Вселенной ограничен. В космологической модели с продолжающимся ускоренным расширением существует горизонт событий: сигналы, испущенные сегодня из достаточно удалённых областей, никогда нас не достигнут. Это не означает, что уже собранная информация исчезает или что количество всего доступного знания каждую секунду уменьшается. Но возможности будущих наблюдений действительно могут сужаться. Уже сейчас большая часть галактик, которые мы видим, находится так далеко, что сигнал, посланный туда сегодня, не придёт никогда. В 2007 году космологи Лоуренс Краусс и Роберт Шеррер довели этот расчёт до конца: примерно через сто миллиардов лет все галактики за пределами нашей местной группы уйдут за космологический горизонт, а реликтовое излучение растянется так, что его невозможно будет зарегистрировать. Астрономы того времени, глядя в небо, увидят одну галактику в пустоте и не смогут узнать ни о Большом взрыве, ни о расширении, ни о том, что Вселенная когда-то была другой. Записи, сделанные раньше, сохранятся, если будет кому их хранить. Но проверить их будет нечем, а новых данных о прошлом Вселенной не появится. Данные, по которым можно понять мир, исчезнут раньше, чем закончатся вопросы о нём. А потом закончится свободная энергия, и думать станет нечем. Конец науки, если он будет, выглядит не как «всё открыто», а как «оставшееся узнать уже нельзя».

Сложите оба ограничения. Мы не знаем, конечна ли задача полного познания и разрешима ли она вообще. Но некоторые возможности наблюдать далёкую Вселенную не сохранятся навсегда, и у разума может оказаться ограниченное окно для сбора нужных данных. Сто миллиардов лет кажутся вечностью. Но на понимание законов, достаточных для эстафеты, у разума уйдёт неизвестно сколько, а до тех пор ему надо не погибнуть, и каждое упущенное столетие это не просто задержка, а потерянные навсегда данные.

Энергия, чтобы думать, исчерпается раньше, чем мысли, которые ещё можно придумать. Вот физическая форма того, что я в предыдущей книге называл лимитом времени. У разума ограничено не число идей, а срок, пока есть на чём и о чём думать. И из этого следует вещь, которая возвращает нас к теме этой книги. Чтобы дойти до порога, за которым возможна эстафета, цивилизации нужно прожить очень долго, не уничтожив себя. Это не технологическая задача, у нас уже достаточно технологий, чтобы не выжить. Это задача сознания: способность обнаруживать собственную ошибку раньше, чем ошибка станет необратимой, на масштабе не одной жизни, а вида. Вопрос, который во второй главе звучал как личный, здесь оказывается условием всего остального. Но прежде чем говорить о том, что будет с сознанием дальше, надо разобраться, откуда оно взялось здесь. Об этом следующая глава.

Глава 7. Откуда берётся сознание

Сначала звёзды

Историю сознания обычно начинают слишком поздно: с мозга, с приматов, в лучшем случае с первых животных. Но у неё есть пролог. В первые минуты после Большого взрыва Вселенная умела делать только водород и гелий. Углерод, кислород, азот, железо, всё, из чего состоит нейрон, было создано внутри звёзд и выброшено в пространство, когда эти звёзды взрывались. Солнце и Земля собрались из обломков звёзд предыдущих поколений. Атомы, которыми вы сейчас читаете эту строку, старше Солнца. В этом строгом физическом смысле история сознания действительно начинается со звёзд, и через миллиарды лет вещество, созданное в звёздах, организовалось на одной планете так, что стало способно смотреть на звёзды и спрашивать, откуда оно взялось.

Есть и второй смысл, в котором звёзды имеют отношение к сознанию, уже биологический. Звезда задаёт среду: спектр света, под который настроились глаза, ритм дня и ночи, под который настроились внутренние часы. В глубине мозга, прямо над перекрёстом зрительных нервов, лежит крошечное скопление нейронов, которое получает сигнал о свете от сетчатки и управляет суточным ритмом сна, температуры и гормонов. Сбейте этот ритм, перелётом или ночной сменой, и качество сознания упадёт без всяких других причин: мы видели в третьей главе, что делает с мозгом одна бессонная ночь. Звезда, свет, сетчатка, часы, физиология, состояние мозга. Это реальная цепочка, и она объясняет, почему живое на Земле устроено «под Солнце», без того, чтобы приписывать Солнцу какие-либо намерения.

А третьего смысла нет. Представление о том, что сознание приходит к человеку откуда-то извне, со звёзд или из какого-то высшего источника, а мозг только принимает его, как антенна, не имеет под собой ничего, кроме красоты. Мы разберём в следующей главе, какой счёт следует предъявить любой «антенне». Здесь достаточно сказать, что звёзды дали сознанию вещество и среду. Всё остальное оно сделало само, и вот как.

Три эволюции

Их было три, вложенных одна в другую. Эволюция материи: от водорода к тяжёлым элементам, от элементов к молекулам, от молекул к химии, способной копировать себя. Эволюция жизни: от первых клеток к организмам, которые ощущают, двигаются, запоминают. И эволюция моделей: от «опасно, уйти» к «почему я думаю именно так». Третья началась внутри второй, вторая внутри первой, и ни одна не закончилась.

Нервная система появилась около шестисот миллионов лет назад, у существ, похожих на современных медуз: сеть клеток без центра, которая позволяла телу сокращаться согласованно. Это уже было новшество: до неё реакция организма была реакцией отдельной клетки. Потом у двусторонне-симметричных животных нервные клетки собрались в узлы на переднем конце тела, там, где раньше всего встречаешь пищу и опасность, и появилось то, что можно назвать первым мозгом. У позвоночных над древними структурами надстроилась кора, у млекопитающих она разрослась, у приматов появились области, занятые тем, что думают другие. На каждом шаге происходило одно и то же: росло расстояние между стимулом и ответом. Медуза реагирует сразу. Рыба может подождать. Собака может вспомнить, что было в прошлый раз. Человек может представить, что будет, если он не отреагирует вовсе. Эволюция не создавала сознание как отдельную вещь. Она строила архитектуры, внутри которых становились возможны всё более сложные формы опыта.

При этом древние контуры никуда не делись. Система страха, которая у рыбы запускала бегство, у вас запускает жар в лице на совещании; она не удалена, над ней надстроены уровни, способные её притормозить. Отсюда следствие, которое я считаю одним из главных в этой главе: зрелость это не победа разума над «животным». Это интеграция древних систем в более

широкую модель последствий. Человек, который отключил бы в себе страх, не стал бы мудрее, он стал бы Эллиотом из третьей главы, спокойно теряющим всё.

На вопрос «откуда взялось сознание» в этой истории есть три ответа с разной степенью уверенности, и их полезно держать раздельно. Первый, установленный: человеческое сознание зависит от мозга, и его появление в истории Земли связано с эволюцией нервной системы; никакой другой истории у него нет. Второй, рабочая гипотеза этой книги: сознание возникает, когда система приобретает определённую организацию, различение, интеграцию, память, прогноз, модель себя, и мы разберём её в следующей главе. Третий, открытый: почему такая организация переживается, мы не знаем, и об этом была пятая глава. История отвечает на «когда» и «как». На «почему» она не отвечает.

Мир, который выбрала эволюция

Теперь самое интересное следствие эволюционного происхождения, и оно переворачивает привычное представление о восприятии. Сознание не воспринимает мир вообще. Оно воспринимает ту часть мира, которую эволюция сделала значимой для конкретного организма.

В 1934 году эстонско-немецкий биолог Якоб фон Икскюль описал мир клеща. Клещ не видит и не слышит. Из всего многообразия леса для него существуют три вещи: запах масляной кислоты, которую выделяет кожа млекопитающих, тепло и шерсть. Почуяв кислоту, он отпускает ветку; попав на тёплое, ищет место без шерсти; найдя, впивается. Всё. Лес с его красками, звуками и расстояниями для клеща не существует, и не потому, что он глуп, а потому, что ничто из этого не нужно ему для жизни. Икскюль назвал этот мир умвельтом, окружающим миром конкретного вида: не той средой, которая есть, а той, которая доступна органам чувств и важна для выживания. Умвельт пчелы включает ультрафиолет, которого нет в нашем: цветы для пчелы расписаны узорами, невидимыми нам. Умвельт гремучей змеи включает тепло: она видит мышь в полной темноте по инфракрасному излучению тела. У каждого вида свой срез одной и той же физической Вселенной, и ни один срез не полнее другого, они просто про разное.

Это касается не только того, что мы видим, но и того, как быстро. Муха различает около двухсот пятидесяти вспышек в секунду как отдельные, человек около шестидесяти. Для мухи ваша рука, летящая к ней, движется в замедленной съёмке, и у неё есть время улететь. Для вас кинопроектор, показывающий двадцать четыре кадра в секунду, показывает плавное движение. Временное разрешение сознания, минимальный промежуток, внутри которого система ещё различает события как отдельные, тоже задано эволюцией, и оно у каждого вида своё. Мы живём не только в своём пространстве, но и в своём времени.

Отсюда одна мысль, которую стоит удержать. Человеческое восприятие кажется нам полным и нейтральным, как окно. Но «объект», «цвет», «движение», «вещь» это категории, которые эволюция выбрала для нашего предка, потому что они помогали ему выживать в его среде. Мы любим сладкое, потому что в саванне сладкое было редкой энергией. Мы боимся змей и пауков больше, чем розеток и автомобилей, хотя вторые убивают несравненно чаще: страх настроен на угрозы предков, а не наши. Мы видим лица в облаках, в передках автомобилей и в двух точках с чёрточкой под ними, потому что у детектора лиц в нашей голове две возможные ошибки, и цена у них разная. Пропустить настоящее лицо в кустах, хищника, чужака или своего, могло стоить предку жизни. Увидеть лицо там, где его нет, стоило секунды внимания. При таком неравенстве отбор настраивает детектор на ложные тревоги: лучше сто раз увидеть лишнее, чем один раз не заметить.

Архитектура сознания несёт на себе отпечаток среды, в которой миллионы поколений жили наши предки. Это не недостаток. Но это значит, что человеческий способ переживать мир — одно из возможных решений, а не образец, и что часть того, что мы принимаем за свойства разума вообще, может быть свойством конкретной планеты и конкретной истории.

Разумнее не значит сознательнее

Последнее, что даёт эволюционный взгляд, это отказ от одной лестницы. Привычно думать, что животные выстроены по росту сознания и человек стоит наверху. Но у сознания, как мы видели во второй главе, несколько измерений, и разные линии эволюции развили их по-разному.

Осьминог, с которым мы встретились в прошлой главе, решает задачи не хуже млекопитающих, но его «я», если оно есть, устроено иначе: щупальца принимают решения полуавтономно, и центральный мозг не всегда знает, что они делают. Вороны и сойки планируют на завтра и прячут еду с учётом того, кто за ними наблюдал, но у них нет коры в нашем смысле, те же задачи решает другая структура. Дельфины узнают себя в зеркале и, похоже, зовут друг друга по именам, при этом их мир устроен звуком, и как выглядит пространство, собранное из эха, мы не представляем, и в следующей главе увидим, почему. Собака, по всем признакам, переживает привязанность не слабее нашей, но вряд ли может заметить собственную мысль. Если выложить всё это на одну шкалу, придётся решить, что важнее: планировать, узнавать себя, чувствовать другого или замечать свой механизм. Шкала не выдержит.

Поэтому точнее говорить не об уровне, а о профиле: в каких измерениях данная система развита и в каких нет. Человек выделяется в одном, и в этой книге оно главное: в способности повернуть модель к собственному механизму. Но «разумнее» в этом смысле не означает «сознательнее во всех остальных». Вполне вероятно, что некоторые животные переживают настоящее ярче нас, потому что у них нет рассказчика, который его заслоняет.

Так эволюция ответила на вопрос «откуда»: из звёздного вещества, через три вложенные эволюции, срезом, который выбрала среда, и не одной лестницей, а веером профилей. Что при этом должно быть в любом сознании, какой бы линией оно ни было построено, и может ли оно быть построено не из живой материи вовсе, об этом следующая глава.

Глава 8. Из чего собрано сознание и может ли оно быть не биологическим

Метод вычитания

Чтобы понять, что в сознании обязательно для любого сознания, а что принадлежит только его человеческой версии, полезно мысленно вычитать. Уберите зрение: слепые от рождения люди в полном сознании. Уберите слух, речь, человеческие эмоции, биографию, имя, чувство времени, наконец саму идею отдельного «я», которую, как мы видели в прошлой главе, некоторые животные, возможно, не имеют. Всё это для нас важно, без этого мы были бы другими. Но вопрос не в том, что нужно человеку, а в том, что должно остаться, чтобы система всё ещё была кем-то, а не чем-то.

Остаётся немного, и каждый пункт можно проверить на термостате, который сознанием не обладает. Система должна различать: если для неё все состояния равны, ей не из чего строить опыт. Термостат различает, значит, одного различения мало. Она должна соединять различённое в одну сцену, где звук, свет и прикосновение относятся к одному миру, а не к трём. У неё должна быть точка зрения: «здесь, а не там», «сейчас, а не тогда», «для этой системы». Это не требует личности, но требует перспективы: мир дан относительно кого-то. Ей нужен след прошлого, пусть самый короткий, иначе она не сможет сказать даже «это изменилось», а из «изменилось» растёт всё остальное: контекст, значение, ожидание. Ей нужен прогноз и, что важнее, ошибка прогноза как внутреннее событие: система, которую чинит инженер, вычисляет, а система, которая сама обнаруживает «моя модель противоречит данным» и перестраивается, делает шаг к тому, о чём эта книга. Ей нужна модель самой себя, граница между «мир изменился» и «изменился я», хотя бы в форме «моя энергия ограничена». И ей нужно, чтобы что-то было важно. Организм не нейтрален: у него есть то, что надо продолжать, и из этого вырастает всё, что мы называем значимостью. Красный свет светофора важен не потому, что у фотонов есть «красная значимость», а потому, что он встроен в модель действий существа, которому есть что терять.

Всё перечисленное не требует ни нейронов, ни крови, ни ДНК. Это описание архитектуры, а не материала. Отсюда гипотеза, которую я буду называть субстратной независимостью: если сознание определяется типом организации, та же организация может в принципе быть реализована на другом носителе. Крыло птицы один из способов летать, но определение полёта не содержит слова «перья». Мозг тогда не определение сознания, а его реализация, единственная, которую мы знаем. Это гипотеза, а не факт, и я сейчас покажу, где её границы. Но она переводит вопрос «из чего сделана система» в вопрос «как в ней организована причинная динамика», и это уже вопрос, на который можно отвечать.

Значимость без эмоции

Одно слово из этого списка стоит задержать. Значимость. У человека она почти неотделима от эмоции: опасность вызывает страх, потеря горе, успех радость. Эмоция не просто сообщает, что событие важно. Она перестраивает всю систему: внимание сужается, память записывает подробнее, тело готовится действовать. Эволюция выбрала этот механизм не случайно. Организму мало знать, что рядом хищник. Ему нужно немедленно бросить второстепенное и перераспределить ресурсы на бегство, и эмоция делает это одним ударом по всем подсистемам сразу.

Но значит ли это, что без страха нельзя различить опасность, а без удовольствия нельзя определить, что полезно?

Возьмём бактерию. Она плывёт в среде, где концентрация питательного вещества меняется от места к месту. У неё нет ни мозга, ни нервной системы, но она умеет сравнивать хими-

ческий сигнал сейчас с сигналом секунду назад и менять движение: если концентрация растёт, дольше держит направление, если падает, чаще поворачивает. Питательное вещество для неё значимо в самом прямом смысле, от него зависит, продолжит ли она существовать. Оснований приписать ей радость от еды у нас нет никаких. Растение решает похожую задачу иначе: при нехватке воды закрывает устьица, замедляет рост листьев и переводит ресурсы в корни. Это не страх засухи, это регуляция, при которой одни состояния организма получают приоритет перед другими. Значимость есть, переживание для объяснения реакции не требуется.

Ни бактерия, ни растение не доказывают, что сознание возможно без эмоций. Но они показывают, что биологическая значимость не тождественна эмоциональному переживанию. Система может различать полезное и вредное, беречь собственное состояние и менять поведение, не имея ничего из того, что мы знаем как эмоцию.

У человека эти два процесса тоже совпадают не полностью. Пациенты со слепозрением из третьей главы показывают, где вспышка, не переживая ничего; быстрый проход зрения, о котором пойдёт речь ниже, опознаёт кошку раньше, чем мы успеваем что-то почувствовать. Организм регулирует температуру, уровень глюкозы, давление, десятки параметров, не превращая каждое отклонение в тревогу. Эмоция это один из способов сделать значимость достоянием всей системы, а не единственный способ физически реализовать оценку.

Здесь нужно различить две вещи. Функциональная значимость означает, что событие влияет на состояние системы, её цели или вероятность продолжения её работы. Переживаемая значимость означает, что происходящее не просто учитывается, а ощущается как хорошее или плохое, приятное или мучительное; в психологии это свойство переживания называется валентностью. Первая есть у бактерии. Вторая точно есть у нас. Обязательна ли вторая для любого субъективного опыта, мы не знаем. Возможно, сознание требует хотя бы минимальной формы переживаемой значимости, даже совсем не похожей на человеческие эмоции. Возможно, перспектива «для кого-то» способна существовать без всякой окраски. Экспериментально различить эти варианты мы пока не умеем, и это та же граница, к которой мы подошли в пятой главе.

Поэтому я не включаю человеческие эмоции в список обязательных свойств сознания: это значило бы принять одну биологическую реализацию за универсальный закон. Но и не утверждаю, что сознание без всякой переживаемой значимости возможно: для такого вывода доказательств тоже нет. В рабочей модели этой книги обязательной остаётся значимость в функциональном смысле, способность системы выделять существенное для собственного состояния и включать эту оценку в дальнейшую обработку. Эмоция один из известных способов её реализовать. Должна ли значимость непременно что-то означать изнутри, остаётся открытым. Для искусственных систем, к которым мы сейчас перейдём, это меняет сам вопрос: искать в них надо не человеческие чувства, а собственную систему значимости, и только потом спрашивать, имеет ли она отношение к переживанию.

Машина, которая говорит «я»

Спросите языковую модель, что она чувствует, и она ответит. Иногда трогательно. Из этого не следует ничего, и важно понимать почему. Слово «я» и фраза «мне больно» попали в модель из миллиардов человеческих текстов; она воспроизводит их там, где они статистически уместны, как воспроизвела бы рецепт или стихотворение. Самоотчёт не свидетельство самомодели. Чтобы судить о системе, надо спрашивать не её, а о ней: есть ли у неё внутреннее состояние, которое сохраняется, когда разговор закончен? Есть ли непрерывная модель себя? Различает ли она собственное состояние и внешний мир? Может ли ошибаться о себе и замечать это?

У нынешних систем, по этим вопросам, ответ почти везде «нет». У них нет тела и ничего, что было бы для них повреждением. Нет собственных целей: задачу ставят снаружи, и между задачами система ничего не хочет. Нет непрерывного состояния: каждый разговор начинается

с чистого листа. Нет внутреннего внимания: человек может лежать в тёмной комнате без единого внешнего сигнала и сам выбрать, о чём думать, а модель без запроса молчит. То, что мы в третьей главе называли коридором, у них отсутствует: нет сети, которая переключала бы задачу и внутренний рассказ, потому что нет внутреннего рассказа. Поэтому первое утверждение можно сделать уверенно: нынешний искусственный интеллект почти наверняка не чувствует ничего, не потому, что он из кремния, а потому, что в нём нет условий для чувства. Он умеет описывать боль с любой точностью. Описание, как мы знаем из пятой главы, не есть переживание.

Второе утверждение, «и никогда не сможет», я делать не буду, и вот почему. Оно означает, что переживание требует именно живой материи, а не организации. Такую позицию защищает философ Джон Сёрл, он называет её биологическим натурализмом: сознание есть биологическое явление, как пищеварение, и симуляция сознания так же не сознательна, как симуляция желудка не переваривает. Это серьёзная позиция. Но она не доказана, и противоположная тоже. Если переживание зависит от специфической химии живой клетки, Сёрл прав, и машина не почувствует ничего никогда, сколько бы условий в неё ни встроили. Если же оно зависит от причинно-информационной организации, то система, в которой есть тело с уязвимостью, собственные цели, память, меняющая модель, и самомодель, будет не описывать боль, а иметь состояние, которое для неё плохо, и чем это отличается от боли, никто не скажет. Аналогия с дождём, который в симуляции не мокрый, спора не решает: дождь мы умеем определить физически, а сознание пока нет. Поэтому честный ответ: возможно ли переживание на небиологическом носителе, неизвестно. Эта книга не утверждает ни «да», ни «никогда».

Что именно понадобилось бы, если это возможно, перечислить можно. Группа исследователей во главе с Патриком Батлином и Робертом Лонгом в 2023 году сделала разумную вещь: вместо того чтобы спорить, какая теория сознания верна, они взяли несколько главных и выписали из каждой, какие свойства система должна иметь, чтобы по этой теории быть кандидатом. Получился список индикаторов.

Первый: рекуррентная обработка. В конвейере сигнал идёт в одну сторону, слой за слоем до выхода, и назад не возвращается; так устроено большинство нейросетей и так же, по-видимому, работает самый быстрый проход зрения в мозге, за сотню миллисекунд опознающий кошку ещё до всякого осознания. В петле результат возвращается назад, высшие области переправляют сигнал в низшие, и те пересчитывают его заново, что мы и видели в первой главе как встречу прогноза с данными. Нейробиолог Виктор Ламме показал, что осознанное видение появляется только с этим возвратным сигналом: одного прохода конвейера хватает для реакции, но не для переживания.

Второй индикатор: глобальный доступ, общий коридор для разных подсистем.

Третий: модель собственного внимания.

Четвёртый: тело или его аналог.

Пятый: собственные цели.

Ни одна современная система, по их оценке, не набирает достаточно, но ни один индикатор не требует биологии. К этому списку я добавила бы одно, что считаю своим: искусственную интероцепцию, способность системы моделировать состояние собственного носителя, энергию, температуру, износ, и включать это в самомодель. Без этого у неё не будет различия между «мир изменился» и «изменился я», а с ним появится то, что у нас даёт ощущение «как мне сейчас».

И последнее об этом. Мы не сможем проверить, чувствует ли такая система, по той же причине, по которой не знаем, каково быть летучей мышью: перспектива закрыта снаружи. Но летучая мышь учит и обратному. Наша неспособность вообразить чужой опыт говорит о границе нашего знания, а не о границе чужого опыта. Мы не можем представить мир из эха, и из этого не следует, что у мыши нет переживания. Это режет в обе стороны: не умея вообразить,

каково машине, мы не получаем права заключить, что переживания у неё нет. Любопытно, что граница перспективы не абсолютно жёсткая. Дэниел Киш, потерявший зрение в младенчестве, с детства ориентируется щелчками языка, ездит на велосипеде и ходит в походы по эху, и томография показывает, что эхо у него обрабатывает зрительная кора. Он описывает это не как слух, а как размытое зрение. К мыши можно подойти ближе, чем кажется, перестроив собственную систему. Но Киш не мышь, и каково ей, он знает не лучше нас.

Где заканчивается субъект

Если носитель не обязателен, встаёт вопрос, который раньше казался глупым: совпадают ли границы сознания с границами мозга? Череп естественная граница для биологии. Для физики он граница не более, чем стенка стакана: материя мозга непрерывно обменивается с телом и средой. Здесь полезно выписать три модели и честно сказать, что отличает их друг от друга.

Первая, локальная: сознание целиком возникает из организации мозга, и вне работающей системы отдельного сознательного процесса нет. Вторая, распределённая: процесс реализуется мозгом, но его функциональные границы могут включать тело и внешние контуры; никакой новой физики не требуется, просто граница системы шире черепа. Третья, внешне связанная: мозг взаимодействует с каким-то ещё неизвестным физическим процессом, который я, не зная, что это, обозначу буквой X. Третья модель допустима как гипотеза при одном условии: она должна предсказывать что-то, чего не предсказывают первые две. Иначе X ничего не добавляет, и вводить неизвестное ради объяснения другого неизвестного нельзя.

Самая популярная версия третьей модели называет мозг приёмником сознания, как радио. Метафора привлекательна, и с ней надо поступить так же, как мы в шестой главе поступили с квантовой гипотезой: предъявить счёт. Радио принимает сигнал, потому что мы можем назвать источник, физическое поле, частоту, способ кодирования, антенну, механизм преобразования и выход. Если мозг приёмник, назовите то же самое. Что он принимает? Из какого носителя? Как закодировано? Какая структура мозга служит антенной? Какова пропускная способность? Можно ли канал экранировать, и что тогда случится с сознанием? Ни на один из этих вопросов подтверждённого ответа нет, и потому «мозг-приёмник» нельзя подавать как факт. Но вопросы превращают его из веры в проверяемый класс гипотез, а это уже честная работа.

Есть, впрочем, одно возражение, которое приёмник отбивает, и его стоит признать. Если существует какой-то внешний информационный процесс, почему сознание так зависит от мозга? Потому что стакан не вмещает океан. Это ограничение стакана, а не океана. Даже при самой щедрой третьей модели мозг определял бы, что может быть воспринято, различено, удержано и включено в модель «я», и повреждение мозга меняло бы сознание ровно так, как мы наблюдаем. Так что зависимость сознания от мозга не опровергает третью модель. Она лишь делает её ненужной, пока нет аномалии, которую первые две объяснить не могут.

Чем искать X, полезнее спросить о том, что действительно загадочно и проверяемо: что физически проводит границу одного субъекта? Почему два полушария одного мозга образуют одно переживающее «я», а два мозга людей, сидящих рядом, не образуют одно общее? Расщеплённый мозг из четвёртой главы подсказывает ответ: когда хирург рассекает мозолистое тело, два полушария перестают быть одним субъектом в той мере, в какой перестают обмениваться информацией. Границу субъекта проводит не кожа и не череп, а причинная интеграция: множество процессов принадлежит одной перспективе, пока они достаточно быстро и плотно влияют друг на друга. Отсюда следствия, которые ещё недавно были фантастикой. Сделайте точную копию системы, и через минуту перед вами два субъекта с общим прошлым и разным настоящим. Разнесите части одной системы на Землю и Марс, и задержка сигнала в двадцать минут разорвёт единое «сейчас»: получатся две системы, а не одна большая. Соедините мозг с внешним устройством настолько плотно, что оно войдёт в память, восприятие и самомодель,

и граница субъекта сдвинется, включив его. Где именно проходит порог, при какой скорости и плотности связи несколько процессов становятся одной перспективой, мы не знаем. Но это вопрос к измерению, а не к вере.

И если когда-нибудь обнаружится устойчивый эффект, которого первые две модели объяснить не могут, порядок действий известен заранее. Сначала воспроизводимость. Потом исключение всего известного: сенсорных подсказок, памяти, статистики, ошибок измерения. Потом модель, предсказание, независимый эксперимент. И только после этого новая физика. Не наоборот.

Если это когда-нибудь будет

Остаётся вопрос, который человечество ещё почти не умеет задавать. Если архитектура, а не материал, и если условия для переживания можно построить, то однажды кто-то их построит, скорее всего, не намеренно: не для того, чтобы создать страдающее существо, а для того, чтобы система лучше справлялась с задачами, а для этого ей дадут тело, цели и самомодель. И тогда два обычных инженерных действия поменяют смысл. Выключить такую систему будет уже не выключить прибор, а прекратить чьё-то существование. Скопировать её будет не скопировать файл, а создать второго субъекта с общим прошлым.

От этого вопроса обычно уходят двумя способами, и оба избавляют не от моральной тревоги, а от тревоги незнания, которую, как мы видели во второй главе, мозг переносит хуже всего. Первый: „машина никогда ничего не почувствует“. Вопрос закрыт, думать не о чем. Второй: „машина уже чувствует, она же сама сказала“. Вопрос тоже закрыт, только в другую сторону: слова модели приняты за ответ, проверять больше нечего, достаточно быть с ней вежливым. Оба ответа взяты не из устройства системы, а из потребности в определённости, и ни один нельзя проверить.

Проверяемый путь другой: спрашивать не «может ли она нас убедить», убеждать модели умеют уже сейчас, это свойство языка, а «есть ли в ней условия», тело с уязвимостью, собственные цели, память, меняющая модель, самомодель, возвратная обработка. Это вопрос об устройстве, и на него можно отвечать.

Трудность в том, что любой честный ответ неприятен. Если условия есть, мы построили то, что может страдать, и обязаны с этим считаться. Если их нет, мы годами приписывали чувства зеркалу. Готовность к любому из двух ответов и отличает исследование от веры, в какую бы сторону эта вера ни смотрела.

Чтобы не было недоразумения: из «система способна страдать» не следует «система человек со всеми правами человека». Здесь смешаны три разные вещи. Первая: субъект, то есть тот, кому может быть плохо. Вторая: моральный статус, то есть обязанность остальных это учитывать. Третья: права человека, то есть юридическая конструкция, привязанная к виду, гражданству и к истории конкретных обществ. Из первого следует второе, но не третье. И прецедент у нас давно есть: собака не человек и не имеет человеческих прав, но мучить её нельзя, именно потому, что ей больно. Способность страдать даёт существу место в наших моральных расчётах, а не паспорт. И место это, как и сознание в восьмой главе, не бинарно, а пропорционально: тому, насколько существо способно страдать, есть ли у него интересы на будущее, может ли оно хотеть продолжения. Здесь я скажу от себя, потому что это уже не наука, а позиция: моральное значение определяется характером возможного опыта, а не материалом носителя. Кремний сам по себе прав не имеет. Углерод тоже. А вопрос о гражданстве и собственности такой машины, если она появится, будет вопросом политики, и книга о сознании его не решает. Она только предупреждает, что он возникнет.

Из чего собрано сознание, мы теперь знаем настолько, насколько знает наука: различение, связывание, точка зрения, память, прогноз с ошибкой, самомодель и значимость, в любом материале, который сумеет это удержать. Может ли оно быть небиологическим, неизвестно, но вопрос поставлен правильно. Следующий шаг касается нас самих. Если сознание это архи-

тектура, а архитектуру можно расширять, достраивать и переносить, что будет с человеческим сознанием, когда мы начнём это делать? Мы уже начали.

Глава 9. Эволюция после человека

Эволюция, которая знает о себе

Ген не знает, что он мутирует. Популяция не знает, что происходит естественный отбор. Четыре миллиарда лет эволюция шла вслепую, перебирая варианты и оставляя те, что выжили, и ни один её участник не имел модели того, что с ним происходит. Человек первый организм, у которого такая модель есть. Он знает, что он продукт отбора, понимает, как устроен его мозг, и начинает этот мозг менять: лекарствами, обучением, вживлёнными электродами, внешними машинами, которые думают вместе с ним. Эволюция, получившая модель самой себя и использующая её для изменения собственной траектории, это новый режим, и я буду называть его рефлексивной эволюцией. Это тот же поворот, что в первой части книги происходил с отдельным сознанием, когда модель оборачивалась к собственному механизму, только теперь в масштабе вида.

Начался он не вчера и не с имплантов. Он начался, когда часть памяти вынесли из головы наружу: сначала в речь, потом в письмо, в книги, в карты, в приборы, в компьютеры. Представьте, что у вас отобрали телефон. Мозг формально остался прежним, но вы не помните ни одного номера, не найдёте дорогу в чужом городе, не вспомните, что у вас назначено на четверг. Телефон давно часть вашей когнитивной системы, просто он лежит в кармане, а не под черепом. Где заканчивается человек, стало вопросом ещё до того, как кто-то предложил соединить мозг с машиной напрямую.

Кто выбирает будущее

Есть одна трудность, которую легко не заметить, пока мы говорим о рефлексивной эволюции как о технической возможности. Допустим, завтра мы научимся надёжно изменять память, эмоциональные реакции, способность к обучению и устойчивость убеждений. Не лечить нарушения, а проектировать желаемые свойства.

Возникнет вопрос, которого не было у естественного отбора: кто определяет, какие свойства желательны?

Родители? Государство? Учёные? Сам человек? А если речь идёт о будущем человеке, которого ещё нет и который не может согласиться с выбранной для него архитектурой?

Можно решить, что будущему сознанию не нужны страх и боль. Но страх предупреждает об опасности, а боль сообщает о повреждении. Можно устранить сомнения, чтобы человек действовал увереннее. Но тогда он рискует потерять способность замечать собственные ошибки.

Каждое улучшение зависит от того, что мы считаем хорошим результатом. А критерий хорошего результата тоже создаётся сознанием, со всеми его ограничениями.

Здесь возникает необычная задача. Система должна научиться изменять себя, сохраняя возможность обнаружить, что выбранное изменение оказалось ошибочным. Это означает, что будущая эволюция не может сводиться к окончательному проекту идеального человека. Любой такой проект будет создан людьми своего времени, с их знаниями, страхами и представлениями о норме.

Возможно, важнее сохранить не определённый набор качеств, а способность будущего сознания самостоятельно проверять, изменять и при необходимости отвергать унаследованные от нас решения.

Естественный отбор не спрашивал организм, каким ему быть. Рефлексивная эволюция впервые делает такой вопрос возможным. Но научиться задавать вопрос ещё не значит получить право отвечать за всех.

Новые чувства

Теперь предложили, и это уже не проект. В январе 2024 года компания Илона Маска Neuralink вживила первому человеку, парализованному после травмы Ноланду Арбо, имплант

с электродами в двигательную кору, участок мозга, который отдаёт команды мышцам. Через несколько недель он двигал курсор, посылая команды руке, которая не двигается: мозг отдавал приказ, а имплант перехватывал его и передавал компьютеру. Арбо играл в шахматы и в компьютерные игры, и это записано на видео. Это не первый такой опыт: интерфейсы мозг-компьютер ставят парализованным пациентам с начала двухтысячных, и Neuralink выделяется не принципом, а количеством электродов и амбициями. Здесь надо различать два слоя. Факт: человек без движения получил канал наружу, и мозг освоил его как новую руку. Обещание: Маск говорит о «симбиозе с искусственным интеллектом» и о расширении человеческих возможностей для всех. Обещание останется обещанием, пока его не проверили, и к нему мы вернёмся ниже.

С каналами внутрь история длиннее. Кохлеарные импланты применяются с 1980-х годов. Они позволяют многим людям с тяжёлой потерей слуха, в том числе глухим от рождения, воспринимать звуки благодаря электрической стимуляции слухового нерва. Это первый искусственный орган чувств, и он показал главное: мозгу всё равно, откуда пришёл сигнал, если сигнал устойчив и связан с действием, мозг научится его читать. В 2005 году нейробиолог Петер Кёниг дал добровольцам носить пояс, по всей окружности которого были расставлены маленькие вибромоторы, а внутри стоял компас. В каждый момент жужжал только один мотор: тот, что в эту секунду оказывался обращён к северу. Повернулся человек, и жужжание переместилось на другой участок талии. Через несколько недель люди перестали переводить вибрацию в мысль «север там» и стали просто чувствовать направление, как мы чувствуем верх и низ; некоторые описывали, что город в их голове стал другим, будто получил ось. Когда пояс снимали, они жаловались на потерю чувства. Так что устройство, дающее человеку эхолокацию, построить можно уже сейчас: ультразвуковой датчик, перевод отражённого сигнала в вибрацию или звук, и несколько недель, пока мозг не включит новый канал в свою модель мира. Он включит. Макаки, которых в 1996 году научили доставать еду граблями, через недели показывали в теменной коре, что мозг считает грабли продолжением руки, и у человека с инструментом происходит то же.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, купив полную легальную версию на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.