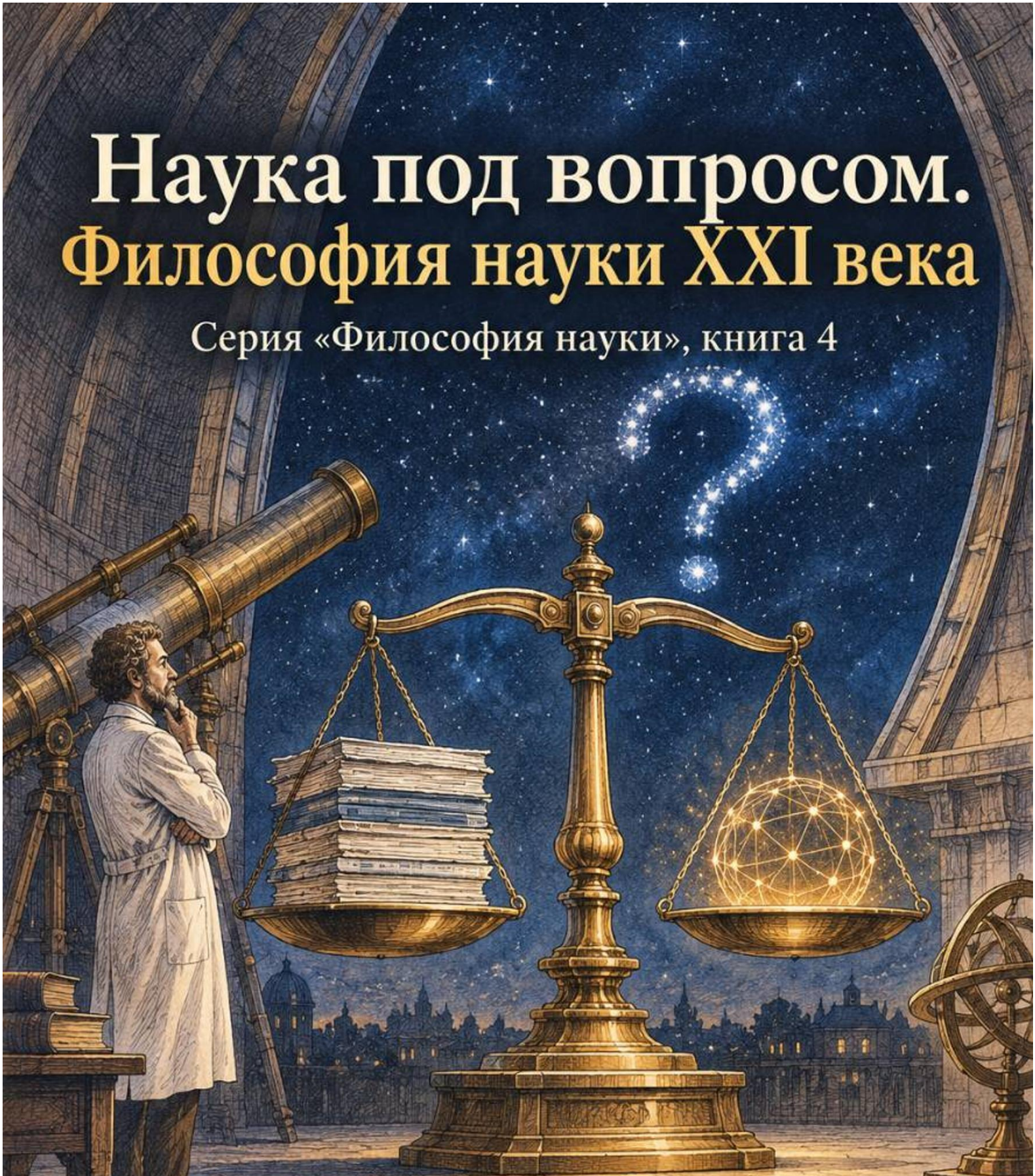
Наука под вопросом.
Философия науки XXI века

Серия «Философия науки», книга 4



Ярослав Золотарёв

Ярослав Золотарёв

Наука под вопросом. Философия науки XXI века

«Автор»
2026

Золотарёв Я.

Наука под вопросом. Философия науки XXI века /
Я. Золотарёв — «Автор», 2026

Из ста повторённых психологических опытов подтвердилась лишь треть. Машина получает Нобелевскую премию за задачу, которую люди не могли решить полвека, и никто не может объяснить, как она это сделала. Математики десятилетие спорят о доказательстве, которое не могут ни принять, ни опровергнуть. Значит ли это, что наука не лучше гадания? Заключительная книга серии «Философия науки» рассказывает о том, как наука XXI века проверяет сама себя: от байесовской революции и причинности Перла до квантовой механики, языковых моделей, научных войн и новой философии лженауки. Почему фраза «учёные доказали» ничего не доказывает? Чем инструменталист отличается от релятивиста, а настоящий скептик от торговца сомнением?В конце серии автор впервые говорит от своего имени и отвечает на вопрос, с которого начинался первый том: чем астроном отличается от астролога.

© Золотарёв Я., 2026
© Автор, 2026

Содержание

Введение. Сто повторённых опытов 5
Август 2015 года 5
Два неверных вывода 6
Откуда мы пришли 8
Что изменилось в XXI веке 9
Пять вечных вопросов в новой редакции 11
Маршрут тома 12
Часть I. Как взвешивать свидетельства 14
Глава 1. Байесовская революция 14
Священник, прилив и Юм 14
Что значит «данные поддерживают гипотезу» 15
Формула словами 17
Что такое вероятность 19
Что байесианство объясняет 22
Возражения: откуда берётся уверенность 25
Итог и вопрос читателю 28
Глава 2. Кризис воспроизводимости 30
Предвидение в ведущем журнале 30
Что означает «значимо» 31
Как получить любой результат 33
Масштаб и причины 37
Возражения и что осталось 39
Итог и вопрос читателю 44
Глава 3. Причина возвращается 46
«Что» и «почему» 46
Как причину изгнали из науки 47
Лестница причинности 49
Миры, которых нет 55
Конец ознакомительного фрагмента. 57

Ярослав Золотарёв
Наука под вопросом.
Философия науки XXI века

Введение. Сто повторённых опытов

Август 2015 года

28 августа 2015 года журнал Science напечатал статью, под которой стояло 270 фамилий. Такого количества авторов не наберёт ни один роман-эпопея, а для психологии это было примерно то же, что для физики коллайдер: большая машина, собранная из множества людей. Руководил работой американский психолог Брайан Носек, а задача выглядела скучнее некуда. Участники взяли три ведущих психологических журнала, отобрали сотню исследований, опубликованных в 2008 году, и повторили их. Никаких новых теорий, никаких открытий. Те же процедуры, по возможности те же материалы, выборки достаточные, чтобы эффект, если он есть, можно было заметить. Всё, что требовалось, — проверить, получится ли у других людей то, что уже напечатано.

Есть занятия, которые вознаграждаются очень плохо, и повторение чужих опытов стоит среди них на одном из первых мест. За него не дают премий, редакторы его не любят, а коллега, чей результат вы не подтвердили, не скажет вам спасибо. Поэтому годы работы большого коллектива выглядели почти как подвижничество, и ждали от них, по всей видимости, спокойного подтверждения: да, кое-что не воспроизводится, но в целом дисциплина стоит на твёрдой почве.

Вышло иначе. В исходных статьях статистически значимые результаты были в 97 случаях из 100. В повторениях — в 36. Когда самих участников спросили, подтвердился ли, по их мнению, исходный результат, положительно ответили примерно о 39 процентах работ. Средняя величина эффекта при повторении оказалась примерно вдвое меньше заявленной. Речь шла не о брошюрах из киоска при вокзале и не о журнале «Тайны потустороннего». Речь шла о лучших журналах науки, которая преподаётся в университетах всего мира и на выводы которой опираются учителя, врачи, рекламщики и законодатели.

Читатель прошлых томов серии вправе насторожиться. Он прошёл с нами долгий путь и видел, как от Пифагора до Лаудана люди пытались ответить на один вопрос: чем наука отличается от всего остального, что претендует на знание. И главным ответом всё время оставалась проверка. Астроном говорит, когда и где будет комета, и либо она появляется, либо нет. Гадалка говорит, что в вашей жизни скоро произойдут перемены, и опровергнуть её не удастся никогда. На этом различии держится вся серия. И вот самое простое из возможных применений этого различия — возьми опубликованное и проверь — выдаёт в одной из почтенных наук две трети неудач.

Именно с этой новости стоит начать последний том, потому что она собирает вместе почти все его вопросы. Что считать подтверждением? Что значит «статистически значимо» и почему это выражение оказалось таким податливым? Почему научное сообщество годами публиковало то, что не повторяется? Кто решает, какой опыт считается удачным? И если сбой обнаружен, что он нам говорит: что наука больна, или что она работает?

Два неверных вывода

Первый вывод немедленно сделали те, кто давно его ждал. Представим себе потомственного целителя, участника телевизионного шоу о людях со сверхспособностями, которому годами твердили о доказательной науке и контрольных группах. Ему попадается заголовок: «Большинство психологических исследований не подтвердилось». Теперь у него есть ответ на всё. Учёные сами признали, что их результаты не воспроизводятся. Значит, их опыты ничем не лучше его опыта, а его опыт насчитывает три поколения бабушек. Значит, наука — такая же вера, только с дипломами и грантами. Сходным образом рассуждает и производитель чудодейственной пищевой добавки, и лектор, разоблачающий «официальную медицину», и политик любой страны и любого строя, которому не нравится чей-нибудь неудобный отчёт: раз учёные ошибаются, можно не слушать и этих.

Вывод этот рассыпается от первого вопроса. Кто проверил сто психологических опытов? Психологи. На чьи деньги, какими методами, в каком журнале? На научные деньги, научными методами, в научном журнале. Сбой обнаружен той самой процедурой, которую целитель объявляет бесполезной. Покажите нам коллектив из 270 астрологов, которые взяли сто опубликованных гороскопов, проверили их по единому протоколу, напечатали результат в главном журнале своей профессии и назвали цифру: столько-то не сбылось. Покажите целителей, которые повторили лечение своих коллег и сообщили, что две трети заявленного не воспроизводятся. Такой работы нет и быть не может, причём не по злому умыслу: в этих занятиях просто отсутствует то, что можно было бы воспроизвести. Ошибки есть и у тех, и у других. Но только у одних они записаны в виде числа.

Второй неверный вывод делают защитники. Он гораздо приличнее и поэтому опаснее. Ничего особенного, говорят они, не произошло. Весной 2016 года в том же Science появился технический комментарий, который написали гарвардский психолог Дэниел Гилберт и его соавторы. Они утверждали, что статистика Носека содержит ошибки: повторения не всегда точно следовали исходным процедурам, случайных неудач при такой постановке следовало ожидать немало, и если всё это учесть, то данные не дают оснований говорить о низкой воспроизводимости, а совместимы и с обратным заключением. Спор о том, что именно считать успешным повторением, идёт до сих пор, и мы вернёмся к нему во второй главе. Гилберт поднял законные вопросы: цифру 36 нельзя просто прочесть как долю ложных открытий.

Однако вывод «всё в порядке» из этих поправок не следует. Сам факт, что о таких вещах нужно было спорить в 2016 году, говорит о многом. Если специалисты не могут договориться, подтвердился ли опыт, значит, понятие подтверждения в рабочей практике науки оказалось куда менее ясным, чем думали её защитники и чем обещают учебники. Если величина эффекта при повторении падает вдвое, то весь корпус литературы завышает то, о чём сообщает, и завышает систематически, а значит, ошибка сидит не в отдельных лабораториях, а в устройстве всего производства.

Есть и третий участник этого спора, о котором редко говорят: обычный читатель новостей. Почти каждую неделю ему сообщают, что учёные доказали пользу или вред кофе, шоколада, красного вина, раннего подъёма и долгого сна, а через год сообщают обратное. Он постепенно усваивает полезное, как ему кажется, правило: они там сами не знают. Правило удобно, потому что освобождает от необходимости разбираться, и оно частично верно, потому что исследования такого рода действительно часто не повторяются. Но вместе с плохими работами оно отбрасывает и прививку от оспы, и расчёт орбиты, и данные о том, что курение вызывает рак. Читатель, который не умеет отличить слабое свидетельство от сильного, в итоге верит либо всему, либо ничему, а выбор между этими двумя состояниями обычно определяется не доводами, а тем, откуда он сегодня узнал новость.

Между этими двумя выводами и лежит тот, ради которого написан этот том. Наука отличается от гадания не тем, что она не ошибается. Об этом мы могли догадаться ещё в первом томе, читая Юма, и в третьем, читая Поппера. Наука отличается тем, что её ошибки можно найти, посчитать и назвать, а метод находить их сам становится предметом исследования и улучшается. Но это утверждение нужно заработать, и прежними средствами его заработать уже нельзя. Простая формула «проверь на опыте» перестала быть простой. Оказалось, что проверка сама состоит из множества решений: каких людей отобрать, что считать результатом, какой порог принять, что напечатать и что положить в стол, кто платит. В каждом из этих мест проверку можно незаметно подменить чем-то другим. Весь том, по существу, об этих подменах и о том, как их замечать.

Откуда мы пришли

Тому, кто открывает серию на этом томе, полезно знать, какой путь остался позади. Читателю предыдущих книг хватит напоминания.

Первый том рассказывал о рождении метода, от пифагорейцев, которые искали числа в музыке и небе, до Юма и Канта. Юм показал, что из прошлых наблюдений логически не следует ничего о будущих. Кант попытался спасти науку, поместив её основания в устройство нашего собственного познания. К концу XVIII века было ясно главное: наука работает, а почему она работает, никто доказать не может.

Второй том был веком позитивизма, от Конта до Пуанкаре. Наука стала образцом для всякого знания и едва ли не новой религией. Зато и вопросы о ней стали точнее: Милль искал правила индукции, Мах настаивал, что теория есть экономное описание ощущений, Дюгем объяснил, что опыт проверяет не отдельную гипотезу, а всю систему допущений сразу, Пуанкаре заметил, что от смены теорий остаются отношения между явлениями, хотя сами образы рушатся. Этот том закончился накануне большой физической революции.

Третий том был самым толстым, и не случайно. XX век предложил все главные ответы и сам же почти все их разрушил. Венский кружок хотел провести границу науки при помощи логики и критерия проверки, и граница расползлась у него в руках. Карнап строил индуктивную логику и не достроил. Австрийско-британский философ Карл Поппер заменил подтверждение опровержением, но его критерий оказался слишком простым для живой науки. Американский историк и философ науки Томас Кун показал, что учёные работают внутри общих образцов и не бросают их при первой неудаче, и из его книги многие вычитали то, чего он, по его собственным заверениям, не хотел сказать: что смена теорий подобна обращению в новую веру. Мы отнеслись к этому скептически. Имре Лакатос попытался спасти рациональность смены теорий, а Пол Фейерабенд объявил, что допустимо всё, и за ним пошли сторонники сильной программы в социологии знания и постструктуралисты, для которых научное утверждение оказалось всего лишь продуктом интересов и текстов.

В конце века на столе остались два развороченных спора. Один касался реальности: говорят ли успешные теории что-то о том, что существует помимо наших наблюдений, или они только хорошо предсказывают? Другой касался границы: в 1983 году американский философ Ларри Лаудан, уже знакомый нам по третьему тому, объявил о кончине проблемы демаркации. Все предложенные критерии, писал он, провалились, а значит, сам вопрос о том, что считать наукой, следует оставить. Мы закончили тем, что с этим приговором не согласились, но пообещали вернуться к нему с новыми доводами. Этот том и есть исполнение обещания.

Из истории серии стоит удержать одно наблюдение. Каждый раз, когда философы объявляли науку объяснённой, она менялась так, что объяснение устаревало. И каждый раз, когда объявляли, что никакой особой научной рациональности нет, это говорилось в зале с исправным микрофоном, на конференции, куда докладчик прилетел самолётом, рассчитанным по совершенно определённым уравнениям.

От прежних томов этот отличается ещё в одном отношении. Там время успело поработать за нас: кто такой Кант, кто такой Мах и что осталось от Венского кружка, история более или менее рассудила. Здесь почти все герои либо живы, либо ушли совсем недавно, а их споры не окончены. Поэтому мы будем стараться отмечать, где заканчивается то, с чем согласны почти все специалисты, и где начинается открытая дискуссия, и не станем притворяться, будто будущий учебник у нас уже в руках.

Что изменилось в XXI веке

Философия науки нового века отличается от прежней не только списком имён. Изменилось само положение дел, и три сдвига определяют всё, что будет дальше.

Первый сдвиг: от логики науки к её практике и прежде всего к её числам. Венский кружок разбирал предложения, Поппер — логическую форму гипотез. Спор шёл так, будто учёный сидит над формулой и сравнивает её с протоколом наблюдения. Сегодняшний учёный сидит над таблицей данных и статистической программой, и между его гипотезой и заключением лежит цепочка расчётов, которую ни один логик начала прошлого века не стал бы рассматривать. Вопрос «что подтверждает гипотезу?» превратился в вопросы о том, что такое вероятность, как взвешивать свидетельства, чем отличается наблюдаемая связь двух величин от причины. На эти вопросы теперь отвечают и философы, и статистики, и специалисты по информатике, нередко не читая друг друга. Отсюда первая часть тома: байесовская революция, кризис воспроизводимости и возвращение причинности, которую статистика более полувека считала неприличным словом. Сюда же относится поворот к эксперименту. Канадский философ Ян Хакинг заметил в 1983 году, что философы рассуждают о теориях так, будто учёные только и делают, что созерцают, и спросил, что изменится, если посмотреть на людей, которые вмешиваются, чинят приборы и распыляют на шарик из ниобия электроны.

Насколько велик этот сдвиг, можно показать на одном числе. Десятилетиями результат в психологии, биологии и медицине считался достойным публикации, если вероятность получить его случайно не превышала пяти сотых. Порог этот возник как практическое удобство из таблиц начала XX века, а превратился в ворота, через которые проходила карьера. Ни один логик не занимался этими воротами, потому что они не выглядели философским вопросом. Они выглядели технической подробностью. Сегодня ясно, что эта подробность решала, что будет считаться знанием, и что вокруг неё сложилось целое производство способов пройти в ворота, не нарушая ни одного явного правила. Философский вопрос теперь нередко скрыт в настройках статистической программы.

Второй сдвиг: от физики к разнообразию наук. Философия XX века молчаливо считала физику-теоретика образцом всякого учёного, а всё прочее — незрелой физикой. Но биолог не ищет всеобщих законов там, где всё устроено исторической случайностью. Геолог и палеонтолог не могут поставить опыт над тем, что случилось шестьдесят шесть миллионов лет назад. Экономист и психолог имеют дело с предметом, который читает их собственные выводы и меняется от них. И даже сама физика, как показывает американский философ Нэнси Картрайт, выглядит куда менее единообразно, когда смотришь не на уравнения в учебнике, а на то, как их приспосабливают к конкретному прибору. Вопрос о том, есть ли один научный метод или целое семейство родственных практик, из отвлечённого стал рабочим. Он же прямо касается демаркации: если единого метода нет, что общего у всего, что мы называем наукой?

Третий сдвиг: машины и организованное недоверие. Эти явления, на первый взгляд далёкие, поставили один вопрос: кому и почему мы верим, когда сами проверить не можем. В 2024 году половину Нобелевской премии по химии вручили за предсказание строения белков, которое выполняет обученная программа. Результат проверяется лабораторией, предсказания работают, а вот объяснить, почему программа даёт именно такой ответ, не может полностью никто, включая её создателей. Что здесь считать пониманием? С другой стороны, языковые модели научились писать тексты, которые нередко трудно отличить от рассуждений специалиста, и вернули к жизни старый спор Соссюра, Блумфилда и Куайна о значении слова.

А в то же самое время повсюду окрепло умелое производство сомнений. В середине XX века табачные компании обнаружили, что для защиты товара не нужно ничего доказывать: достаточно поддерживать ощущение, что учёные ещё спорят. Этот приём оказался переноси-

мым на любую тему, и им пользуются в любой стране и при любом строе, когда вывод науки мешает чьим-то деньгам или чьей-то власти. Симметричным образом и слова «наука доказала» стали инструментом: ими можно закрыть обсуждение там, где никакой проверки не было, а было только мнение влиятельных людей. Ссылка на «научное сообщество» звучит убедительно, но сообщество состоит из людей, которым кто-то платит. Если проверку объявляют одной из традиций, её место занимает сила. Если вместо проверки ссылаются на авторитет сообщества, сила приходит с другого хода, но приходит та же самая. Четвёртая часть тома целиком посвящена этому раскладу.

Пять вечных вопросов в новой редакции

Во всех томах серии мы держали перед глазами пять вопросов, которые задаёт себе всякая философия науки: о методе, об объяснении, о реальности, о развитии и о границе. Посмотрим, как они звучат сегодня. Звучат они так, что ни Бэкон, ни Конт, ни даже Карнап сразу не узнали бы своего вопроса в новой одежде.

Метод. Прежде спрашивали, ведёт ли к знанию индукция, или лучше выдвигать смелые гипотезы и опровергать их. Теперь спрашивают иначе: как взвешивать свидетельства, если один и тот же результат можно получить честным опытом и подгонкой, а со стороны они выглядят одинаково? Как должна меняться наша уверенность от нового факта? Сколько в этом расчёте от данных, а сколько от того, во что мы верили до них?

Объяснение. Гемпель требовал вывести явление из закона. Теперь появились системы, которые предсказывают явление, ни из какого закона его не выводя. Нужна ли причина, если модель и так предсказывает? Что мы теряем, когда получаем правильный ответ без понимания, и можно ли вообще отличить понимание от очень хорошего предсказания?

Реальность. Что остаётся от мира после квантовой механики и нескольких смен теорий? Эксперименты с запутанными фотонами, за которые присудили Нобелевскую премию 2022 года, перевели старый спор Эйнштейна и Бора из разговора за чаем в лабораторный протокол. Но они вовсе не закрыли вопрос о том, что считать реальным, а скорее обострили его. Есть ли в наших теориях что-то, кроме структуры уравнений? И есть ли смысл спрашивать о реальности, которой никак нельзя проверить, если всё, что у нас есть, — это связи между наблюдениями и работающие модели? Спор о реализме идёт и сегодня, и в этом томе обе его стороны получат слово.

Развитие. Кун спрашивал, как сменяются образцы научной работы. Сегодня спрашивают ещё и о том, кто платит за эту работу и как устроено её разделение. Кому выгодны те или иные темы? Могут ли научные журналы, гранты и премии сами производить систематическую ошибку? Что изменится в науке, когда значительную часть её будут делать машины?

Граница. Лаудан утверждал, что её нет, потому что нет единого критерия. Новая философия лженауки отвечает: единого критерия действительно нет, но из этого ничего подобного не следует. У слова «болезнь» тоже нет одного определения, что не мешает отличать чуму от здоровья. Можно ли провести границу без единого признака? Кто её охраняет, и что происходит, когда охранять её берётся власть?

Эти пять вопросов не идут каждый по своей главе. Они переплетаются, и, надо признаться, ни на один из них философия XXI века не дала окончательного ответа. Но некоторые ответы теперь исключены, и знать, какие именно, уже немало. Читатель, который поймёт, почему байесианец не верит потомственному целителю, почему опыт Белла не доказывает, что сознание творит мир, и почему отсутствие единого критерия не уравнивает астрономию с астрологией, будет защищён лучше, чем читатель, выучивший десяток определений науки.

Маршрут тома

Том состоит из четырёх частей.

Первая часть называется «Как взвешивать свидетельства». Мы начнём с Байеса, преподобного математика XVIII века, чьё имя стало знаменем целого направления, и с вопроса о том, что значит быть уверенным на столько-то процентов. Затем вернёмся к сотне повторённых опытов, разберём, как устроена подгонка результатов и почему учёные годами обманывали самих себя без всякого намерения обманывать. И закончим причинностью, понятием, изгнанным из статистики и возвращённым в неё американским специалистом по информатике Иудеей Перлом.

Вторая часть, «Эксперимент, закон, реальность», самая философская. В ней Хакинг покажет, что можно быть уверенным в электронах, не будучи уверенным в теории электрона, Картрайт объяснит, в каком смысле законы физики лгут, а сторонники структурного реализма предложат сохранить от теорий только уравнения. Потом мы дойдём до квантовой механики и выясним, что в действительности следует из опытов Белла и Аспе, а что к ним приписали любители чудес. Заключительная глава части даст слово тем, кто вообще не считает нужным говорить о мире, скрытом за наблюдениями, и покажет, что у этой старой линии Маха и Пуанкаре есть новые доводы, в том числе из науки о мозге.

Третья часть, «За пределами физики», посвящена наукам, которые философия долго не замечала. Как работает наука о прошлом, где нельзя поставить опыт? Что такое математика, которая ничего не наблюдает и всё предсказывает? И что происходит с научным знанием, когда в нём участвуют машины?

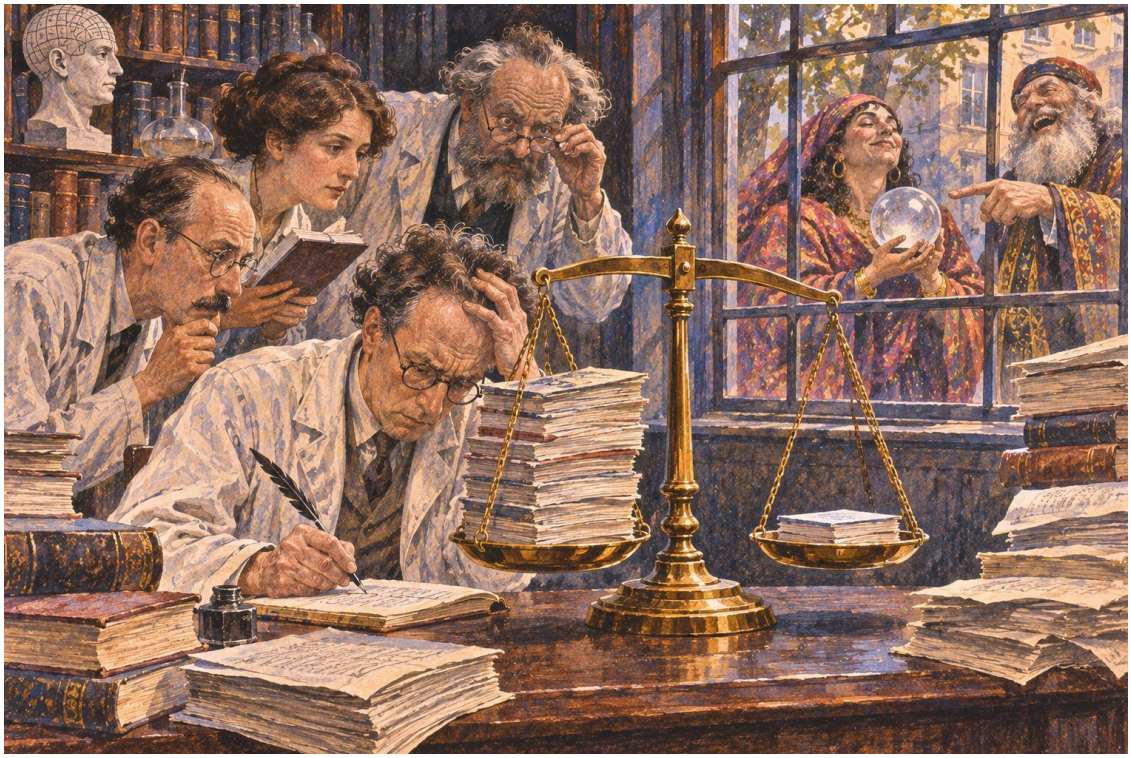
Четвёртая часть, «Наука и общество», возвращает нас к исходной ситуации этого введения. Она о том, как работает сообщество учёных и насколько ему можно доверять, о научных войнах конца прошлого века, о фальшивой статье, которую физик Алан Сокал отправил в журнал гуманитариев, о табачной индустрии и производстве сомнений. Заканчивается она новой философией лженауки: понятием-кластером, «пари Локи» и судебным делом в маленьком пенсильванском городке.

У этого тома есть ещё одна задача, не вполне философская. Большинство авторов, о которых пойдёт речь, плохо известны русскому читателю. Многие книги не переведены вовсе, другие переведены поздно и малым тиражом, а слова вроде «постпозитивизм» у нас часто значат то, чего никто из самих постпозитивистов в виду не имел. Кто читал Хакинга, знает, что он совсем не похож на своё расхожее изложение. Картрайт нередко цитируют как разоблачительницу физики, хотя ничего подобного она не хотела. Байесианство известно у нас преимущественно как формула из учебника или как модное слово, а Перла только начинают читать. Поэтому мы будем излагать этих людей подробно и по возможности их собственными доводами, даже там, где не соглашаемся с ними, а о реальной критике будем говорить как о критике, а не как о приговоре.

В конце тома нас ждёт заключение всей серии. В первом томе мы задали простой вопрос: чем астроном отличается от астролога? После четырёх томов мы ответим на него ещё раз, зная всё, что можно возразить на любой простой ответ. В том же заключении автор один раз скажет от собственного имени то, что до сих пор только подразумевалось: откуда в этой книге столько насмешки над теми, кто путает модель с миром, и почему эта насмешка не противоречит уважению к науке.

Читателю этого тома понадобится терпение. Здесь будут вероятности, хотя и без сложных формул, будет немного квантовой механики, будет статистика. Всё будет объяснено с самого начала, и ни одна глава не потребует ничего сверх школьной математики и готовности

подумать. Взамен мы постараемся дать то, чего не даёт ни один гороскоп: способ проверить, насколько мы правы, включая те случаи, когда очень хочется быть правым.



Часть I. Как взвешивать свидетельства

Глава 1. Байесовская революция

Священник, прилив и Юм

В начале 1767 года лондонский священник-диссентер Ричард Прайс выпустил книгу под названием «Четыре рассуждения». Последнее из них было направлено против самого опасного сочинения своего времени: эссе Дэвида Юма «О чудесах», напечатанного девятнадцатью годами раньше. В пылу спора Прайс назвал довод Юма жалкой софистикой. Потом, остыв, написал Юму письмо: он сожалеет о резком выражении и уберёт его из следующего издания. 18 марта того же года Юм ответил. Он поблагодарил и признал, что Прайс представил спор в свете новом, правдоподобном, остроумном, «и, возможно, основательном». По воспоминаниям современников, больше всего философа, которого половина Англии считала безбожником, удивила такая щепетильность в лице духовного лица. Так выглядела в XVIII веке полемика о том, воскресал ли Христос: два человека, расходящиеся в самом главном, обмениваются извинениями. Читатель, знакомый с сегодняшними спорами о науке и вере в интернете, может отложить книгу и немного помолчать.

Но нас интересует не только тон. В сноске и в рассуждениях Прайса содержалось нечто, чего в споре о чудесах прежде не было: расчёт. Прайс ссылался на работу своего покойного друга, который, как он писал, нашёл способ определять точную вероятность выводов, основанных на индукции. Этим другом был Томас Байес.

О Байесе известно немного, и это почти само по себе анекдот, если учесть, как часто теперь повторяют его имя. Он родился около 1701 года, был пресвитерианским священником, то есть принадлежал к тем протестантам, которых не допускали в Оксфорд и Кембридж, учился в Шотландии, служил в курортном городке Танбридж-Уэллс, был избран в Королевское общество и при жизни почти ничего не напечатал. Единственный его портрет, который кочует сегодня по учебникам и презентациям, изображает, по всей вероятности, вообще другого человека. Байес умер 7 апреля 1761 года. Прайс разыскал среди его бумаг рукопись, потратил почти два года, чтобы её дополнить и исправить, написал вступление и приложение с примерами, и 23 декабря 1763 года работу зачитали в Королевском обществе. Напечатана она была в следующем году в томе за 1763 год под нейтральным заголовком «Опыт решения одной задачи в учении о случайностях». Зато на отдельном оттиске, как обнаружил уже в нашем веке американский историк статистики Стивен Стиглер, стояло другое название: «Метод вычисления точной вероятности всех заключений, основанных на индукции». Если учесть, кто в ту пору громче всех утверждал, что заключения, основанные на индукции, ничем обосновать нельзя, намёк очевиден.

Юм нам уже хорошо знаком по первому тому. Там мы видели его главный вывод: из того, что солнце всходило до сих пор, логически не следует, что оно взойдёт завтра. В эссе «О чудесах» он применил свой скепсис иначе и практичнее. Чудо есть нарушение закона природы. Закон природы подтверждён единообразным опытом всех людей во все времена. Свидетельство о чуде опирается на слова немногих людей, которые могли ошибиться, поддаться страсти или соврать. Поэтому благоразумный человек всегда взвешивает: что вероятнее, что мёртвый воскрес или что рассказчик ошибается? Никакое свидетельство, заключал Юм, не может установить чудо, кроме такого, ложность которого была бы ещё большим чудом. Этот

довод с тех пор служит всем, кто разоблачает ясновидящих, и, надо сказать, служит неплохо. Прайс, однако, увидел в нём две слабости.

Первая слабость касалась самого понятия единообразного опыта. Пусть, рассуждал Прайс, прилив наблюдали миллион раз подряд, и он приходил всякий раз. Следует ли отсюда, что сбой невозможен? Юмовское рассуждение звучит так, будто следует. Но по расчёту, основанному на результате Байеса, из миллиона удач вытекает только то, что вероятность сбоя очень мала: примерно от одного случая на шестьсот тысяч до одного на три миллиона. Это очень мало. Но это не ноль, и разница между «очень мало» и «ноль» имеет принципиальное значение, мы увидим это в конце главы. Прошлый опыт, сколь угодно длинный, даёт степень уверенности, а не запрет.

Вторая слабость оказалась серьёзнее. Юм рассуждал так, будто на весы кладут один закон природы и одно свидетельство. А что будет, спрашивал Прайс, если свидетелей несколько и они независимы друг от друга? Показания ненадёжных людей перемножаются. Если каждый из свидетелей ошибается редко и они не могли сговориться, то вероятность того, что все они ошиблись одинаково, падает очень быстро, а значит, даже очень невероятное событие может стать вероятным. Американский философ Джон Эрман через двести с лишним лет посвятил этому целую книгу с заголовком, не оставляющим сомнений в его симпатиях: «Позорный провал Юма». Он доказывал, что Прайс в этом пункте прав, а прославленный довод Юма как логическое рассуждение не выдерживает проверки.

Предупредим сразу: из этого не следует, что чудеса случались. Прайс был верующим человеком и надеялся защитить Писание, но расчёт, как обычно, служит не тому, кто его придумал. Всё зависит от слова «независимо». Как только мы спросим, действительно ли свидетели независимы, не пересказывают ли они друг друга, не хотят ли все они одного и того же, не было ли среди них автора единственного первоисточника, весь чудесный результат тает. Это знает любой следователь. Многочисленные отзывы об одной и той же чудодейственной добавке на сайте её производителя выглядят как сто независимых свидетелей и стоят примерно столько же, сколько один продавец, переодевшийся сто раз. К этому слову мы ещё вернёмся.

Спор о мотивах самого Байеса не закончен. Стиглер считает вероятным, что Байес начал работать над задачей именно после эссе Юма: даты совпадают. Но он сам признаёт это предположение самым спекулятивным звеном своей истории, потому что в статье Юм не упомянут, а записи Байеса почти не сохранились. Другие историки полагают, что Байес вышел на задачу, читая Абрахама де Муавра, и никакой богословской цели не преследовал. Прайс — другое дело. У него цель была вполне определённой. Поэтому мы и будем говорить здесь о работе Байеса и о споре Прайса с Юмом как о разных вещах, хотя популярные книги, включая книгу Перла, о которой речь впереди, часто сливают их воедино.

Так или иначе, из богословского спора XVIII века вышло самое влиятельное в современной философии науки учение о том, как следует менять свои убеждения. Сегодня его применяют в фильтрах почтового спама, в поиске пропавших судов и самолётов, в судебной экспертизе и в машинном обучении. А философы подтверждения, которые к концу XX века устали от неудач Карнапа и Гемпеля, всё чаще говорят: это единственная теория свидетельства, которая у нас есть. Тот же Эрман назвал другую свою книгу в форме пари: «Байес или крах?». В этой главе мы разберём, что это за теория, почему её так полюбили и где она даёт сбой.

Что значит «данные поддерживают гипотезу»

Чтобы понять, почему байесианство вызвало такую надежду, надо вспомнить, в каком состоянии оставил этот вопрос XX век. На протяжении всего третьего тома мы наблюдали череду неудач.

Задача выглядела скромно. Логика умеет сказать, когда заключение с необходимостью следует из посылок. Но учёный почти никогда не находится в таком положении. У него есть гипотеза и несколько наблюдений, которые логически её не доказывают. И всё же он считает, и мы все считаем, что после удачного опыта гипотеза стоит прочнее, чем до него. Значит, нужна теория этого «прочнее». Венский кружок и его наследники пытались её построить как логику, только особую, ослабленную, в которой посылки не гарантируют заключения, а только поддерживают его в какой-то степени.

Из этого ничего не вышло. Гемпель предложил самый естественный принцип: всякий чёрный ворон подтверждает утверждение, что все вороны чёрные. И сразу получил белый ботинок, подтверждающий его же. Простой гипотетико-дедуктивный рецепт, согласно которому гипотеза подтверждается, когда сбываются её следствия, рухнул ещё быстрее: если гипотеза Н предсказывает наблюдение Е, то Е предсказывает и составная гипотеза «Н и на Луне живут гномы», так что гномы подтверждаются вместе с законом. Гудмен с его зелубыми изумрудами показал, что подтверждение вообще нельзя определить только по логической форме рассуждения: два вывода одного вида могут быть один хорошим, другой нелепым. Наконец, Карнап посвятил десятилетия индуктивной логике, основанной на вероятностях. В его системе вероятность всякого всеобщего закона при бесконечном мире оказывалась равной нулю, сколько ни наблюдай, и все дополнительные допущения, вводимые для устранения этого и других неудобств, только множили вопросы. Американский философ Питер Годфри-Смит, чей учебник служит нам образцом, называет всю эту историю повестью о горестях.

Поппер извлёк из неудач радикальный вывод: никакого подтверждения нет и не нужно. Наука движется догадками и опровержениями, а теория, выдержавшая проверки, всего лишь ещё не опровергнута. Этот ответ обладает героической простотой и одним неудобством, которое мы уже обсуждали в третьем томе. Когда инженер строит мост, он берёт теорию, многократно прошедшую проверку, а не только что придуманную и ещё ни разу не опровергнутую. Он прав. Но если прошлые удачи ничего не говорят о будущем, то он прав неизвестно почему. Понятие «подкреплённости», которое ввёл Поппер, описывает только прошлое, как зачётная книжка студента, и не содержит рекомендации на будущее. Либо подкреплённость всё-таки даёт основание для большего доверия, и тогда это подтверждение под другим именем, либо не даёт, и тогда инженер ничем не лучше того, кто строит мост по вчерашней догадке соседа.

А что же сами учёные? Они не ждали окончания этого спора и пользовались статистикой, и здесь надо сказать о вещи, которая для понимания всей главы важнее любой формулы. Мы уже видели в девятой главе третьего тома, как английский статистик Рональд Фишер и его оппоненты Ежи Нейман и Эгон Пирсон построили методы проверки гипотез. Эти методы устроены следующим образом. Предположим, что гипотеза верна, например, что лекарство не действует вообще. Тогда мы можем посчитать, насколько вероятно получить такие или ещё более резкие данные. Если вероятность мала, то мы говорим, что результат значим.

Обратите внимание на направление вопроса. Метод отвечает: какова вероятность данных при условии гипотезы. А учёному и его читателю нужен обратный ответ: какова вероятность гипотезы при условии данных. Это разные вещи, причём разные очень сильно. Вероятность того, что случайно выбранный папа римский окажется мужчиной, равна единице. Вероятность того, что случайно выбранный мужчина окажется папой римским, составляет примерно один на четыре миллиарда. Перепутать эти два числа в таком примере никто не может, но то же самое рассуждение, изложенное языком статистики, путают постоянно, в том числе и сами учёные, и мы увидим во второй главе, чего это стоит.

Именно этот вопрос о направлении и был темой статьи Байеса. Прямую задачу математики XVIII века решать умели. Если известно устройство игральной кости, можно посчитать, как часто выпадет шестёрка. Обратная задача гораздо труднее: у нас есть только результаты бросков, и мы хотим узнать, какова кость. Её называли задачей об обратной вероятности. В

ней-то и заключена вся проблема индукции в её прикладном виде, потому что наука всегда находится в положении человека, видящего результаты бросков. Причины нам не показывают. Показывают только следствия.

Условия, которым должна удовлетворять хорошая теория свидетельства, теперь можно перечислить. Она должна признавать степени: одни гипотезы лучше подтверждены, чем другие. Она должна менять эти степени по определённому правилу, не оставляя всё на произвол настроения. Она должна работать в самых разных случаях, от ворон до кварков, без отдельной логики для каждого. Она должна объяснять то, что учёные и так делают правильно: почему неожиданное предсказание весит больше ожидаемого, почему подгонка задним числом вызывает подозрение, почему повторение одного и того же опыта через некоторое время перестаёт что-либо добавлять. И наконец, она должна объяснять, на каком основании мы говорим об индукции вообще, после Юма.

Байесианство утверждает, что выполняет все эти условия, причём при помощи одной-единственной формулы, которую можно вывести за две строчки. Для философии, уставшей от неудач, это звучит почти как обещание чудесного исцеления, и потому к обещанию стоит отнестись так, как учит сама эта теория: посмотреть, откуда оно исходит, и проверить.

Формула словами

Теорема Байеса — одна из самых простых теорем математики. Математику для неё знать не нужно, достаточно умения делить. Её простота обманчива только в одном: сама теорема ни у кого не вызывает сомнений, а споры начинаются сразу после неё.

Запишем её словами. Обозначим гипотезу буквой H, а новое свидетельство — буквой E. Выражение P(H) означает вероятность гипотезы до того, как мы узнали о свидетельстве. Её называют априорной. Выражение P(E при H) означает вероятность увидеть такое свидетельство, если гипотеза верна. Её называют правдоподобием. Выражение P(H при E) есть то, что мы хотим узнать: вероятность гипотезы после того, как свидетельство получено. Её называют апостериорной. Теперь сама теорема:

P(H при E) = P(H), умноженная на P(E при H) и делённая на P(E).

Знаменатель P(E) — это общая вероятность увидеть такое свидетельство с учётом всех возможностей: и того, что гипотеза верна, и того, что она ложна. То есть он складывается из двух частей: вероятность E при истинной гипотезе, умноженная на вероятность её истинности, плюс вероятность E при ложной гипотезе, умноженная на вероятность её ложности.

Вся теория сводится к правилу пользования этой формулой. Прежде чем увидеть данные, у вас есть некоторая степень уверенности в гипотезе. Данные пришли. Новая степень уверенности равна той, которую подсказывает формула. Сегодняшняя апостериорная вероятность становится завтрашней априорной, и так до тех пор, пока приходят новые данные.

Чтобы увидеть, как это работает и почему вообще нужно считать, возьмём пример, разобранный в книге американского специалиста по информатике Иудеи Перла, с которым мы подробно познакомимся в третьей главе. Женщине сорока лет, без всяких симптомов, сделали маммографию, и результат оказался положительным. Что это значит? Числа такие: риск рака груди у женщины этого возраста в ближайший год составляет примерно один из семисот; при наличии опухоли обследование обнаруживает её примерно в 73 случаях из 100; у здоровых женщин оно ложно показывает тревогу в 12 случаях из 100.

Что подсказывает интуиция? Большинство людей, включая, как показывают опросы, немало врачей, отвечают, что вероятность болезни теперь велика, и называют что-нибудь вроде 70 процентов. Посчитаем теперь не в процентах, а в людях, как советуют психологи. Возьмём три тысячи таких женщин. Среди них больны примерно четыре. Обследование обнаружит заболевание примерно у трёх из них. Здоровых почти три тысячи, и 12 процентов из них, около

360 человек, тоже получат положительный результат. Итого положительный результат получили 363 женщины, а больны из них три. Вероятность болезни при положительном результате составляет меньше одного процента.

Повторим, что здесь произошло. Правдоподобие было велико: 73 процента. Свидетельство подняло вероятность болезни примерно в шесть раз, что совсем немало: с одного случая из семисот до одного примерно из ста шестнадцати. Но априорная вероятность была так мала, что и после шестикратного роста болезнь остаётся маловероятной. Обследование от этого не становится бесполезным: положительный результат всерьёз повышает вероятность и требует дополнительной проверки, а у женщины с наследственным риском один из двадцати тот же самый положительный результат поднимает вероятность болезни почти до одной трети. Одно и то же свидетельство при разных исходных условиях значит разное. Какие из этого выводы для конкретных обследований, решают врачи, а не философы, и не будем делать вид, будто глава о Байесе способна заменить консультацию. Для нас важно другое. Почти никто не может посчитать это в голове правильно. Мы плохо чувствуем обратную вероятность и всё время подставляем вместо неё прямую.

Из этого примера и формулы следуют три урока, и все три прямо касаются демаркации.

Первый урок: важна исходная частота. Ошибка, о которой мы только что говорили, встречается и в суде. В 1999 году в Великобритании была осуждена за убийство двух своих младенцев Салли Кларк. Важную роль сыграла цифра, названная педиатром-экспертом: вероятность двух случаев внезапной младенческой смерти в одной благополучной семье составляет один на 73 миллиона. Цифра содержала сразу две ошибки. Во-первых, эксперт получил её, просто умножив вероятность одной смерти на саму себя, то есть считал два случая независимыми, хотя у двух детей одной семьи может быть общая наследственная или бытовая причина. Вот то самое слово, о котором мы говорили у Прайса. Во-вторых, из малой вероятности невинного объяснения присяжные были склонны заключить, что столь же мала вероятность невиновности. Но ведь двойное убийство собственных младенцев тоже чрезвычайно редко. Сравнивать нужно не одну редкость с единицей, а две редкости друг с другом. Королевское статистическое общество выступило с публичным заявлением об ошибке, а через несколько лет приговор был отменён, хотя и по совокупности причин, не только статистических.

Второй урок: неожиданное весит больше. В формуле P(E) стоит в знаменателе. Чем менее вероятным казалось свидетельство до опыта, тем сильнее оно поднимает гипотезу, которая его предсказала. Если теория предсказывает то, чего никто не ждал, и это сбывается, её акции взлетают. Если она предсказывает то, что случается всегда, ничего не меняется. Перл выразил это формулой, с которой Байес и Прайс, пожалуй, согласились бы: чем чудеснее чудо, тем достовернее гипотеза, которая объясняет его появление. Но разумеется, чудо надо сначала установить, что возвращает нас к Юму.

Третий урок: важно, как часто такое свидетельство появляется, если гипотеза ложна. Перепишем формулу в другом виде, удобном для запоминания. Отношение шансов гипотезы после опыта равно отношению шансов до опыта, умноженному на так называемый фактор Байеса: на то, во сколько раз вероятнее свидетельство при истинной гипотезе, чем при ложной. Именно этот множитель и есть вес свидетельства. Если он равен единице, то есть свидетельство одинаково вероятно при обоих вариантах, то оно не весит ничего, каким бы впечатляющим ни выглядело.

Теперь посмотрим на гороскоп. «На этой неделе вас ждёт важный разговор, будьте внимательны к деньгам». Какова вероятность, что это сбудется, если астрология верна? Высокая. А если астрология ложна? Тоже высокая, потому что такое случается почти со всеми почти всегда. Множитель равен единице. Сбывшийся гороскоп не добавляет астрологии ровно ничего, и его сбывание можно наблюдать хоть каждую неделю до конца жизни. Вот точный смысл того, что Поппер называл отсутствием риска: теория, которая ничем не рискует, не может и выиг-

рать. Байесианство даёт этому рассуждению количественное выражение и попутно объясняет, почему «подтверждения», на которые ссылаются астрологи, не имеют никакой силы. Дело не в том, что подтверждений нет. Дело в том, что их вес равен нулю.

То же относится к телевизионному ясновидящему, который на глазах у миллионов находит спрятанный в багажнике предмет. Вопрос Байеса звучит не так: «могло ли это быть при наличии дара?» Вопрос звучит так: «насколько вероятно было увидеть именно этот эпизод на экране, если никакого дара нет?» Если снимали много дублей, если в эфир идут удачи, если участники оборудованы подсказками или хотя бы умеют наблюдать за людьми, то вероятность такой передачи при отсутствии дара близка к единице. И вся впечатляющая сцена, снятая с драматической музыкой, в точном смысле слова ничего не весит. Зритель, который чувствует обратное, просто не видит знаменателя.

Вернёмся теперь к Прайсу и посчитаем его свидетелей. Пусть мы оцениваем событие, вероятность которого составляет один на миллион. Пусть каждый свидетель говорит правду в 99 случаях из 100. Один свидетель умножает шансы примерно в 99 раз: чудо всё ещё очень маловероятно, примерно один шанс на десять тысяч. Два свидетеля, если они независимы, умножают уже почти в десять тысяч раз. Три дают примерно миллион, и шансы уравниваются. Четвёртый делает событие вполне вероятным. Юм не учёл именно этого. Но обратите внимание, что весь этот расчёт держится на независимости. Если второй свидетель слышал о случившемся от первого, он не добавляет ничего, множитель для него равен единице. Если сто тысяч человек верят в явление, потому что об этом рассказал один проповедник или одна телепередача, то весит это как один рассказ. Юм прав там, где говорит о человеческой склонности к чудесному, и считать надо не количество верующих, а количество независимых проверок.

Согласно этой логике, известное правило, что чрезвычайные утверждения требуют чрезвычайных доказательств, которое популяризовал американский астроном Карл Саган, не пристрастие скептиков и не способ отмахнуться от неудобного. Это следствие арифметики. Чем ниже априорная вероятность, тем больший множитель нужен, чтобы её поднять. Именно поэтому, когда кто-то заявляет о вечном двигателе, разумно требовать большего, чем когда кто-то заявляет о новой марке двигателя автомобильного.

На этом месте возникает очевидный вопрос. Априорная вероятность рака груди известна по медицинской статистике. Но откуда мы берём априорную вероятность общей теории относительности, или гипотезы о том, что астрология верна? Что вообще означает вероятность утверждения, которое либо истинно, либо ложно, и никакого множества повторений для него нет? Без ответа на этот вопрос формула останется формулой. Ответ занял почти двести лет и стоил математикам немалых нервов.

Что такое вероятность

В начале XX века на этот вопрос было два ответа, и оба для наших целей не годились. Первый мы встречали во втором томе у Джона Венна, а в третьем у Райхенбаха: вероятность есть предельная частота. Когда мы говорим, что у монеты вероятность орла равна половине, мы имеем в виду, что при очень большом числе бросков орёл выпадает примерно в половине случаев. Частотное понимание удобно для страховых таблиц, игральных костей и физических коллективов. Но для единичных утверждений оно ничего не даёт. У теории относительности нет частоты. Она либо верна, либо нет, и нельзя бросить вселенную тысячу раз и посмотреть, в какой доле случаев теория оказалась правильной. Поэтому для частотника выражение «вероятность гипотезы» бессмысленно, и именно отсюда растут методы Фишера, Неймана и Пирсона, старательно избегающие этого выражения. Логично, но учёному от этого не легче.

Второй ответ давала логическая теория, которую развивал английский экономист Джон Мейнард Кейнс в «Трактате о вероятности» (1921), а потом на свой лад Карнап. Вероятность,

по Кейнсу, есть объективное логическое отношение между посылками и заключением, нечто вроде частичного следования. Такое отношение одно на всех, и разумный ум способен его усмотреть.

В 1926 году эту теорию подверг уничтожающей критике молодой кембриджский математик и философ Фрэнк Рамсей. Читатель третьего тома помнит его по предложению Рамсея, о котором мы ещё будем говорить в шестой главе этого тома. Рамсей был человеком, о котором современники говорили с некоторым смущением, потому что его способности не укладывались в обычные мерки. За несколько лет он внёс основополагающий вклад в логику, основания математики, теорию налогообложения и теорию экономического роста. Умер он в январе 1930 года, не дожив до двадцати семи. Статью «Истина и вероятность» он написал в конце 1926 года и прочитал в Кембридже в Клубе моральных наук, а напечатана она была только посмертно, в 1931 году.

Возражение Рамсея Кейнсу было обезоруживающе простым: он не видит этих логических отношений, и он подозревает, что другие тоже их не видят, потому что никто не может договориться о них в самых простых случаях. Зато существует нечто, что несомненно есть у каждого человека: степень уверенности. И её можно измерить. Обычный способ узнать, насколько человек в чём-то уверен, старше всякой философии: предложить ему пари.

Предположим, вы считаете справедливыми такие условия: вы получаете сто рублей, если завтра будет дождь, и отдаёте сто, если не будет. Значит, ваша уверенность в дожде равна половине. Если вы готовы поставить триста рублей против ста, значит, ваша степень уверенности равна трём четвертям. В общем виде: готовность рискнуть суммой X ради выигрыша единицы означает степень уверенности X, делённое на X плюс один. Разумеется, у настоящих людей есть склонность к азарту и страх перед риском, а сто рублей для бедного и для богатого значат разное. Рамсей это прекрасно понимал и поэтому построил более тонкую конструкцию, в которой мера уверенности и мера ценности благ определяются вместе из предпочтений человека. Но суть остаётся: уверенность определяется не интроспекцией и не логической интуицией, а тем, как человек готов действовать. Вероятность перестала быть свойством мира или логики и стала свойством носителя убеждения.

Но тут немедленно возникает возражение: если вероятность — просто степень чьей-то уверенности, то любой безумец прав. Один считает возвращение пришельцев вероятным на девяносто процентов, другой на одну миллиардную, и кто им судья? Ответ Рамсея и есть ключевой ход всей теории. Законы теории вероятностей, говорил он, не говорят, во что вам следует верить. Они говорят, какие наборы убеждений согласованы, а какие нет. В этом они подобны обычной логике: логика тоже не проверяет, истинны ли ваши посылки, она только сообщает, что из них следует, и запрещает противоречие. Если вы считаете, что Сократ человек и что все люди смертны, вы не можете считать Сократа бессмертным. Если вы считаете дождь вероятным на семьдесят процентов, вы не можете считать его отсутствие вероятным на шестьдесят.

За нарушение этого запрета в логике полагается только упрёк в непоследовательности. А здесь следует вполне ощутимое наказание. Его называют голландской книгой. Слово «книга» здесь означает букмекерский набор ставок. Почему книга голландская, в точности никто не знает, и Годфри-Смит в своём учебнике извиняется перед читателями-голландцами. Вот простейший пример в его изложении, только в рублях. Предположим, что вы оцениваете вероятность орла при броске монеты в 0,6 и вероятность решки тоже в 0,6. В сумме получается больше единицы, и правила вероятностей нарушены, но вас это пока не беспокоит. При степени уверенности 0,6 вы считаете справедливой ставку 150 рублей против 100 в обоих случаях. И вот приходит букмекер и предлагает вам обе ставки сразу. Вы их принимаете: обе кажутся вам справедливыми. Монета падает. Если выпал орёл, вы выигрываете сто рублей на первой ставке и проигрываете 150 на второй. Если выпала решка, то же самое наоборот. В любом случае вы

теряете 50 рублей, что бы ни произошло в мире. Вас обыграли не потому, что вы ошиблись насчёт монеты. Вас обыграли потому, что ваши убеждения несогласованы между собой.

Математически доказывается и общее утверждение. Если степени уверенности человека нарушают правила теории вероятностей, то против него можно составить такой набор ставок, а если соблюдают, то нельзя. А отдельное рассуждение того же типа оправдывает и само правило обновления: тот, кто меняет свои оценки при поступлении данных не по формуле Байеса, тоже уязвим для букмекера, действующего в несколько ходов.

Пример с букмекером часто вызывает у читателя законное недоумение. Что за дело физику до пари? Никто не заставляет его заключать ставки, да и голландских букмекеров на свете немного. Но, как и подчёркивает Годфри-Смит, суть не в букмекере. Букмекер здесь такая же наглядная выдумка, как поезда у Эйнштейна. Настоящий смысл в том, что произвольные на первый взгляд правила теории вероятностей оказываются условиями внутренней последовательности человека, которому вообще приходится действовать при неопределённости. А такому положению подвержены все.

Следующий шаг сделал итальянский математик Бруно де Финетти (1906–1985). Он пришёл к тем же идеям независимо от Рамсея, изложил их на конгрессе математиков в Болонье в 1928 году и окончательно в лекциях, напечатанных в 1937 году. Де Финетти был радикальнее. Для него никакой вероятности помимо степени уверенности вообще нет, и в предисловии к итоговой книге он написал прописными буквами: вероятности не существует. Это заявление надо понимать буквально. Он утверждал, что вероятность объективная, какое-то свойство монеты или мира, есть такое же мнимое понятие, как флогистон или эфир, а те, кто говорит о «случайности» монеты, просто неудачно выражают собственное незнание. Логика вероятного, писал он, не добавляет от себя ничего, а только помогает увидеть то, что уже содержалось в наших исходных оценках.

Но что же тогда делать с частотами и с очевидным успехом статистики? Здесь де Финетти предъявил математический результат, из-за которого его подход и стал работоспособным. Этот результат известен как теорема о представлении. Если человек считает порядок испытаний неважным (то есть для него важно, сколько выпало орлов, а не в какой последовательности), то при поступлении всё новых данных он будет рассуждать в точности так, как если бы у монеты была неизвестная объективная частота орла, и его оценка будет к этой частоте приближаться. Иначе говоря, поведение, которое кажется признанием объективной вероятности, можно получить без всякой объективной вероятности. Понятие случайности оказывается удобной моделью для упорядочивания наших ожиданий, а не сведением о том, что существует в монете.

Читатель прошлых томов может заметить здесь знакомую интонацию. Юм говорил, что мы не видим необходимой связи, а видим только привычку ожидания. Мах говорил, что сила в механике есть описание, а не вещь. Де Финетти говорит это о случайности. Причём для этого подхода вопрос о том, что в действительности стоит за нашими оценками, просто не возникает. Степени уверенности оцениваются по тому, согласованны ли они и хорошо ли срабатывают. Такое понимание вероятности вполне устраивает того, кто считает теории инструментами для связи одних наблюдений с другими. И в то же время оно оскорбляет другую часть философов, которые считают, что есть факт, а уверенность должна с ним сообразовываться. К этому столкновению мы ещё вернёмся.

Третьим в этом ряду обычно называют американского математика Леонарда Сэвиджа, чья книга «Основания статистики» (1954) соединила рамсеевскую идею с полной математической теорией решений. Сэвидж показал: если предпочтения человека удовлетворяют нескольким требованиям последовательности, то его поведение можно описать так, как будто у него есть определённая система вероятностей и определённая шкала ценностей, и он выбирает наибольшую ожидаемую выгоду. С этого времени такой подход стал рабочим инструментом эко-

номистов и специалистов по решениям. И с этого же времени начинается то, что называют неклассической, а чаще просто байесовской статистикой.

Вся вторая половина XX века прошла в статистике под знаком войны двух лагерей. Частотники напоминали, что наука должна быть объективна, и обвиняли байесианцев в том, что те подмешивают в анализ субъективные мнения. Байесианцы отвечали, что субъективность есть у всех, только частотники прячут её в выборе модели и порога значимости, и насмешливо перечисляли логические нелепости частотных процедур. Практика при этом шла своим путём. Методы, связанные с обратной вероятностью, применялись тихо там, где результат важнее чистоты теории. По некоторым свидетельствам, ими пользовался во время войны в Блетчли-парке Алан Тьюринг, взламывая немецкие шифры, а потом их применяли при поиске подводных лодок и в прогнозах погоды. Настоящее распространение пришло с компьютерами, потому что байесовские расчёты для сколько-нибудь сложных задач вручную почти невозможны. Сегодня статистики всё чаще относятся к обоим подходам как к разным инструментам, а не вероисповеданиям.

Как устроены эти методы изнутри, что такое значимость, доверительный интервал и мощность критерия и откуда взялись все эти сорта вероятности, мы здесь разбираем лишь настолько, насколько этого требует спор о свидетельстве. Математическая статистика заслуживает своего подробного и столь же нескучного разговора, и популярную книгу о ней мы тоже скоро напишем: это наши ближайшие планы. А здесь пойдём дальше по философской линии и посмотрим, что байесианство способно объяснить в работе учёных.

Что байесианство объясняет

В 1989 году английские философы Колин Хоусон и Питер Урбах выпустили книгу «Научное рассуждение», которая выдержала три издания и стала главным манифестом байесианства в философии науки. Их замысел был смелым: взять не отдельные примеры, а весь накопленный XX веком список загадок о подтверждении и показать, что они решаются одной формулой, причём без дополнительных допущений. Посмотрим, что из этого вышло. Для начала вернёмся к тем загадкам, что остались от Гемпеля.

Вот белый ботинок. По Гемпелю, он подтверждает гипотезу о чёрных воронах. Байесианец отвечает на это неожиданно спокойно: да, может подтверждать, но вес его ничтожен, а иногда и вовсе обратен. Всё зависит от того, что нам уже известно. В статье о подтверждении, написанной американскими философами Аланом Хайеком и Джеймсом Джойсом, разобрано несколько случаев. Если мы знаем, что воронов либо очень мало и все они чёрные, либо их много и попадаются белые, то встреча с чёрным вороном говорит скорее в пользу второго варианта и потому снижает вероятность гипотезы о том, что все они чёрные. На это ещё в 1967 году указал английский статистик Ирвинг Гуд. Если же положение такое, как в действительности, и нечёрных предметов в мире неизмеримо больше, чем воронов, то и чёрный ворон, и белый ботинок поднимают вероятность гипотезы, только в очень разной мере. Проверить одного ворона значит серьёзно рискнуть гипотезой: он мог оказаться белым. Проверить один нечёрный предмет почти ничем не рисковать: он почти наверняка окажется не вороном. Отсюда и наше ощущение, что орнитология не делается в гардеробной. Интуиция Гемпеля и интуиция здравого смысла примиряются без всякой логической хирургии.

Вот гномы, которых мы приклеили к закону. Если H предсказывает E, то и «H и гномы» предсказывает E, и E действительно поднимает вероятность этой составной гипотезы. Логика ничего не может с этим поделать. Но формула показывает, насколько поднимает: прирост умножается на вероятность гномов при условии H. А эта величина мала, так что гномам от чужого подтверждения почти ничего не достаётся. Глупое следствие оказывается не ошибкой теории, а верным, но ничтожным численным фактом.

Интереснее обстоит дело с проблемой Дюгема, которую мы подробно обсуждали во втором томе. Опыт проверяет не гипотезу, а гипотезу вместе со всеми вспомогательными допущениями: что прибор исправен, что образец чист, что наблюдатель не ошибся. Когда предсказание не сбывается, логика говорит лишь, что что-то неверно, но не говорит, что. Дюгем ссылался на «здравый смысл» учёных, не поясняя, как он устроен. Байесианец предлагает расчёт. Если у нас есть априорные вероятности основной гипотезы и вспомогательного допущения, то формула определяет, какая из них пострадает от неудачи сильнее. Иногда неудачный опыт понижает основную гипотезу и почти не трогает допущение. Иногда бывает наоборот: допущение падает, а гипотеза стоит. Возможен даже случай, когда гипотеза от опровержения слегка выигрывает.

Хоусон и Урбах разбирают случай английского врача и химика Уильяма Праута. В 1815 году он предположил, что атомные веса всех элементов кратны весу водорода. Многие элементы подтверждали это прекрасно. Но хлор упорно давал вес около 35,5. По Попперу, гипотезу надо было отбросить. А химики XIX века долго этого не делали, полагая, что дело в недостаточной очистке образцов и неточности методов. Это выглядит как нечестная защита от опровержения. Но если оценить, насколько уверены были тогда в красивой гипотезе и насколько в методах очистки, расчёт показывает, что основной удар должны были принять на себя методы. Химики были, в этом смысле, рациональны. А в XX веке обнаружилось, что и гипотеза, и данные были по-своему верны: природный хлор есть смесь изотопов с разным весом.

Отсюда же получается и объяснение тому, что так смущало в Куне. Кун показал, что нормальный учёный не бросает основную теорию из-за аномалии, а ищет ошибку у себя. Его последователи увидели в этом иррациональность. Но если основная теория имеет высокую априорную вероятность, потому что за ней много успехов, а вспомогательные допущения имеют более низкую, то именно такое поведение и предписывается формулой. Перестаёт быть загадкой и обратное: когда аномалии накапливаются, вероятность основной теории всё-таки падает, и в какой-то момент рационально начать искать другую. Лакатосовский защитный пояс вокруг твёрдого ядра тоже получает численное истолкование. Для философии, которая полвека билась над этими вопросами, это выглядело сильно.

Та же логика объясняет, почему разнообразие свидетельств ценится выше простого повторения. Первое повторение опыта поднимает вероятность гипотезы, но и делает следующее повторение менее неожиданным. Если результат уже получен сто раз одним и тем же способом, то сто первый почти ничего не добавит, потому что его и так все ждали. Зато результат другим способом, другим прибором, в другой области, для которого прежнее подтверждение почти ничего не предсказывало, добавит много. Заметим одну важную оговорку, которая понадобится нам в следующей главе: сколько ни повторять опыт с одной и той же систематической ошибкой, вероятность того, что результат вызван этой ошибкой, не снизится. Перл вспоминает, как Фишер, защищая табак, язвительно замечал, что девятнадцать одинаково смещённых исследований не лучше одного. В этом пункте Фишер был прав, хотя в главном вопросе и он, и табак ошибались.

А как быть с Поппером? Байесианство включает его в себя как предельный случай. Если гипотеза H логически влечёт E, а наблюдается отсутствие E, то вероятность гипотезы по формуле падает точно до нуля. Это и есть опровержение: его можно получить как частный случай правила, при совершенно точном и надёжном наблюдении. То, что Поппер называл смелой гипотезой и строгой проверкой, тоже находит точное выражение: чем менее вероятно предсказанное без гипотезы, тем сильнее его удача её поднимает. Гипотеза, пережившая проверку, грозившую её погубить, становится вероятнее, и чем больше была угроза, тем заметнее. Но Поппер считал, что ничего подобного происходить не может, и всю жизнь с этим боролся. Его главный довод заключался в том, что всякий всеобщий закон, охватывающий бесконеч-

ное число случаев, имеет нулевую вероятность. Байесианцы отвечают, что это следует только из определённого способа распределять вероятности, вроде карнаповского, а никто не обязан распределять их так. Инженер, выбирающий проверенную теорию для моста, теперь получает то, чего ему не дал Поппер: основание считать проверенную теорию более вероятной. Именно поэтому многие физики и биологи, называющие себя попперианцами, рассуждают в действительности как байесианцы, сами о том не догадываясь.

Наконец, простота. Почему из бесконечного множества кривых, проходящих через те же точки, мы выбираем самую простую? Английский геофизик Гарольд Джеффрис, одна из центральных фигур байесианства в период его непопулярности, заметил, что сложная кривая с большим числом настроек способна подогнаться под что угодно. Если мы будем менять закон с каждым новым наблюдением, писал он, то исчезнет сама идея, что закон что-либо значит за пределами исходных данных. В байесовском изложении это выглядит так: гипотеза, которая могла объяснить любой результат, распределяет свою вероятность между очень многими исходами и каждому достаётся мало. Гипотеза узкая ставит на один исход. Если исход случился, то узкая гипотеза выигрывает. Это рассуждение прямо применимо к демаркации, поскольку теория, объясняющая что угодно, расплачивается за всеядность весом каждого отдельного объяснения. Психоаналитик или конспиролог, объясняющий любой исход, в итоге получает от любого исхода почти ничего. Правда, Джеффрис и его наследники вынуждены здесь вводить дополнительные допущения о простоте, и их обоснованность остаётся спорной. Ниже мы к этому вернёмся.

Добавим ещё один пример, который показывает одновременно силу теории и слабость нашей интуиции. В сентябре 1990 года американская журналистка Мэрилин вос Савант ответила читателю популярного журнала на задачу из телевикторины. Перед вами три двери. За одной автомобиль, за двумя козы. Вы выбираете дверь. Ведущий, который знает, где что находится, открывает другую и показывает козу. Стоит ли теперь поменять выбор? Вос Савант ответила: да, это удваивает шанс на выигрыш, с одной трети до двух третей. Она получила больше десяти тысяч писем, многие от людей с учёными степенями, и писали ей с раздражением и поучениями. Венгерский математик Пал Эрдёш, говорят, поверил только компьютерной симуляции. Расчёт по Байесу вполне прост: дверь, которую вы выбрали, не подвергалась никакой опасности, потому что ведущий не мог её открыть. А третья дверь подверглась и выжила. Но самое важное обнаруживается в варианте задачи. Если ведущий не знает, где автомобиль, и открывает дверь наугад, причём случайно оказывается коза, то данные выглядят точно так же, а менять выбор бесполезно: вероятности равны.

Те же самые факты, та же открытая дверь и та же коза, а ответ другой. Вес свидетельства зависит не только от самого факта, но и от того, каким путём он до нас дошёл. Как заметил Перл, способ получения сведений важен не меньше самих сведений. Запомним это. Именно этим различием статистика Фишера пренебрегала, называя свою задачу сокращением данных. И именно этого не замечают те, кто выбирает из десятка анализов удачный и докладывает его как единственный. Но об этом следующая глава.

Впечатляющий список, не правда ли? Подтверждение, вороны, гномы, Дюгем, Кун, Поппер, простота. В конце XX века многие всерьёз считали, что центральная проблема философии науки решена. И здесь нам пора вспомнить совет самой теории: когда кто-то предъявляет объяснение всего при помощи одной формулы, спроси, не слишком ли дёшево оно досталось. Обратите внимание на слова, которыми заканчивались почти все перечисленные решения: «если у нас есть априорные вероятности». Откуда?

Возражения: откуда берётся уверенность

Сам Годфри-Смит, чьё изложение мы использовали выше, прямо пишет, что не принадлежит к байесовскому лагерю, хотя признаёт силу этих идей. Вот что заставляет его и многих других колебаться.

Откуда брать априорные вероятности? На этот вопрос у байесианцев два ответа, и оба неудобны. Объективисты, такие как Джеффрис и американский физик Эдвин Джейнс, считают, что существуют правильные исходные распределения, и надо лишь найти правило для их выбора. Самое старое правило называется принципом безразличия: если нет оснований предпочесть один исход другому, считай их равновероятными. Кажется очевидным, но ведёт к противоречиям, поскольку результат зависит от того, в каких словах описать варианты. Пусть о кубике известно только то, что ребро его имеет длину от одного до двух сантиметров. Значит, объём от одного до восьми кубических сантиметров. Безразличие по длине ребра говорит, что с равной вероятностью ребро окажется короче полутора сантиметров или длиннее. Ребру в полтора сантиметра соответствует объём примерно 3,4. Безразличие по объёму говорит, что равновероятно окажется меньше или больше объёма 4,5. Два ответа на один вопрос, при одинаковом незнании. Как замечает Хоусон, выбор исходного разбиения на равные части зависит от выбора языка описания. Джейнс предложил более изощрённое правило максимальной неопределённости, но там, где никаких исходных данных нет, оно возвращается к тому же равномерному распределению со всеми его бедами. Джеффрис предлагал давать большую вероятность более простым законам, но что считать простым, снова зависит от языка.

Субъективисты, как Рамсей и де Финетти, разрубают узел: не надо никакого правила. Выбирайте какие хотите априорные вероятности, лишь бы они были согласованы между собой. Логика тоже не проверяет посылок. Этот ответ снимает все вычислительные трудности. Но посмотрите, что за него приходится платить. Вообразим человека, который начал с вероятности 0,99 в пользу того, что положением планет определяется судьба. Он согласован, следует формуле, ни один букмекер его не поймает. С точки зрения строгого субъективиста упрекнуть его не в чем. Теория, которая должна была объяснить, чем наука рациональнее астрологии, объявляет астрологию рациональной, если только она последовательна в своих вычислениях.

Смоет ли данные всю разницу? Байесианцы отвечают, что всё это не страшно. Начальные вероятности со временем смываются поступающими данными. Существуют точные теоремы о схождении мнений: если два человека начинают с разных априорных вероятностей и видят одни и те же данные, то их оценки сближаются и в пределе становятся сколь угодно близкими. Астролог и астроном, стоит им только наблюдать достаточно долго, должны сойтись. Вот объяснение объективности науки, и притом без всякой объективной вероятности. Красиво.

Возражений несколько. Сближение может занять сколько угодно времени, и никто не знает сколько. Американский философ Генри Кайберг любил напоминать обратную сторону той же теоремы: при любом количестве данных найдутся такие исходные оценки, что двое и в конце останутся несогласными. Хоусон указывает ещё на математическое обстоятельство: сильнейшие теоремы о схождении опираются на спорную аксиому счётной аддитивности, которую сам де Финетти отвергал как принятую ради математического удобства. Но важнее возражение Годфри-Смита. Все эти доказательства предполагают, что два человека, расходясь в априорных вероятностях гипотезы, совпадают в правдоподобиях: оба одинаково оценивают, какова вероятность каждого наблюдения при истинной гипотезе и при ложной. В задаче с монетой это естественно. Но в настоящем споре правдоподобие зависит от всего, во что человек верит.

Возьмём сторонника теории заговора. Для него отсутствие доказательств заговора вполне ожидаемо, раз заговорщики хорошо заметают следы. Опровержение от учёных тоже ожида-

емо, раз они подкуплены. Настоящие улики ещё более ожидаемы. Его правдоподобия устроены так, что любое возможное наблюдение либо поднимает гипотезу, либо оставляет как есть. Он согласован. Он честно применяет формулу. И он не сойдётся со своим оппонентом никогда. Бельгийский философ Маартен Будри, о котором подробно пойдёт речь в последней главе, называет подобные приёмы защитными механизмами системы убеждений. Настоящая разница между наукой и её имитацией оказывается не в формуле обновления, а в том, откуда взялись сами правдоподобия и можно ли их проверить независимо от гипотезы. Этого теорема о схождении дать не может.

Догматик. Рассмотрим крайний случай. Если априорная вероятность гипотезы равна единице или нулю, то формула перестаёт работать совсем: ноль после любого умножения останется нулём, и никакое свидетельство его не сдвинет. Английский статистик Деннис Линдли, один из вождей байесианства второй половины века, назвал это правило именем Кромвеля. 3 августа 1650 года, перед сражением при Данбаре, Оливер Кромвель писал пресвитерианам шотландской церкви, совершенно уверенным в своей правоте: заклинаю вас, подумайте, что можете ошибаться. Правило Кромвеля гласит: никогда не присваивай эмпирическому утверждению вероятность ноль или единицу. Оставь немного, говорил Линдли, хотя бы одну миллионную, и на то, что Луна сделана из зелёного сыра, иначе армия астронавтов с образцами сыра ничего для тебя не изменит.

Вот это, пожалуй, самое важное, что байесианство может сказать о демаркации. Важно не содержание убеждений. Важно, оставлен ли зазор, через который данные могут войти. Прайс и его миллион приливов учили тому же: очень мало и ноль — вещи разной природы. Потомственный ясновидящий, которого ничто не может разубедить, и разоблачитель, которого тоже ничто не может разубедить, стоят с этой точки зрения рядом. У науки, если говорить коротко, нет привилегированного содержания, зато есть требование не закрывать дверь. Но заметим и другое: правило Кромвеля не следует из теории вероятностей. Это дополнительное требование, и притом нормативное. Согласованность сама по себе его не даёт, а вероятность в одну стомиллиардную мало чем отличается от нуля на практике. Придётся сказать, что к логике согласованности уже что-то добавлено со стороны, и добавлено именно то, что нас интересует.

Старые свидетельства. В 1980 году американский философ Кларк Глимур, о котором мы подробно поговорим в третьей главе, указал на затруднение, которое кажется то пустяковым, то смертельным. С середины XIX века астрономы знали, что перигелий Меркурия смещается не совсем так, как положено по Ньютону. В ноябре 1915 года Эйнштейн обнаружил, что его новая теория даёт как раз нужную величину, и, по его собственным воспоминаниям, несколько дней не помнил себя от радости. Физики считали это одним из сильнейших свидетельств в её пользу, вероятно, более сильным, чем предсказанное отклонение света. Но смещение перигелия было уже известно. Значит, его вероятность равна единице, а значит, формула не даёт никакого прироста: и при данной теории, и при всех других вероятность известного факта равна единице. Получается, что старый факт ничего не подтверждает. А ведь именно старый факт и убедил физиков.

Предложено несколько выходов. Может быть, подтверждает не сам факт, а открытие, что теория его влечёт. Хоусон возражает: логические отношения трудно считать случайными величинами, да и не всякое подтверждение устроено так. Может быть, надо считать по той вероятности, какую факт имел бы, если бы мы о нём не знали. Это правдоподобно, но как определить такое знание за вычетом одного факта, если он переплетён со всем остальным? Хайек и Джойс предлагают различать два понятия опоры: прирост, который свидетельство даёт сейчас, и вес, какой оно имеет в общей совокупности данных. Старые свидетельства не дают прироста, но имеют вес. Вопрос этот не закрыт, а его значение в том, что формула, как обнаруживается, описывает не логическую связь гипотезы и свидетельства, а биографию одной головы, и в этой

биографии есть порядок поступления сведений. Но ведь у науки биография другая, она коллективна.

Логическое всеведение. Правила вероятностей требуют, чтобы всякой логической истине приписывалась единица. Значит, идеальный байесианец знает все следствия из своих убеждений и все теоремы математики. Настоящий учёный ничего подобного не знает, а открытие того, что из его теории следует неожиданный вывод, есть основная часть его работы. Хоусон замечает, что упрёк теряет силу, если считать аксиомы описанием согласованности, а не предписанием к действию. Но тогда теория рациональности теряет немало своей предписывающей силы.

Гипотеза, которой ещё никто не придумал. В знаменателе формулы стоит вероятность свидетельства при всех остальных возможностях. Что входит в эти остальные возможности? Всё, что ещё не пришло никому в голову. Каким числом Ньютон или Лаплас могли бы оценить вероятность движения планет при условии, что истинна неньютоновская теория, если никто не имел о ней понятия? Обычно поступают просто: сравнивают уже имеющиеся теории и молча считают, что других нет. Но это означает, что апостериорная вероятность в 0,99 говорит только о том, какая из имеющихся теорий лучше. Появление же новой теории в байесовской схеме вообще не предусмотрено: распределение было задано заранее на всём, что мыслимо, а новая идея по определению в него не входила. Ни одно правило не говорит, как перераспределить уверенность, когда кто-то выдвигает гипотезу, о которой прежде не думали. А ведь это и есть самое интересное событие в истории науки. Разговор об этой трудности мы продолжим в восьмой главе у Кайла Стэнфорда, для которого она стала основанием целой позиции.

Зелубые изумруды никуда не делись. Стандартный байесовский ответ Гудмену таков: обе индукции законны, и зелёная, и зелубая. Наблюдения поддерживают обе одинаково. Но у зелёной гипотезы большая априорная вероятность, поэтому она и после наблюдений остаётся впереди. Годфри-Смит спрашивает: а почему? Что мешает кому-то начать наоборот? Строгий субъективист не может сказать ничего. Загадку не решили, а только перенесли в то место теории, где правила молчат. И это место, заметим, то же самое, что и в предыдущих случаях. Вся содержательная работа совершается в априорных вероятностях и в правдоподобиях, а сама формула лишь аккуратно передаёт её дальше.

Строгая проверка. Самая основательная альтернатива пришла со стороны, которая вообще не желает приписывать гипотезам вероятности. Американский философ Дебора Мейо в книге «Ошибка и рост экспериментального знания» (1996), а затем в большой книге о статистических спорах (2018), предлагает переформулировать сам вопрос. Важно не то, насколько мы уверены в гипотезе, а то, насколько тщательно её проверили. Свидетельство подтверждает гипотезу, если она выдержала строгую проверку: такую, что при ложной гипотезе эта процедура с большой вероятностью дала бы результат, хуже с ней согласующийся. В основе лежат всё те же частоты, но вычисляются они не для гипотез, а для процедур. Какова надёжность этого способа проверки, как часто он обманывает?

Одно из заглавий Мейо передаёт суть различия: высоковероятная гипотеза и хорошо проверенная гипотеза не одно и то же. Гипотеза может быть очень вероятной, потому что в неё сильно верили заранее, и при этом никто и никогда её по-настоящему не пытался опровергнуть. В такой картине Поппер слышится отчётливее, чем Байес. Самое чувствительное место здесь следующее. Многие байесианцы принимают так называемый принцип правдоподобия: всё, что данные говорят о гипотезе, содержится в их правдоподобии, а какими были намерения экспериментатора, например, когда он решил бы прекратить набор данных, значения не имеет. Мейо настаивает, что это катастрофа. Если исследователь проверяет результат после каждого нового участника и останавливается, как только значение стало удачным, то при частотном подходе он обязан это учесть, поскольку у такой процедуры высока вероятность обмануть. А правдоподобию самих данных до этого нет дела. Байесианцы отвечают, что подобные проце-

дуры надо учитывать в модели, а не в философии, и что априорная вероятность автоматически защищает от случайной удачи. Этот спор далеко не окончен, и он станет главной темой второй главы, где выяснится, что вопрос о том, каким путём были получены данные, стоит очень дорого. Заметим только, что с Мейо не вполне соглашаются и сами частотники. Но проделанная ею работа сильно мешает всем, кто считает, что вопрос можно решить одной формулой.

Вопрос о причине. Наконец, самое неожиданное возражение исходит от человека, который в 1980-е годы внёс едва ли не наибольший вклад в практические успехи байесианства, то есть от Перла. Он создал байесовские сети, позволившие считать по этой формуле задачи с тысячами переменных, и потом сам назвал себя отступником. Вероятности, заметил он, описывают только то, что мы видим, и не различают, что чем вызвано. Распределение вероятностей одинаково хорошо записывается и так, что огонь повышает вероятность дыма, и так, что дым повышает вероятность огня. Но зажечь огонь — способ получить дым, а напустить дыма в комнату не значит её поджечь. Более того, выводы о причинах не становятся менее субъективными с ростом данных, тогда как априорные вероятности со временем смываются. Если Перл прав, то байесианство, каким бы совершенным оно ни было, неспособно выразить главное, что нас интересует в науке. Об этом третья глава.

И последнее замечание. Психологи Даниэль Канеман и Амос Тверски показали, что живые люди не считают по Байесу, и пример с маммограммой это хорошо подтверждает. Обычно это не возражение, ведь теория говорит, как нужно рассуждать, а не как рассуждают. Но у этого ответа есть обратная сторона. Учёные тоже не приписывают числа своим гипотезам и вряд ли смогли бы это сделать, если их спросить о вероятности общей теории относительности. Ещё Годфри-Смит замечает, что наука, возможно, сильна не потому, что в голове каждого учёного работает маленький правильный вычислитель, а потому, что определённые процедуры и определённое устройство сообщества заставляют идеи соприкасаться с наблюдением. Может быть, байесианство прекрасно описывает, как должна рассуждать идеальная одинокая голова. Но наука — не одинокая голова. Вопросу о том, что такое коллективная рациональность, и о том, можно ли ей доверять, посвящена двенадцатая глава.

Итог и вопрос читателю

Что же остаётся от байесовской революции после всех возражений? Остаётся довольно много, но не то, что обещали её самые восторженные сторонники.

Решением проблемы индукции байесианство не стало. Юм мог бы прочесть все теоремы о схождении и спокойно спросить, откуда известно, что мир и завтра будет выдавать те же правдоподобия. Не стало оно и готовым критерием науки: формула равно служит и астроному, и последовательному астрологу, если только оба согласны в своих числах. Но оно дало нечто другое, и притом очень полезное: общий язык, на котором видно, где именно спрятано допущение. Когда двое расходятся в выводах из одних данных, теперь можно спросить, в чём они расходятся: в исходной уверенности, в оценке того, что должно было наблюдаться при каждой из возможностей, или просто в арифметике. Спор о свидетельстве становится разбором конкретных чисел. Часть прежних загадок при этом исчезает, а оставшиеся получают точный адрес.

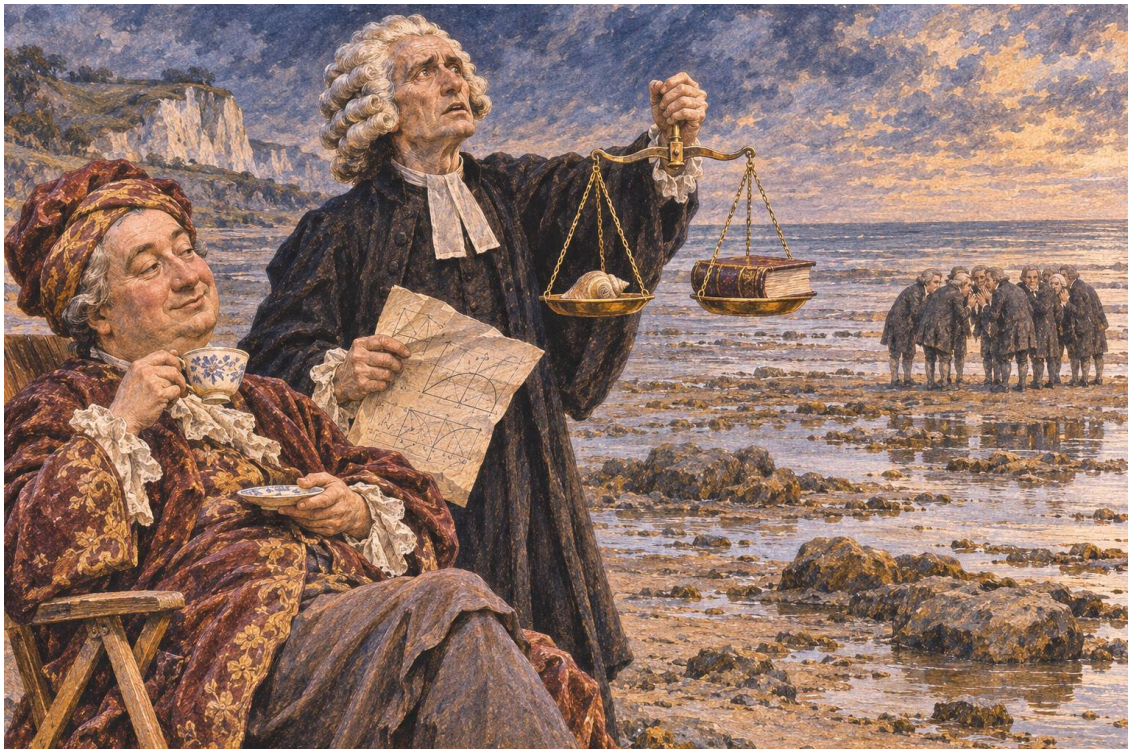
Немаловажно и то, чего эта теория не требует. Де Финетти и Рамсей показали, что можно говорить о рациональной уверенности, не утверждая ничего о природе вещей. Степени доверия не отражают мира как зеркало. Это способ согласовывать ожидания и проверять их на дальнейшем опыте. Перейдёт ли это в более общую точку зрения на науку, мы увидим значительно позже.

А самый практичный её вклад понятен и без формул. Спрашивай не только о том, ожидается ли этот факт при твоей любимой гипотезе, но и о том, ожидался бы он, если гипотеза

ложна. Считай независимых свидетелей, а не громкость голосов. Помни об исходной частоте. И не присваивай ничему вероятность единицы, если речь идёт о мире. Этих четырёх правил хватает, чтобы защититься от большинства экстрасенсов, рекламных роликов и половины новостей, а заодно они довольно неприятно действуют на собственные любимые мнения.

Однако теория оставила нерешённым вопрос, ради которого мы начали этот том. Она говорит, как менять уверенность при поступлении данных, если эти данные получены честно и обозначены полностью. Но что если сама процедура получения данных устроена так, что систематически выдаёт ложные удачи? Если публикуется только то, что вышло, а опыты с отрицательным результатом лежат в столе? Тогда самый дисциплинированный байесианец будет исправно обновлять свои оценки по искажённой выборке, хотя никто сознательно ничего не искажал. К этому мы и переходим.

Вопрос читателю. Назовите убеждение, в котором вы уверены почти полностью: о своём здоровье, о близком человеке, о том, как устроено общество и кто в нём виноват. Какое конкретное наблюдение снизило бы вашу уверенность хотя бы вдвое? Попробуйте описать его так, чтобы другой человек мог сказать, случилось оно или нет. Если такого наблюдения придумать не удаётся, то данные здесь уже ничего не решают, и стоит честно признаться себе, почему вы держитесь за это убеждение.



Глава 2. Кризис воспроизводимости

Предвидение в ведущем журнале

В 2011 году в «Журнале личности и социальной психологии», одном из самых уважаемых изданий своей области, появилась статья под названием «Чувствуя будущее». Её автор, почётный профессор Корнеллского университета Дэрил Бем, был известен не как маргинал, а как один из видных социальных психологов своего поколения, создатель теории, которую до сих пор излагают в учебниках. Статья сообщала о девяти опытах с участием более тысячи студентов. В восьми из девяти получен статистически значимый результат. Вывод: люди способны, пусть слабо и бессознательно, предчувствовать будущие события.

Устройство опытов было остроумным. Бем брал хорошо изученные психологические эффекты и переворачивал их во времени. Всем известно, что слова, которые человек повторял, он запоминает лучше. В девятом опыте участникам показывали список слов и просили вспомнить как можно больше. А уже потом программа случайным образом отбирала часть слов и заставляла их выписывать. И оказалось, что участники лучше вспоминали те слова, которые им только предстояло повторять. Упражнение как будто действовало на память в обратном направлении. В другом опыте студент угадывал, за какой из двух занавесок на экране спрятана картинка, причём компьютер выбирал место уже после ответа. Для картинок эротического содержания угадывание превышало случайный уровень.

На следующий день заголовки сообщали то, чего всегда ждали потомственные ясновидящие: наука признала. Передачи о людях со сверхспособностями могли бы показывать страницы журнала прямо в заставке. Вот журнал, вот рецензенты, вот значимость. Чего ещё вы хотели от нас, господа скептики, все эти годы?

Редакторы журнала прекрасно понимали, что печатают. Они сопроводили статью комментарием, в котором отметили, что их собственные убеждения тут ни при чём. Работа соответствовала принятым в дисциплине требованиям и прошла обычное рецензирование, а значит, отказывать ей только за неправдоподобность вывода было бы неправильно. Вот это, собственно, и есть центральный момент всей истории. Если бы Бем нарушил правила, его можно было бы поймать и забыть. Но он их соблюдал.

Дальше события приняли характер притчи. Бем, к его чести, сам призвал других повторить опыты и предоставил все материалы. Трое британских исследователей, Стюарт Ричи, Ричард Уайзман и Кристофер Френч, откликнулись на призыв. Они трижды независимо повторили девятый опыт, тот самый, что дал наибольший эффект, заранее зарегистрировав свой план. Ничего не вышло, память упрямо не желала работать в обратную сторону. Статью они послали в тот же самый журнал. И журнал отказал, даже не отправив её на рецензию. Отказ, как потом рассказывали авторы, объяснялся простым правилом: журнал не печатает повторений, ни удачных, ни неудачных.

Теперь посмотрим на положение дел в целом. Ведущий журнал охотно печатает поразительное открытие и отказывается напечатать проверку, потому что проверка не является новостью. Правила устроены так, что поразительный результат проходит, а его опровержение нет. Журналу, разумеется, никто не хотел зла. Он следовал своим правилам, и те казались разумными: журналы хотят нового, читатели хотят нового, повторять чужое скучно. Статья Ричи и его коллег вышла в итоге в другом журнале, открытого доступа. Позже и этот журнал смягчился: в 2012 году он всё-таки напечатал работу Джонатана Галака и его соавторов, которые провели семь повторений с участием 3289 человек, втрое больше, чем было у Бема, и не нашли ничего.

Читатель, вероятно, догадывается, что заявленным результатам Бема мало кто поверил. Но именно здесь и возникает настоящая трудность. Психологи читали статью и говорили, что выводу они не верят. А на вопрос, что же в ней не так, отвечать приходилось с большим трудом. Методы те же самые. Статистика та же самая. Значимость та же самая. Точно такими же способами получены тысячи работ о том, как влияет на поведение цвет стены, поза, название продукта, недавно прочитанное слово. Если этот инструмент подтвердил ясновидение, то возможны только два вывода. Либо ясновидение существует, либо инструмент способен подтвердить то, чего нет. А тогда как быть со всем остальным, что им подтверждено?

В этом смысле Бем оказал психологии услугу, о которой не просил и на которую, вероятно, не рассчитывал. Неправдоподобность его вывода сработала как контрольный опыт. Обычно исследователь не знает заранее правильного ответа и поэтому не может проверить прибор. А здесь большинство было уверено в ответе заранее, и прибор выдал другой. Как мы говорили в первой главе, неожиданное весит больше. В этом случае неожиданным оказалось поведение самой дисциплины. Именно 2011 год принято считать началом того, что потом назовут кризисом воспроизводимости. Через четыре года вышла та самая работа Носека о сотне повторённых опытов, с которой началось введение к этому тому. Теперь посмотрим, что за этими цифрами стоит.

Что означает «значимо»

Почти весь обсуждаемый здесь спор вращается вокруг одного слова, которое в обычной речи значит одно, а в статистике другое. Когда журналист пишет, что результат значимый, читатель понимает: важный, серьёзный, достоверный. Статистик имеет в виду нечто гораздо более скромное и странное. Мы видели в первой главе, что статистические методы, вошедшие в обиход в XX веке, отвечают на вопрос, поставленный в определённом направлении. Теперь разберём его подробнее.

Уже знакомый нам Рональд Фишер предложил такую процедуру. Возьмём «нулевую гипотезу», утверждение, что никакого эффекта нет: лекарство не действует, удобрение бесполезно, эротические картинки не предчувствуются. Предположим, что она верна, и посчитаем, насколько вероятно при этом получить такие данные или ещё более резкие. Это число называется p-значением. Если оно очень мало, значит, либо случилось редкое совпадение, либо нулевая гипотеза ошибочна. Фишер считал такое число мерой того, насколько данные удивляют нас при отсутствии эффекта, предлагал порог в одну двадцатую как удобный ориентир и настаивал, что единичный опыт ничего не решает: явление установлено, когда мы знаем, как поставить опыт, который будет редко подводить.

Ежи Нейман и Эгон Пирсон подходили к делу иначе. Их интересовало не удивление, а решение. Нужно установить заранее правило, которое при длительном применении будет редко ошибаться двумя способами: объявлять эффект там, где его нет (ошибка первого рода), и не замечать его там, где он есть (ошибка второго рода). Вероятность заметить настоящий эффект назвали мощностью критерия. Отдельный опыт в этой картине не несёт никакой степени доказательства, важны только долгосрочные свойства процедуры, как у контролёра на заводе. Фишер и Нейман враждовали десятилетиями и не сходились почти ни в чём.

А затем случилось то, что может оценить любой лингвист, знакомый с историей смешения языков. Авторы учебников для психологов и биологов, которым нужна была простая инструкция, взяли от Фишера нулевую гипотезу и p-значение, от Неймана и Пирсона жёсткий порог и формулировку решения и соединили всё это в одну процедуру. Ни Фишер, ни Нейман такой процедуры не предлагали и вряд ли бы её одобрили. Немецкий психолог Герд Гигеренцер назвал её гибридом и сравнил с ритуалом. Выглядит он примерно так: придумай нулевую гипотезу об отсутствии всякого эффекта, посчитай p, сравни с 0,05, если меньше, значит, твоя

теория подтвердилась, иди печатать. Не требуется ни конкретной альтернативы, ни величины эффекта, ни расчёта мощности. Весь спор о свидетельстве, о котором мы говорили в первой главе, заменён одним вопросом с двумя возможными ответами.

Эта процедура поощряла сразу несколько заблуждений. Значимый результат принимали за доказательство гипотезы, хотя p-значение ничего о вероятности гипотезы не говорит. Незначимый результат принимали за доказательство отсутствия эффекта, хотя он мог означать лишь маленькую выборку. Величину p принимали за размер открытия. Эти заблуждения были настолько распространены, что в 2016 году Американская статистическая ассоциация сочла нужным выпустить официальное заявление: профессиональное сообщество статистиков разъясняло, что значение p не измеряет вероятность истинности гипотезы и само по себе не является хорошей мерой свидетельства. Обычно такие организации не выступают с заявлениями о смысле базовых понятий своей дисциплины. Для сравнения представим себе Международный астрономический союз, разъясняющий, что звёзды — не дырочки в небесной сфере.

Самое неловкое в этой истории то, что предупреждали давно. Ещё в 1959 году американский статистик Теодор Стерлинг подсчитал, что более девяти десятых статей в психологических журналах сообщают об успехе проверяемой гипотезы, и заметил, что так не бывает, если печатать всё подряд. В 1962 году психолог Джейкоб Коэн просмотрел подшивку одного журнала и установил: при эффекте средней величины типичный опыт имел шансы его обнаружить немногим меньше половины. Иначе говоря, авторы тратили месяцы на опыт, равносильный подбрасыванию монеты, и сами этого не знали.

А в 1967 году американский психолог и философ Пол Мил напечатал в журнале Philosophy of Science статью, которую сегодня цитируют как пророчество, под названием «Проверка теорий в психологии и физике: методологический парадокс». Парадокс состоял вот в чём. Физическая теория предсказывает конкретное число. Чем точнее прибор, тем уже допуск и тем труднее теории уцелеть. А психологическая теория обычно предсказывает только направление: одна группа окажется выше другой. Но две настоящие группы людей почти никогда не бывают абсолютно равны, какой-нибудь микроскопической разницей они всегда отличаются. Значит, при всё большей выборке нулевая гипотеза будет отвергнута почти наверняка, причём в нужную сторону с вероятностью примерно одна вторая, даже если теория ни на что не годится. Улучшение метода в физике усложняет жизнь теории, а в психологии облегчает. Мил написал, что статистическая значимость играет в психологии логическую роль, обратную той, что она играет в физике. Читатель узнает здесь Поппера, применённого к статистике: испытание, которое теория проходит почти при любом положении дел, ничего не значит.

Статью уважали и продолжали работать как прежде. В 2002 году немецкие исследователи Хайко Халлер и Штефан Краусс предложили психологам в университетах шесть распространённых неверных толкований значимости и обнаружили, что заблуждения широко распространены даже среди тех, кто сам преподаёт статистику студентам. Понадобились больше полувека предупреждений и один почтенный профессор, доказавший ясновидение, чтобы эти мысли перестали быть уделом кабинетных зануд.

Теперь мы можем разобрать самую цитируемую статью всей этой истории. В 2005 году греко-американский врач и специалист по методологии Джон Иоаннидис напечатал в журнале PLoS Medicine очерк под заглавием, которое не требует дополнительной рекламы: «Почему большинство опубликованных результатов исследований ложны». В нём нет новых экспериментов и почти нет новых данных. В нём есть простой расчёт, знакомый нам по задаче о маммограмме.

Представим себе область науки, где проверяют тысячу гипотез. Многие хорошие идеи неверны, поэтому предположим, что истинна из них сотня, а девятьсот ложны. Порог значимости стандартный: 0,05. Из девятисот ложных гипотез случайно значимый результат дадут примерно сорок пять. Это те самые пять процентов, о которых написано в учебнике. А что с

истинными? Здесь всё решает мощность. Если опыт хорошо спланирован и мощность равна 80 процентам, то из ста истинных гипотез будут обнаружены восемьдесят. Всего значимых результатов 125, из них ложных 45. Каждый третий с лишним напечатанный «факт» неверен, хотя все правила соблюдены и никто не жульничал.

Но мощность в 80 процентов — это учебник. В психологии и нейронауках она часто оказывалась гораздо ниже. Если она равна 35 процентам, то истинных находок будет 35 против тех же 45 ложных, и большинство результатов окажутся ложными. Читателю первой главы всё это знакомо: число 0,05 есть вероятность данных при гипотезе, а не вероятность гипотезы при данных. Исходная доля истинных гипотез есть та самая априорная вероятность, которую статистические ритуалы велели игнорировать. Чем смелее и неожиданнее область, чем реже в ней верны проверяемые идеи, тем больше среди её открытий пустых, даже при безупречной честности. Самые сенсационные темы оказываются самыми ненадёжными. Но как раз они и публикуются в самых заметных журналах.

Иоаннидис сделал ещё один шаг, и в нём главное. До сих пор мы считали, что исследователи работают честно и точно по плану. Он добавил в расчёт смещение: долю анализов, которые не должны были бы дать результат, но в итоге дали его из-за особенностей планирования, обработки и подачи данных. Если таких случаев даже немного, картина резко ухудшается. А ещё прибавляется конкуренция: если десять лабораторий проверяют одно и то же и каждая публикует только свою удачу, то вероятность случайной удачи хотя бы у одной из них велика. Статья Иоаннидиса вызвала и возражения: некоторые статистики указывали, что его допущения взяты произвольно, а подсчёты по опубликованным p-значениям давали для медицинских журналов более скромные оценки доли ложных находок. Слова «можно доказать, что большинство результатов ложны» звучат сильнее, чем позволяет модель с настраиваемыми числами. Но главная логика никем не опровергнута: значимость сама по себе мало что говорит, если не знать, сколько попыток стояло за ней и какова была исходная вероятность удачи.

Есть и обратная ловушка, о которой ещё в 1957 году писал уже знакомый нам Деннис Линдли. При очень большой выборке результат может оказаться сколь угодно значимым и в то же время свидетельствовать в пользу отсутствия эффекта. Причина понятна после первой главы: значимость спрашивает лишь о том, редки ли такие данные при отсутствии эффекта, и не спрашивает, насколько они вероятны при его наличии. При огромной выборке ничтожное отклонение от нуля действительно редко, если эффекта нет, но ещё реже оно встречается, если эффект имеет сколько-нибудь заметную величину. В эпоху, когда компании и исследователи располагают данными о миллионах людей, значимым можно сделать почти любое различие, и это уже не математическая экзотика.

И здесь возникает вопрос, ответ на который занял следующие несколько лет. Что такое это самое смещение? Как оно работает в лаборатории, в которой честные люди делают свою работу? Могут ли они подтвердить что угодно, ни в чём не солгав? Ответ оказался неприятным и очень забавным.

Как получить любой результат

В том же 2011 году, когда вышла статья Бема, журнал Psychological Science напечатал работу американских исследователей Джозефа Симмонса, Лейфа Нельсона и Ури Симонсона. В ней сообщалось о поразительном открытии. Двадцать студентов Пенсильванского университета слушали либо песню группы «Битлз» «When I'm Sixty-Four», либо нейтральную инструментальную пьесу «Калимба», входившую в набор звуков одной из версий Windows. Потом участники указывали дату своего рождения и возраст отца. Анализ показал, что те, кто слушал песню о старости, стали почти на полтора года моложе, чем те, кто слушал «Калимбу»: 20,1 года против 21,5. Значение p составило 0,040.

Стоит вчитаться. Не почувствовали себя моложе. Стали моложе: по дате рождения в паспорте. Ни один студент не помолодел, разумеется, и авторы это прекрасно понимали. Опыт был проведён по всем правилам, данные настоящие, статистика правильная, ни одной строки подделано не было. Они просто сделали то, что считалось обычной практикой, и описали результат так, как его обычно описывают. А затем раскрыли, что именно было сделано на этом пути.

Они собрали сведения о многом: на сколько лет участник себя ощущает, сколько лет его отцу и матери, и ещё целый ряд посторонних вопросов. Была и третья песня, так что сравнивать можно было любую пару групп. Данные смотрели по мере поступления: проверяли после каждой порции в десять человек, остановившись, как только результат стал значимым. Возраст отца использовали как поправку, потому что с ней выходило лучше. А в статье сообщили только об удачном сочетании, как будто оно и планировалось с самого начала. Каждое такое решение по отдельности совершенно невинно и может быть даже разумным. Отчего бы не добрать участников, если результат неясен? Отчего бы не учесть возраст родителей, если он связан с возрастом детей? Отчего бы не исключить группу, в которой что-то пошло не так?

Вернёмся на минуту к Бему и посмотрим на его статью новыми глазами. Критики указали на то, чего читатель журнала сразу увидеть не мог. Чем больше участников было в опыте, тем меньше оказывался эффект, что естественно для набора данных до удачного момента. Голландский психолог Эрик-Ян Вагенмакерс и его соавторы, чей отклик вышел в том же журнале, показали, что многие проверки были в сущности разведочными, а представлены как подтверждающие. По словам других комментаторов, порядок опытов в статье не совпадал с порядком их проведения, а процедуру кое-где меняли по ходу дела. Бем считал всё это мелочами, ни одна из которых сама по себе не меняет вывода. Возможно, так оно и есть. Но в этом и состоит механизм: каждая отдельная вольность выглядит мелочью, а их произведение уже нет.

Авторы назвали это степенями свободы исследователя и посчитали, чего они стоят. Компьютерное моделирование на пятнадцати тысячах искусственных выборок, где никакого эффекта заведомо нет, дало следующее. При заранее заданном единственном анализе доля ложных удач равна положенным пяти процентам. Если из двух родственных показателей выбирать удачный, то это почти десять процентов. Если добавлять наблюдения после подглядывания, то около восьми. Если использовать одну поправку по выбору, то почти двенадцать. Если выбирать, какие группы сравнивать, то около двенадцати. А при совместном применении всех этих вольностей результат становится значимым в 61 проценте случаев. То есть вероятность «доказать» отсутствующий эффект оказывается выше, чем при бросании монеты. Заголовок статьи был прямым и по-своему отчаянным: «Ложноположительная психология». В первых её строках авторы написали, что опубликовать статистически значимое свидетельство в пользу любой гипотезы недопустимо легко.

Очень важно понять, чем это отличается от обмана. Подлинный обман есть и о нём мы ещё скажем, но он большая редкость. Здесь происходит нечто иное и гораздо более распространённое. Вот учёный, искренне верящий в свою гипотезу. Он провёл опыт, эффекта нет. Он думает: может быть, мало людей? Добирает. Может быть, опрос в конце был неудачным? Смотрит другой показатель. Может быть, те, кто плохо понял задание, всё смазали? Отбрасывает их. На каждом шаге он находит разумную причину, и он эту причину не выдумал. Так и работает обычное человеческое мышление: когда результат приятен, мы его принимаем, а когда неприятен, начинаем выяснять, в чём дело. В 2012 году психолог Лесли Джон и её соавторы анонимно опросили около двух тысяч американских психологов, и выяснилось, что заметная доля признавала за собой такие приёмы, причём многие не считали это нарушением. Это не кучка мошенников. Это нормальная лаборатория, какой она была.

Выразительнее всего это показал случай, когда сам участник рассказал обо всём как о добром примере. Брайан Вансинк, профессор Корнеллского университета, был одним из самых популярных исследователей питания в мире: его работы о том, что люди съедают больше из

большой тарелки и не замечают дна бездонной миски супа, разошлись по журналам, книгам и рекомендациям для школьных столовых. В ноябре 2016 года он опубликовал в своём блоге поучение молодым учёным под названием «Аспирантка, которая никогда не говорила нет». Там рассказывалось, как он дал новой сотруднице данные неудавшегося опыта: цена в итальянском ресторане со шведским столом не повлияла на количество съеденного. Он сказал ей, что данные дорого обошлись и что-то в них должно найтись. Девушка согласилась, в отличие от другого сотрудника, отказавшегося от этого занятия, и вскоре вышло несколько статей: сравнивались мужчины и женщины, обедающие и ужинающие, сидящие в одиночку и в компании. Автор блога гордился её трудолюбием. Читатели пришли в ужас. Другие исследователи заново просмотрели его работы и обнаружили множество несостыковок. К сентябрю 2018 года было отозвано тринадцать его статей, университет признал нарушение академической этики, и Вансинк подал в отставку, продолжая настаивать на ценности своей работы. Важнее всего здесь то, что написать такое в открытом блоге мог только человек, вообще не видевший в этом ничего дурного. И десятилетиями в его дисциплине этого действительно почти никто не видел.

Американские статистики Эндрю Гельман и Эрик Локен продвинулись ещё дальше. Они указали, что нет нужды даже в сознательном переборе. Исследователь мог вообще провести только один анализ. Но какой именно, определилось уже после того, как он увидел данные. Если бы данные были другими, он бы логично выбрал другой анализ, и у него нашлась бы для этого столь же убедительная причина. Гельман и Локен назвали это садом расходящихся тропок, позаимствовав образ у Борхеса. Человек проходит по саду только одной тропинкой и искренне уверен, что выбора не было. А расчёт вероятности предполагает, что маршрут был задан заранее, вне зависимости от того, что встретится по пути. Отсюда важное следствие: честность сама по себе от этого не лечит. Ни добрые намерения, ни отсутствие подгонки в обычном смысле не помогают, если выбор анализа зависел от данных. Ведь выше мы уже видели задачу о трёх дверях: одна и та же открытая дверь весит по-разному в зависимости от того, каким правилом она открыта. Здесь то же самое, только двери открывает сам учёный и не замечает этого.

Как широк этот сад, удалось измерить прямо. В 2018 году группа под руководством Рафаэля Зильберцана раздала один и тот же набор данных 29 независимым командам, всего 61 специалисту. Вопрос был один: чаще ли футбольные судьи показывают красную карточку темнокожим игрокам? Никто ни с кем не соревновался за публикацию, все работали открыто и добросовестно. Двадцать команд нашли значимый эффект, девять не нашли. Оценки разошлись от почти полного отсутствия различий до почти трёхкратного, а наборы учитываемых поправок образовали 21 разное сочетание. Ни убеждения аналитиков, ни их опыт, ни оценки качества работы, выставленные коллегами, этого разброса не объясняли. Вот что такое одна опубликованная статья с одним анализом: одна тропинка из большого сада, и читателю обычно показывают только её. Заметьте также, что данные здесь никто не подгонял. Просто честные решения честных людей ведут в разные стороны, а статистика сообщает только о том, что случилось на выбранной дороге.

К этому же относится ещё один приём, которому американский психолог Норберт Керр в 1998 году дал название, ставшее термином: выдвижение гипотез после того, как результат известен. По-английски это звучит как HARKing. Работает так. Лаборатория проверяла одно, получила другое, неожиданное. Интересно! И статья пишется так, будто неожиданное было предсказано с самого начала: во введении излагается теория, которая ведёт именно к этому, и читатель видит безупречную смелую гипотезу, блестяще подтверждённую. Именно против этого предостерегали все теории подтверждения, какие мы знаем. Поппер требовал, чтобы проверка могла опровергнуть теорию, а если теорию придумали после результата, никакого риска не было. В байесовских терминах предсказание задним числом ничего не поднимает: вероятность такого результата при подогнанной гипотезе равна единице потому, что её под него подогнали. Есть даже советы для начинающих психологов, написанные уважаемым авто-

ром, где прямо рекомендуется писать статью о том, что вышло, а не о том, что планировали. Заметим, кстати, что автором одного из самых популярных таких советов был сам Дэрил Бем.

До сих пор мы говорили о том, что происходит в лаборатории. Но есть вторая половина картины, та, что происходит на пути в журнал. В 1979 году американский психолог Роберт Розенталь дал ей имя, которое прижилось сразу же: проблема ящика стола. Представим себе, что двадцать лабораторий независимо проверяют один и тот же несуществующий эффект. В среднем одна из них получит значимый результат. Девятнадцать положат свои неудачи в стол, поскольку журнал такое не примет, а заняться чем-нибудь другим полезнее. Одна опубликует. Читатель журнала видит единственную статью, в которой есть значимый эффект, и совершенно правильная статистика. Откуда ему знать о девятнадцати?

Масштаб этого явления теперь можно оценить. Есть особый формат публикации, о котором мы ещё скажем: журнал принимает статью к печати на основании плана опыта, до получения результата. Это единственный случай, когда отбор по результату невозможен. В 2021 году группа психологов сравнила обычные статьи с такими. Результаты, поддерживающие проверяемую гипотезу, составили в обычных статьях 96 процентов. В статьях с предварительным принятием 44 процента. Даже с учётом того, что подобные работы могут отличаться по темам, разрыв огромен. Литература, которую читают студенты и на которую ссылаются популяризаторы, в которой почти каждый проверяемый замысел удаётся, описывает не науку, а её выставочный зал. Психолог, заглянувший в собственный архив, сразу понимал это, но считал, что неудача случается только у него.

Отбор по значимости порождает и следствие, которое называют проклятием победителя. Если опыт мал, значимым может оказаться только большой эффект, поскольку маленький утонет в шуме. Значит, небольшая выборка не столько находит настоящие большие эффекты, сколько пропускает в печать случайные завышения. Настоящий эффект может быть, но всего вполовину или вчетверо меньше заявленного. Как раз это и получилось в работе Носека: при повторении эффекты оказались вдвое меньше. Тут есть почти комичный парадокс. Чем хуже опыт, тем более поразительный результат он должен дать, чтобы его напечатали. Статьи, прочитанные за завтраком под заголовком «оказывается, мы совсем не те, кем себя считаем», особенно часто бывают именно такого происхождения.

Наконец, обратимся к уже знакомому Иудее Перлу. Он смотрит на всю эту историю с другой стороны, и его упрёк серьёзнее. Для него ошибка не только в том, что результаты отбирают. Ошибка ещё и в том, что считают. Статистики, пишет он, настолько запутались в вопросе о том, какие переменные надо учитывать, что по умолчанию стали учитывать всё, что удалось измерить. Но поправка на лишнюю переменную может не очистить, а исказить результат. Если учесть то, что само является следствием изучаемой причины, или то, что случайным образом связывает две другие величины, можно создать связь там, где её нет. Коэффициент регрессии, сколько его ни уточняй, остаётся описанием статистического тренда и причин сам по себе не показывает. Он приводит поразительную цифру: метод анализа посредствующих переменных, предложенный Бароном и Кенни в 1986 году, к 2017 году собрал 73 тысячи цитирований, больше иных трудов Эйнштейна и Фрейда, хотя в нём, по оценке Перла, вообще нет причинной логики. Тысячи исследователей пользовались им как рецептом. Об этом подробно в следующей главе.

Но один вывод нужно сделать уже здесь. Точная процедура, которую можно применять механически, удобна. Перл саркастически заметил, что учёные всегда предпочитают привычные вычисления методам, которые затрагивают их научное знание. В этом смысле ритуал значимости был не просто ошибкой. Он был договорённостью о том, как не спорить. Если результат прошёл ворота, вопрос закрыт, и можно обойтись без размышлений о том, что именно измерено, сколько было попыток и нет ли другого объяснения. А всякий раз, когда из проверки делают обряд, в ней появляется возможность пройти обряд, не проходя проверки.

Масштаб и причины

Можно было бы подумать, что всё это особенность психологии, науки молодой и зыбкой, где предмет изучения капризен, а эффекты малы. Многие физики и молекулярные биологи так и думали. Однако в 2011 и 2012 годах появились два сообщения совсем из другой области, причём от людей, которым было важно не разоблачение, а деньги.

Фармацевтическая компания, прежде чем вкладывать сотни миллионов в новое лекарство, обычно проверяет у себя лабораторные результаты, на которых оно основано. Группа исследователей немецкой фирмы Байер сообщила, что из 67 внутренних проектов полностью подтвердить опубликованные данные удалось примерно в четверти. Годом позже Гленн Бегли, работавший в американской компании Амджен, и Ли Эллис сообщили ещё более резкую цифру. Сотрудники Амджен попытались воспроизвести 53 работы, которые считались важнейшими в доклинических исследованиях рака, и подтвердили шесть. Это одиннадцать процентов. Оговоримся: у этих двух сообщений есть свои слабые места. Компании не раскрыли, какие именно работы проверялись, так что сами их выводы нельзя проверить, и в этом есть горькая шутка. Отобраны были особо заметные и новаторские работы, а неудача в чужой лаборатории не всегда означает, что результат неверен. Но вывод всё равно очевиден: дело не в одной психологии, и даже без соблазнов общественных наук есть целые области, где напечатанное ненадёжно.

Другие большие проекты дали результаты получше, но не утешительные. В экономике при проверке лабораторных опытов подтвердилось около 60 процентов. В 2018 году группа под руководством американского экономиста Колина Камерера повторила 21 опыт по общественным наукам из самых престижных журналов мира, Nature и Science, используя выборки примерно в пять раз больше исходных. Подтвердились 13 из 21, и величина эффекта составила примерно половину заявленной. У этой работы есть любопытная деталь. Перед повторением участники исследовательского сообщества делали ставки на особом рынке, какие результаты подтвердятся, и в целом угадывали хорошо. Это значит, что специалисты в массе своей могут отличить надёжную работу от сомнительной, просто прочитав её. Тогда почему журналы всё это печатали? И почему коллеги всё это цитировали? Вопрос сам собой переходит с методов на людей.

Цифры становятся нагляднее, если посмотреть, что именно пало. В 1996 году американский психолог Джон Барг сообщил об опыте, вошедшем затем во все учебники. Студенты составляли предложения из наборов слов, среди которых были слова, связанные со старостью, и после этого шли по коридору медленнее, чем контрольная группа. Работу цитировали тысячи раз, а популярные книги пересказывали её как доказательство того, что поведением человека незаметно управляет любая мелочь. В 2012 году бельгийская группа Стефана Дуайена повторила опыт, заменив секундомер в руках экспериментатора автоматическим замером, и эффекта не нашла. Зато он появился, когда самим экспериментаторам внушали, что участники должны идти медленнее. Сразу вспоминается лошадь по имени Умный Ганс, считавшую по незаметным знакам своего хозяина. Барг ответил резко, разгорелся публичный спор. Другой случай: так называемое истощение эго, представление, что сила воли есть нечто вроде мышцы, которая устаёт от работы. На эту идею опиралось множество исследований и популярные книги о самоконтроле. В 2016 году 23 лаборатории повторили опыт по единому заранее утверждённому плану на 2141 участнике, и эффект оказался практически нулевым, хотя сторонники идеи спорят о том, удачной ли была выбранная задача. И, наконец, поза силы: работа 2010 года на 42 участниках утверждала, что две минуты в уверенной позе меняют уровень гормонов и готовность к риску. Об этом с успехом рассказывали на всемирно известных лекциях и в советах соискателям. Более крупные опыты гормональный эффект не подтвердили, а в 2016 году

первый автор исходной статьи, Дана Карни, публично заявила, что больше не считает эффект реальным. Такое признание само по себе редкость, и за него ей стоит отдать должное.

Чтобы не впасть в обратную крайность, надо добавить, что пало не всё. В работе Носека результаты когнитивной психологии подтверждались примерно вдвое чаще, чем социальной: около половины против примерно четверти. То, что слово «врач» ускоряет узнавание слова «медсестра», воспроизводится легко, как и многие другие результаты о памяти и внимании. Устойчиво ведут себя и сильные классические эффекты вроде того, что человеку трудно назвать цвет чернил, которыми написано название другого цвета. Даниэль Канеман, чья популярная книга сильно опиралась на результаты о незаметных влияниях, ещё в 2012 году написал коллегам открытое письмо с предупреждением о надвигающейся катастрофе, а позже прямо признал, что слишком доверял маленьким исследованиям. Различие идёт не по границе дисциплин, а по устройству работы. Там, где эффект велик, процедура стандартна, а один и тот же человек может повторить опыт много раз, проверка срабатывает сама собой. Там, где эффект тонок, а данные дороги и единичны, начинается наша история.

Отвечать на него проще всего так: виноваты отдельные жулики. Такие действительно есть. В 2011 году в Нидерландах выяснилось, что видный социальный психолог Дидерик Стапел годами попросту выдумывал данные: он садился за компьютер и вводил те числа, которые должны были получиться, а потом передавал готовые таблицы аспирантам. Отозвано более пятидесяти его работ. Но и такие случаи, сколько можно судить, встречаются редко, и анонимные опросы показывают, что в явной подделке признаётся около двух процентов учёных. Для объяснения того, о чём мы говорим, никакой подделки не нужно. Дело Стапела важно другим: годами коллеги, рецензенты и журналы ничего не замечали, потому что его результаты всем нравились. Раскрыли его в итоге младшие сотрудники. Проверка этого человека не коснулась, пока рядом не оказалось людей, готовых задать вопрос.

Гораздо интереснее другое объяснение. В 2016 году американский специалист по когнитивным наукам Пол Смалдино и антрополог Ричард Макэлрит опубликовали статью с заглавием в духе Дарвина: «Естественный отбор плохой науки». Их рассуждение строится как модель из теории эволюции. Представим популяцию лабораторий. У каждой есть свой способ работы, более тщательный или менее. Тщательность стоит времени, поэтому тщательные лаборатории публикуют меньше, зато у менее тщательных больше удач, в том числе ложных. Карьера, гранты и новые места зависят прежде всего от числа публикаций в заметных журналах. Выпускники успешных лабораторий заводят собственные и учат студентов так, как учили их самих. Что будет через несколько поколений? В модели менее тщательные методы постепенно вытесняют более тщательные, причём никто никого не обманывает и ни один человек сознательно не выбирает плохую работу. Всё, что требуется, это чтобы награды полагались за количество и новизну. Авторы прибавляют и данные: проанализированные ими обзоры показывают, что средняя мощность исследований в поведенческих науках за шестьдесят лет почти не выросла, хотя статистики всё это время призывали её повышать.

Обратите внимание на то, как это дополняет картину третьего тома. Роберт Мертон описывал нормы науки: всеобщность, бескорыстие, организованный скептицизм. Он считал, что система наград поддерживает эти нормы, поскольку первооткрывателя чтут, а проверяющего в итоге тоже уважают. Однако если награды выдаются за первенство и никогда за повторение, то организованный скептицизм остаётся словами на плакате. Эта система хорошо работает там, где результат сам быстро подвергается проверке, например когда из него делают прибор. Там, где подлинной проверки нет и никто не ставит опыта второй раз, она производит то, за что платит.

А вот с деньгами дело обстоит ещё проще. В клинических исследованиях давно заметили, что работы, оплаченные производителем лекарства, заметно чаще приходят к выгодным для него выводам, чем независимые. Часто это происходит без явной подделки, так как достаточно

выбора того, с чем сравнивать, какой показатель считать главным и что из результатов печатать. По той же причине там, где регистрация испытаний стала обязательной и исказить выводы стало труднее, доля положительных результатов заметно снизилась. Перл в своей книге приводит случай табачной индустрии, которая, как мы увидим в тринадцатой главе, поставила производство сомнений на промышленную основу. Там всё было сознательно. Но механизм рынка действует и без сознательного намерения: приборы, реактивы, чудодейственные тренинги и лекарства приносят деньги, а отрицательный результат их не приносит никогда и никому.

Вот один пример, где можно точно измерить последствия, потому что в этой области существует надёжный контрольный архив. Американский психиатр Эрик Тёрнер, работавший ранее рецензентом в Управлении по надзору за продуктами и лекарствами США, сравнил в 2008 году документы, которые производители обязаны подавать в это управление, с тем, что напечатано в журналах. Он взял все зарегистрированные испытания двенадцати антидепрессантов, в которых участвовало более двенадцати тысяч человек. Судя по журнальным статьям, удачными были 94 процента испытаний. По документам управления удачными были 51 процент. Почти все положительные испытания вышли в печать, а отрицательные либо не были опубликованы, либо поданы так, что выглядели удачными. Средняя величина эффекта в опубликованном виде оказалась примерно на треть выше настоящей. Заметим, что вывод здесь не в том, что эти лекарства бесполезны: обычно они всё же превосходят плацебо. Вывод в другом: врач, добросовестно прочитавший всю литературу и следующий ей, располагал искажённой картиной. И он никак не мог этого знать.

Что всё это значит для вопроса, с которого начался том? Обычно, когда говорят о силе, вытесняющей проверку, подразумевают внешнюю власть: начальство, запрещающее неудобное. Пример Бема показывает другой её вид, гораздо менее заметный. Журнал, рецензенты, авторитетные коллеги, комиссия по присуждению грантов — всё это и есть то, что принято называть научным сообществом. В сообщество обычно предлагают верить. Если мнение сообщества считать последней инстанцией, то выводы о том, как поведение человека незаметно меняется от случайно прочитанного слова, следовало принимать и в 2005 году, поскольку они были в учебниках и на них ссылались тысячи работ. Не сообщество установило их ненадёжность. Её установили конкретные люди, которые взяли и проверили, зачастую вопреки сообществу, рискуя репутацией и получая упрёки в недоброжелательности и в «методологическом терроризме». Это выражение тоже прозвучало. Голос сообщества здесь ничего не проверяет: это только свод того, что уже прошло через ворота, которые могут быть устроены плохо. В любой стране и при любом строе у таких ворот есть свои хозяева, у которых есть свои интересы. Мы вернёмся к этому в двенадцатой главе. Но пока важно отметить, что наш вывод касается всей картины: на чьей бы стороне ни находилось большинство голосов, вопрос решается не их количеством.

Здесь полезно вспомнить и о другом. Заблуждения, о которых идёт речь, были не только в головах ведущих учёных. Они были встроены в программу подготовки: во всех странах студентов обучали статистике как обряду. Это обстоятельство может объяснить, почему реформа так медленно идёт и почему многие из тех, кто мог бы её поддержать, встречают её с раздражением: она объявляет ошибочным то, чему они сами учили, и то, на чём выросла их карьера.

Возражения и что осталось

Есть ли вообще кризис? Итальянский исследователь науки Даниэле Фанелли, много лет изучавший смещения в научной литературе, в 2018 году выступил в журнале Национальной академии наук США с неожиданным заявлением. Рассказ о кризисе, считает он, сам по себе преувеличен. Ненадёжность распределена очень неравномерно: она высока в одних областях и у одних типов работ, и невелика в других. Нет надёжных свидетельств того, что поло-

жение ухудшается со временем. А подделки редки, и связь смещений с давлением публикаций не столь проста. Рост числа отзывов статей говорит скорее об улучшении контроля. И спрашивает, нужна ли вообще риторика кризиса: что если учёные, повторяющие на каждом углу, что наука сломана, лишь снабжают аргументами тех, кому она мешает? Мы видели, что Гилберт оспаривал и саму статистику Носека. Дополнительно указывают, что повторение никогда не бывает точным. Опыт на студентах в Вирджинии 2013 года иной, чем на студентах в другом месте и в другое время, и в психологии такая разница может значить многое.

Эти возражения следует принять, но принять в точных границах. Единой цифры ненадёжности для всей науки не существует. Слова «вся наука сломана» ложны, и это очень важно для следующих глав. Спектроскопию, расчёт орбит, картирование геномов и большую часть инженерных основ они вообще не затрагивают. Ссылка же на тонкие различия условий имеет свою цену: если эффект исчезает при незначительной перемене места, времени и людей, то объявленный общий закон поведения человека надо по меньшей мере сильно сузить. Если в статье написано «люди склонны к тому-то», а не «определённые студенты в определённом кампусе в определённом году», то автор сам взял на себя обязательство, что его вывод переносим.

Тут особенно полезно сравнение с физикой, потому что её часто приводят как образец. У физиков принят гораздо более строгий порог для объявления открытия: так называемые пять сигм, что соответствует вероятности примерно один на три с половиной миллиона. Именно такой уровень был достигнут при объявлении бозона Хиггса в 2012 году. Но и у них всё не так уж просто. В сентябре 2011 года коллаборация OPERA сообщила, что нейтрино прилетают из Женевы в итальянскую лабораторию на шестьдесят наносекунд раньше, чем позволяет скорость света. Статистическая значимость была выше шести сигм. Авторы не утверждали открытия и прямо просили проверить их работу. Причина обнаружилась через несколько месяцев: плохо подключённый оптоволоконный кабель и неточный тактовый генератор. Значимость в сколько угодно сигм, как видим, не спасает от систематической ошибки, и вот о чём предупреждал Фишер, говоря о повторении смещённого опыта. Разница с психологией не в статистике, а в другом: физики сразу знали, что результат противоречит огромному массиву других данных, что считать его верным стоит только при исключении всего остального, и могли проверить это на других приборах. Априорная вероятность была низкой, и её никто не смел не учитывать.

И у физиков бывало самообольщение коллективного масштаба. В 1903 году французский физик Рене Блондло, серьёзный учёный, открыл новый вид излучения, названный им N-лучами. За пару лет появились сотни статей десятков авторов, которые тоже их наблюдали, а регистрировались лучи по слабому изменению яркости, то есть в значительной мере на глаз. Всё кончилось, когда американский физик Роберт Вуд, приехавший в лабораторию, незаметно вынул в темноте из прибора призму, без которой ничего не должно было работать, и наблюдатели продолжали видеть то же самое. Так что дело не в том, что психологи глупее физиков. Дело в том, что для физиков такая проверка оказалась сравнительно дешёвой и что призму можно вынуть, а поведение людей настолько зависит от обстоятельств, что вынуть там почти нечего.

Что считать повторением. Поппер ещё в «Логике научного открытия» писал, что невоспроизводимые единичные события не имеют для науки значения. Звучит как трюизм, но на деле из этого не следует, что понятно, чем одно событие отличается от другого. Психологи различают прямое повторение, при котором как можно точнее воспроизводят прежнюю процедуру, и концептуальное, при котором проверяют ту же идею иным способом. Концептуальные повторения считались более почётными: они показывают широту идеи. Но у них есть коварное свойство. Если такой опыт удался, он считается подтверждением. Если не удался, можно сказать, что новая процедура просто не подходит, и никто не сочтёт это опровержением. Значит, при такой практике концептуальное повторение может только прибавить, но никогда не убавить. Эта асимметрия означает, что целое направление могло строить сотни взаимно

подтверждающих работ и ни разу не поставить свой исходный результат под настоящую угрозу. Вот почему прямые повторения, казавшиеся скучными, оказались столь важны: не потому, что они заменяют разнообразие свидетельств, о котором мы говорили в первой главе, а потому, что только они способны дать неудачу, которую нельзя списать на различие условий.

Регресс экспериментатора. Есть возражение гораздо более глубокое, и оно уходит в философию. В 1970-е годы британский социолог науки Гарри Коллинз изучал попытки обнаружить гравитационные волны. Американский физик Джозеф Вебер заявил, что поймал их на своих массивных цилиндрических антеннах. Другие построили подобные приборы и не увидели ничего. Кто прав? Вебер отвечал: ваши приборы плохо настроены. Как проверить, хорошо ли работает прибор? Он работает хорошо, если показывает правильный результат. Но какой результат правилен, мы и пытаемся выяснить при помощи прибора. Коллинз назвал этот круг регрессом экспериментатора (1985). Значит, повторение не может быть окончательным судьёй, ведь само решение, считать ли опыт грамотным повторением, зависит от того, что мы думаем о результате. По Коллинзу, круг разрывается не логикой, а переговорами внутри сообщества: кто-то оказывается убедительнее, у кого-то больше веса.

Сегодня это возражение звучит очень знакомо. Каждый автор неудавшегося опыта может сказать: вы неправильно повторили, вы не знали всех тонкостей, у вас неподходящие участники. И часто он прав, ведь наука состоит не только из протоколов, но и из умения. Читатель сразу узнает здесь проблему Дюгема в рабочем халате: неудача проверяет не гипотезу, а всю систему вместе с мастерством лаборанта. Если остановиться на Коллинзе, то вывод окажется неприятным: судьба результата решается тем, кто сильнее. Именно на это обычно ссылаются те, кому дорога мысль, будто наука есть разновидность политики.

Однако решающий ответ дал не философ. Американский физик и историк науки Аллан Франклин подробно разобрал то же самое дело Вебера и показал, что исход спора определялся не только авторитетом. Были доводы: калибровка приборов известными сигналами, найденная ошибка в программе у одной из групп, расхождение обнаруженных сигналов с тем, что требует теория для такого источника, независимость разных приборов. Именно совокупность этих доводов перевесила, и судить об их весе можно было, не зная заранее ответа. Регресс разрывается не потому, что кто-то сильнее, а потому, что есть другие способы проверить прибор, помимо спорного результата. В 2015 году установка LIGO действительно зарегистрировала гравитационные волны, причём совсем не такие сильные, какие сообщал Вебер, и по той причине, что на этот раз независимые основания для доверия были гораздо богаче. Этот случай нам ещё пригодится в главе о Хакинге. Сходимость независимых способов и есть то, что разрывает круг. Так что Коллинз прав, описывая реальную трудность, и неправ, делая из неё приговор. Тот же Коллинз, заметим, был очень далёк от отрицания физики и позже написал целую книгу с большой симпатией к физикам, искавшим эти волны.

Лекарства. Самое простое лечение следует прямо из диагноза. Если беда в том, что анализ выбирается после данных, надо выбирать его до данных и оставлять запись. Это называется предрегистрацией: исследователь заранее и публично фиксирует гипотезу, план, число участников и способ обработки. Следующий шаг сделал британский нейробиолог Кристофер Чемберс, предложивший с 2013 года в журнале Cortex формат зарегистрированных отчётов, о котором мы уже упоминали: журнал рецензирует план и обещает печатать результат, каким бы он ни оказался. Заодно обязательной становится публикация данных и программ, а журналы начинают принимать повторения. Именно формат зарегистрированных отчётов дал те самые 44 процента вместо 96. С точки зрения Поппера это просто запоздалое выполнение его требования: заранее указать, что опровергло бы гипотезу. Медики пришли к этому раньше психологов: клинические испытания обязаны регистрироваться с начала XXI века.

Но и здесь нельзя рассчитывать на механику. Мы уже говорили, что всякий обряд можно пройти, не проходя проверки. Заявленный план может быть туманным, отступления от него

могут не быть оговорены. Исследовательская работа, которая ещё не знает, что ищет, вообще плохо укладывается в предварительный план, и нельзя её запрещать: Коперник и Дарвин ничего не регистрировали. Нужно только честно называть её разведкой и не выдавать за проверку. Самый наглядный урок дала сама реформа. В 2023 году в журнале Nature Human Behaviour вышла статья, подписанная в том числе Носеком и другими видными фигурами движения. Сообщалось, что если соблюдать все новые правила, то повторяемость достигает 86 процентов. Это была радостная новость для реформаторов. Через год журнал статью отозвал: критики и редакторы обнаружили, что главный анализ не был предрегистрирован, хотя статья утверждала обратное, а также нашли другие изъяны. Авторы с отзывом согласились в части неверных утверждений о предрегистрации, с остальными обвинениями нет. Некоторые противники реформы торжествовали. На наш взгляд, ничего более убедительного в её пользу придумать было нельзя: критерий применён к собственным сторонникам, и ошибка найдена именно потому, что документы лежали в открытом доступе.

Спор продолжается о самих числах. В 2018 году большая группа статистиков и учёных предложила понизить порог значимости для новых открытий с 0,05 до 0,005. Другие возразили, что этот порог столь же произволен, что он увеличит трудности с небольшими выборками и что выбор порога надо обосновывать в каждом случае. Третьи предлагают вообще отказаться от слов «статистически значимо», и в 2019 году эту мысль поддержал специальный выпуск журнала Американской статистической ассоциации. Байесианцы предлагают заменить p-значения факторами Байеса. В этом споре мы видим то самое столкновение, которым закончилась первая глава. Когда Вагенмакерс и его коллеги пересчитали данные Бема по-байесовски, они получили, что данные его гипотезу почти не поддерживают. На это Бем с соавторами ответили, что априорные распределения были выбраны неудачно, и предложили свои, при которых результаты оказываются намного лучше. Формула не положила конец спору, а только показала, в чём именно он состоит: насколько вероятен эффект заранее и какой он должен быть величины.

А Дебора Мейо, чья теория строгой проверки уже появлялась в первой главе, видит в этой истории подтверждение своей правоты. Для неё подглядывание в данные и выбор анализа — главное, что следует учитывать, и как раз для этого нужны частоты ошибок процедуры. Значимость в 0,05, полученная перебором, означает нестрогую проверку, и гипотеза ничего не прошла. Байесовский расчёт, не учитывающий правила остановки, по её мнению, не замечает самого существенного. Байесианцы вроде Вагенмакерса отвечают, что и у них есть средства борьбы с подгонкой, а безусловная вера в предрегистрацию ничего не заменяет. Для нас здесь важно не вынесение окончательного приговора, а то, что обе стороны теперь согласны в главном: процесс получения данных принадлежит к данным.

Так что же с ясновидением? Посмотрим, чем кончилось дело, потому что здесь и лежит настоящий ответ на вопрос о границе. Бем не сдался. В 2015 и 2016 годах он с соавторами опубликовал метаанализ девяноста опытов из 33 лабораторий четырнадцати стран и заключил, что в целом эффект подтверждается. Сторонники ясновидения указывают на эту работу и поныне. Но сводить в метаанализ то, что прошло через ящик стола, значит складывать смещения. Тогда в 2023 году вышел результат проекта, который назывался «Прозрачное пси». Венгерский психолог Золтан Кекеш с коллегами взял первый опыт Бема с занавесками и организовал его проверку так, как никто никогда прежде не проверял. План был согласован комиссией из 29 специалистов, в которой пятнадцать были сторонниками Бема, включая его самого, а четырнадцать скептиками, данные поступали в открытый доступ по мере получения, чтобы никто не мог ничего изменить. В десяти лабораториях девяти стран были собраны 37 836 попыток от 2115 участников. Угадывали в 49,89 процента случаев при случайном уровне 50. У Бема было 53,07.

Это и есть образец того, как кончается подобное дело, когда всё сделано честно: стороны заранее согласились, что считать результатом, а потом посмотрели. Как заметил Кристофер

Френч, мировая пресса, обрадованная в 2011 году сенсацией, на этот раз промолчала. Разумеется, и после этого можно предположить, что эффект есть, но он в сотни раз меньше, или что он скрывается, когда за ним наблюдают скептики. Исследователи, изучающие парапсихологов, давно отметили этот приём: когда явление не воспроизводится, говорят, что оно по своей природе тонкое, капризное и неуловимое. Эрих Гуд в сборнике Пильюччи и Будри описывает эту позицию как типичную для всей области. И вот здесь проходит граница, которая нас волнует.

У этого образца есть название. Канеман предложил говорить о состязательном сотрудничестве: противники заранее договариваются, какой опыт и при каких результатах будет считаться решающим, а затем вместе его проводят. Для вопроса о том, кому верить, это особенно важно. Здесь решает не мнение большинства, не авторитет журнала и не голосование специалистов. Сообщество участвует в деле не как судья, а как сторона, которая может согласиться с условиями проверки и этим лишить себя права потом объявить их неправильными. Такая процедура не упраздняет разногласий и не устраняет всех вопросов, но она показывает, как сообщество может работать на проверку, а не вместо неё.

Обе дисциплины столкнулись с невоспроизводимостью. Психология, пусть и с воем, с взаимными обвинениями и разрушенными карьерами, изменила свои правила: ввела предрегистрацию, начала печатать неудачи, признала, что многие учебные результаты ошибочны. Парапсихология такого не сделала, и не сделала потому, что в ней нет результатов, которые могли бы устоять после изменения правил. Заметим: Бем соблюдал те же нормы, что и другие психологи. Методологической границы между ними в 2011 году не было. Граница появилась потом, и проведена она не свойствами метода, а реакцией на неудачу. Мы увидим в последней главе, что современные философы лженауки считают это одним из важнейших её признаков.

Лакатос назвал бы то, что происходит в парапсихологии, вырождающейся программой. Каждая неудача сопровождается новым вспомогательным допущением, которое ничего нового не предсказывает. Эффект есть у верящих и отсутствует у неверящих: это предложили называть эффектом овец и козлов. Эффект исчезает со временем: это назвали эффектом спада. Эффект пропадает в присутствии скептика. Заметим, что подобные допущения сами по себе не абсурдны. Вера участника могла бы, например, влиять на результат. Однако они вводятся задним числом, а не из новых предсказаний, и в итоге защищают исходную гипотезу от любых возможных данных. По той же логике надо оценивать и некоторые ответы психологов на неудавшиеся повторения: скрытые посредники, особенности культуры, неподходящие участники. Такие ответы допустимы и бывают верны, если из них следует новое проверяемое предсказание. Если не следует, они ничем не лучше отговорки медиума, что сегодня духи не в настроении.

Остаётся опасность с другой стороны, о которой говорил Фанелли. Как только учёные сами признают такие вещи, ими начинают пользоваться те, кому неудобны их выводы. Американские табачные компании ещё в 1990-е годы создавали организации в защиту «здравой науки», требуя всё более полных доказательств там, где доказательства были для них невыгодны. В любой стране есть влиятельные люди, которым кстати сообщение, что значительная часть исследований не повторяется, если речь идёт о каком-нибудь неудобном им отчёте. Но заметим, как отвечает на это хорошее понимание предмета. Сказать «наука ненадёжна» значит не сказать ничего. Надо спрашивать: что именно было проверено, сколько раз, кем, независимо ли, сходятся ли разные способы? Результат о том, что на вероятность рака влияет курение, опирается на сотни независимых работ разного типа, и даже всеобщий кризис ему не страшен. А свежая работа на пятидесяти студентах о том, что определённый цвет обоев делает вас щедрее, по-прежнему выглядит хрупко. Кризис воспроизводимости — не довод против науки. Он есть инструкция, как читать её в разных случаях по-разному.

Итог и вопрос читателю

Первая глава учила нас менять уверенность по правилам. Эта показала, что правила ничего не стоят, если неизвестно, как данные попали нам в руки. Значимое число в статье есть конец длинной цепочки решений, а большая её часть остаётся за кадром: сколько групп сравнивали, когда перестали набирать участников, какие показатели выбрали, что положили в стол, за что платили. Проверка опытом не отменена. Она оказалась работой, которую надо делать на всех этих звеньях, а не только на последнем.

Отсюда несколько выводов, приложимых и к чтению газет. Единственное исследование почти ничего не решает, особенно если оно маленькое, а вывод поразительный. Результат, повторённый независимыми людьми другими способами, стоит несравненно больше. Важен не тот вопрос, значим ли результат, а другие: сколько было попыток, был ли план заранее, кто мог выиграть и нашёлся ли кто-нибудь, кто пытался его опровергнуть.

Для философии науки главный итог неудобен, и его придётся удержать. Поппер, Лакатос и Байес писали о том, как должен рассуждать учёный. Но то, что делает наука в целом, зависит не только от рассуждений. Журнальные правила и система наград могут сделать негодным даже верный метод в руках честных людей, и разумный отдельный человек ничего с этим не поделает. Это значит, что вопрос о рациональности науки в значительной части есть вопрос о том, как устроено её сообщество, и что ссылка на него сама по себе ничего не гарантирует. Мы подробно обсудим это в двенадцатой главе.

А пока отметим последнее обстоятельство. Перл настаивает, что вся эта история с ритуалами возникла из более старой ошибки: статистика запретила себе говорить о причинах. Исследователь думает «вызвано», а пишет «связано с», и никакая предрегистрация этого сама не исправит. Безупречно зарегистрированное и точно повторённое исследование может с большой надёжностью устанавливать связь, ничего не говорящую о том, что произойдёт, если вмешаться. Об этом следующая глава.

Вопрос читателю. Вспомните диету, способ воспитания, вложение денег или народное средство, которые «сработали» у вас или у знакомых. Сколько других способов пробовали и забыли? Сколько людей попробовали то же самое и промолчали, потому что вышло плохо? У каждого человека есть свой ящик стола, и самые твёрдые житейские убеждения нередко опираются лишь на то, что из него случайно выпало.



Глава 3. Причина возвращается

«Что» и «почему»

В 1990-е годы американский специалист по информатике Иудея Перл, который вырос в Израиле и в школьные годы штудировал Книгу Бытия вместе с классом, заново прочёл сцену в Эдемском саду и заметил в ней деталь, которой прежде не придавал значения. Бог задаёт Адаму простой вопрос о факте: не ел ли ты от дерева, с которого я запретил тебе есть? Ответ на такой вопрос состоит из одного слова. Но Адам говорит: жена, которую ты мне дал, дала мне плод. Бог спрашивает Еву: что ты сделала? Ева отвечает: змей обольстил меня. Спрашивали «что», а ответили «почему». У обоих есть объяснение, и оба, что характерно, указывают на другого. Первые люди в этой истории только что съели плод познания и немедленно начали с причинного анализа.

Перл извлёк из этой сцены вывод, ради которого написана вся его книга, написанная с Даной Маккензи (2018; по-русски она вышла под заглавием «Думай „почему?“»). Мир, каким его знают люди, состоит не из сухих фактов, а из фактов, связанных сетью причин и следствий. Для нашего мышления объяснение первичнее сообщения. Ребёнку трёх лет понятно, что отбирать игрушку у младшего брата и брат заплакал — вещи связанные, причём связанные определённым образом, и что если бы он её не отобрал, брат бы не заплакал. Для этого ему не требуется ни статистики, ни логики. И вот, продолжает Перл, откройте учебник статистики, науки об извлечении знания из данных. В предметном указателе многих из них вы не найдёте слова «причина». Зато найдёте предупреждение, которое слышал каждый студент: корреляция не есть причинность.

Предупреждение это справедливо, и мы только что, в предыдущей главе, видели, как дорого обходится забвение его. Но сформулировано оно странно. Оно говорит, чем причинность не является, и ничего не говорит о том, чем она является. Врач хочет знать, поможет ли лекарство больному. Законодатель хочет знать, что будет, если изменить налог. Судья хочет знать, погиб бы человек, если бы пожарный выход не был загорожен. Все они задают причинный вопрос. Статистик отвечает им числом, которое на этот вопрос, строго говоря, не отвечает, а затем прибавляет оговорку, что причинных выводов из него делать не следует. В результате причинный вывод делает читатель, но делает его тайком. В статье написано «связано с», а думает автор «вызывается» и надеется, что читатель подумает то же самое.

Это положение дел в XXI веке начало меняться. Перл называет произошедшее причинной революцией и, как всякий революционер, преувеличивает значение своего участия. Но у него есть основания. За несколько десятилетий усилиями небольшого числа информатиков, статистиков, эпидемиологов, а также, что для нас особенно важно, философов был создан язык, на котором можно записать причинный вопрос и выяснить, ответят ли на него данные. Это одно из самых заметных изменений в методологии науки за последние полвека, и в русскоязычной среде о нём знают ещё мало.

Для серии эта глава занимает особое положение. Причина была центральным понятием у Аристотеля, предметом сомнений у Юма, необходимой формой мышления у Канта, а затем её последовательно выводили из науки Мах, Пирсон и Рассел. Читатель первых томов знает, что у этого изгнания были серьёзные основания. Посмотрим сначала, как оно произошло и зачем было нужно, а потом разберём, чем оно кончилось.

Как причину изгнали из науки

Начало положил Юм, и начало это было двойным. В «Трактате» 1739 года он определил причину как предмет, за которым следует другой, причём все сходные предметы сопровождаются сходными. Причина есть постоянное соединение, и больше мы ничего не наблюдаем. Но в «Исследовании» 1748 года он добавил, как будто это то же самое другими словами: или, иначе говоря, если бы первого предмета не было, не было бы и второго. Перл с удовольствием отмечает, что связка «иначе говоря» скрывает здесь пропасть. Первое определение говорит о том, что можно наблюдать. Второе говорит о мире, которого нет. Весь дальнейший спор проходит между этими двумя фразами, причём два с лишним века побеждала первая. Петух всегда кричит перед восходом, значит, в смысле первого определения он его причина. Для второго определения это нелепость: если петух заболеет, солнце всё равно взойдёт.

В XIX и начале XX века позитивисты сделали из первой фразы программу. Мы подробно видели это во втором томе. Мах считал, что понятие причины несёт на себе следы первобытной привычки, а наука должна заменить его функциональной зависимостью между измеряемыми величинами. В 1913 году Рассел, зачитывая доклад в Аристотелевском обществе, произнёс слова, которые цитируют с тех пор все, кто пишет на эту тему. В передовых науках, вроде гравитационной астрономии, слово «причина» вообще не встречается, а закон причинности, по его словам, есть пережиток минувшей эпохи, уцелевший, как монархия, лишь благодаря ошибочной уверенности, что он безвреден. Основание у него было совершенно ясное. Уравнение Ньютона связывает силу, массу и ускорение симметрично. Оно не говорит, что чем вызвано. Его можно прочесть в обе стороны, а все времена в нём равноправны. Если физика, образец всех наук, обходится без причин, то зачем они остальным?

Самое же решительное изгнание произошло там, где оно имело наибольшие практические последствия, и началось оно почти случайно. 9 февраля 1877 года английский учёный Фрэнсис Гальтон, двоюродный брат Дарвина, путешественник и изобретатель дактилоскопии, читал в Королевском институте лекцию о законах наследственности. На лекции он показывал устройство собственной конструкции: доску с рядами штырьков, сквозь которые падают шарики и собираются внизу в колоколообразную кучу. Гальтон искал механизм. Если рост передаётся по наследству с добавлением случайных отклонений, то разброс в каждом поколении должен увеличиваться, как при удлинении доски. Но людей ростом в три метра нет. Почему распределение остаётся устойчивым?

Искал он около восьми лет и обнаружил явление, которому дал название возврата, или регрессии к посредственности. Сыновья очень высоких отцов в среднем ниже отцов. Гальтон сначала принял это за закон наследственности, а потом обнаружил нечто озадачившее его. Отцы очень высоких сыновей тоже в среднем ниже своих сыновей. Та же регрессия идёт и назад во времени, и между ростом и длиной предплечья у одного и того же человека. Но если следствие работает в обе стороны, это не причина. Это свойство самих чисел: всюду, где результат определяется способностью и везением одновременно, за необычно хорошим результатом следует более обычный, потому что удача не наследуется и не повторяется. Канеман позднее выразил это в формуле: успех есть талант плюс удача, большой успех есть немного больше таланта и намного больше удачи.

Здесь стоит задержаться, потому что регрессия к среднему объясняет почти всё в работе целителей. К знахарю, гомеопату и потомственному целителю идут тогда, когда хуже всего. Боль в спине, простуда, тяжёлая полоса, приступ тоски обычно колеблются вокруг какого-то среднего уровня. Человек обращается за помощью на пике, и в подавляющем большинстве случаев через некоторое время ему станет лучше, что бы он ни делал. Целителю не обязательно обманывать. Он наблюдает одно и то же сотни раз: пришли плохо, ушли лучше. Его

многолетний опыт совершенно реален. Нет в нём только контрольной группы. Точно так же тренер хвалит спортсмена после рекорда и ругает после провала, замечает, что после похвалы результат падает, а после ругани растёт, и выводит из этого педагогическую систему. Человек, не знающий о регрессии, увидит причину там, где есть только арифметика, причём увидит её очень убедительно.

Гальтон начал с поиска причин, а закончил понятием корреляции, и его младший современник превратил это в философию. Английский математик Карл Пирсон, близкий по взглядам к Маху и написавший в 1892 году «Грамматику науки», описывал впоследствии чтение гальтоновой книги о наследственности как освобождение: он почувствовал себя пиратом времён Дрейка. Причинность для него была лишь предельным случаем корреляции. Опыт показывает, что такая-то последовательность повторялась. Ни о какой необходимости наука ничего сказать не может. Сила как причина движения, писал он, стоит в одном ряду с богом дерева как причиной роста. Всё, что есть у науки, это таблицы сопряжённости.

Перл здесь не стесняется в выражениях. Для него Пирсон — фанатик, изгнавший причину из науки. Позитивизм, пишет он, считает вселенную продуктом человеческой мысли, а причинность вне головы бессмысленной. Читателю второго тома понятно, что это карикатура: Мах и Пирсон утверждали не то, что мир создан мыслью, а то, что наука описывает связи между наблюдениями и что всякое понятие должно быть оплачено проверкой. Юм и Рассел тоже имели основания. Перл, кстати, сам открыто признаёт, что пишет историю с точки зрения победителя, зная, чем всё кончилось, и объявляет такой подход своей гордостью. Будем помнить об этом, читая его. Но одно последствие запрета он описывает верно, и его можно проверить документально.

Возник так называемый сухой закон на причинность, продержавшийся больше полувека. Шотландский статистик Джордж Удни Юл изучал в 1899 году, как помощь бедным влияет на бедность. Он учёл долю пожилых и другие обстоятельства, то есть поступил как человек, желающий выделить причину. А в сноске к своей работе написал: вместо слов «вследствие» читайте «связано с». Этот образец действовал поколениями. Учёные думали «вследствие», а говорили «связано с». Разумеется, сами эти связи продолжали доставлять неудобства. Пирсон знал о ложной корреляции: он сам обнаружил её в собственных данных о черепах из парижских катакомб, где при смешении мужских и женских черепов появлялась связь, отсутствовавшая в каждой группе. Юл находил корреляцию 0,95 между смертностью в Англии и долей браков в англиканской церкви. Нужно было как-то сказать, что эти корреляции «ложные». Но ложные в сравнении с чем? С настоящими. А настоящие в их языке означали что? Чтобы отличить одни от других, нужно понятие причины, которое только что запретили. Таблица сопряжённости не содержит ответа.

Современный читатель встречает такие корреляции в заголовках чуть ли не ежедневно. Страны, где едят больше шоколада, получают больше Нобелевских премий на душу населения. Такой расчёт действительно был опубликован в серьёзном медицинском журнале, правда, отчасти в шутку. Числа там верны. Все понимают, что шоколад не прибавит таланта, и что за обеими величинами стоит богатство страны. Но почему мы в этом так уверены? Статистика сама этого не скажет. Такая уверенность возникает из причинных представлений, которые мы приносим к таблице сами.

Некоторые философы пытались вернуть причину иначе. Ханс Райхенбах, которого мы подробно обсуждали в третьем томе, в посмертно вышедшей книге «Направление времени» (1956) сформулировал принцип общей причины. Если две величины устойчиво связаны и ни одна не вызывает другую, то у них должна быть общая причина, при учёте которой связь исчезает. Если человек носит зонт и улица мокрая, то, учтя дождь, мы больше не увидим между ними связи. В этом замечательном наблюдении, которое называется экранированием, заключён один из корней всей современной теории. Позже Патрик Суппес и другие

предложили определять причину через повышение вероятности: X есть причина Y, если при X вероятность Y выше, чем без него. Но мороженое повышает вероятность преступлений в городе, поскольку и то и другое бывает летом, и никто не решит проблему уличной преступности запретом мороженого. Значит, надо учесть жару. А какие ещё обстоятельства надо учесть? Нэнси Картрайт в 1983 году дала честный ответ: те, которые имеют причинное отношение к делу. Тем самым определение причины стало содержать слово «причина». Перл называет эту программу разбившейся о подводную скалу смешения факторов. А сам он, как мы увидим, прошёл тот же путь: его байесовские сети 1980-х годов тоже писались на языке вероятностей, и фразу из книги 1988 года о том, что причинность есть удобный язык для разговора о вероятностных связях, он теперь вспоминает со смущением. Как мы отмечали в первой главе, сеть безразлична к тому, вызывает ли дым огонь или огонь дым.

Лестница причинности

Пока большинство статистиков соблюдало запрет, один человек шёл своей дорогой, почти никем не замеченный. Американский генетик Сьюэлл Райт (1889–1988) занимался наследованием окраски у морских свинок. Его интересовал вопрос, какая часть разнообразия окраса вызвана наследственностью, а какая обстоятельствами развития в утробе. Опыт над этим поставить трудно, но у Райта были данные за двадцать лет. В 1920 году он напечатал статью со странным рисунком: буквы, а между ними стрелки, показывающие, что на что влияет. Наследственность родителей, условия развития, окружение после рождения, окрас потомка. Перл считает этот рисунок первой опубликованной причинной диаграммой. Кроме рисунка, Райт дал правило, которое связывало причинные коэффициенты на стрелках с корреляциями, которые можно измерить. У свинок, скрещиваемых случайно, наследственность объяснила 42 процента разнообразия, а развитие 58. В инбредной линии, где наследственное разнообразие почти исчезло, соотношение стало 3 и 92. То есть корреляции всё-таки можно прочесть в обратную сторону и получить причинное утверждение, если заранее известно, как устроен процесс.

Вот пример, с которым Перл связывает всё значение этого метода. В статье 1921 года Райт спрашивает: насколько больше весит при рождении морская свинка, проведшая в утробе лишний день? Прямое сравнение даёт 5,66 грамма в день. Но эта цифра смешивает два процесса, потому что дольше вынашиваются детёныши из маленьких помётов, а они и без того крупнее. Если рост в утробе и размер помёта внести в диаграмму, задать простые уравнения и решить их, то получается 3,34 грамма. Причём сама скорость роста в опыте не измерялась. Первое число есть факт о данных. Второе есть ответ на вопрос о мире, в котором можно было бы подержать детёныша лишний день. Заметьте также, что на диаграмме особенно важны стрелки, которых нет: каждое такое отсутствие есть сильное допущение, которое можно оспорить.

Ответ пришёл из лагеря Пирсона в 1921 году. Молодой учёный Генри Найлз заявил, что противопоставлять причинность и корреляцию незачем, поскольку причинность есть просто совершенная корреляция, а вся затея Райта сомнительна: она заранее предполагает то, что должна доказать. Райт ответил, что и не утверждал, будто его метод позволяет выводить причины из данных без предварительного знания. Смысл в другом: из допущений и данных получается неочевидный вывод, который можно проверить. Метод игнорировали десятилетиями. Фишер, безраздельно властвовавший в статистике с 1920-х по 1950-е годы, его не принимал и видел в Райте врага по другим вопросам. Только через шестьдесят лет, в 1983 году, Райту, которому было уже за девяносто, пришлось ещё раз защищать свой метод перед видным стэнфордским статистиком, требовавшим анализа без модели. Анализа без модели, ответил он, не бывает, а тем, кому очень хочется его провести, он посоветовал подавить это желание и заняться чем-нибудь другим.

Язык, которого не хватало Райту, был разработан в конце XX века. Опираясь на него, Перл предложил образ лестницы с тремя ступенями. Это деление на самом простом уровне определяется тем, на какие вопросы способна ответить система.

На первой ступени стоит наблюдение, ассоциация. Вопрос звучит так: что я узнаю, если увижу X? Какова вероятность, что покупатель зубной пасты купит зубную нить? Здесь работают сова, выслеживающая мышь, традиционная статистика и, как утверждает Перл, современные обучаемые машины. Они могут делать великолепные предсказания, но хорошие предсказания, замечает он, не нуждаются в хорошем объяснении. Сове незачем понимать, почему мышь бежит из точки А в точку Б.

На второй ступени находится вмешательство. Вопрос: что произойдёт, если я сделаю X? Что станет с продажами зубной нити, если удвоить цену на пасту? На такой вопрос данные о прошлом сами по себе не отвечают. В прошлом цены могли быть высокими, например, из-за дефицита, а дефицит сам влиял на спрос. Вместо этого нужен опыт или модель, которая скажет, что произойдёт, когда мы вмешаемся. Для этого различия Перл ввёл особую запись. Если обычная условная вероятность P(Y при X) говорит о вероятности Y среди тех случаев, когда мы увидели X, то P(Y при do(X)) говорит о вероятности Y в том случае, когда мы сделали X принудительно, независимо от всех обычных причин. Английское слово do здесь значит просто «сделать». Классическая статистика не имела для этого обозначения. Вот различие, в котором, по Перлу, заключена вся суть дела: увидеть и сделать — разные вещи.

Вот и барометр, с которым мы встречались в третьем томе и которому там было обещано продолжение. Показание барометра связано с давлением через уравнение, которое можно написать как угодно: B равно k, умноженному на P, или P равно B, делённому на k. Алгебра равнодушна к направлению. Падение стрелки надёжно предсказывает бурю. На уровне наблюдения связь безупречна. Но если схватить стрелку пальцем и сдвинуть вниз, бури не будет. Или возьмём мир, в котором нет различия между P(Y при X) и P(Y при do(X)). В таком мире, как замечает Перл, люди перестали бы ходить к врачу, чтобы реже болеть, потому что среди посещающих врача больше больных, а города распустили бы пожарных, потому что где много пожарных, там много пожаров.

Третья ступень — это воображение, контрфактуал. Вопрос: что было бы, если бы я поступил иначе? Голова прошла после аспирина. Прошла бы она без него? Об этой ступени поговорим отдельно в следующем разделе.

Главное утверждение Перла теперь можно выразить так. С низшей ступени на высшую нельзя подняться только за счёт данных. Сколько ни накапливай наблюдений, они не скажут, что будет при вмешательстве, потому что данные суть факты, а факты не содержат ответа на вопрос о другом положении дел. Данные, говорит он, глубоко глупы. Нужно что-то ещё: нужна модель того, что чем вызвано. Причинная диаграмма и есть такая модель. Точки на ней обозначают величины, а стрелка от A к B утверждает: B слушает A. Определение метафорическое, и Перл это признаёт.

Откуда в нас такая способность? Перл предполагает, что она составляет отличие человека от других животных и что ей примерно сорок тысяч лет, если судить по фигурке человека с головой льва из пещеры в Германии: чтобы вырезать её из бивня мамонта, надо было вообразить то, чего нет. Это смелая догадка, и проверить её нелегко. Более проверяемы опыты американского психолога Элисон Гопник с маленькими детьми. Ребёнку показывают коробочку, которая загорается, когда на неё кладут некоторые предметы. Уже дошкольники не только замечают связь, но и по немногим примерам делают вывод, какой предмет действует, а какой просто лежал рядом, и, желая включить коробочку, правильно выбирают именно действующий. Иначе говоря, они уверенно отличают то, что видели, от того, что надо сделать. Разделение труда между врождённым и приобретённым тут остаётся предметом спора. Но если эти результаты верны, то причинное мышление не позднее изобретение философов, которое Рас-

сел мог выкинуть за ненадобностью, а способ, которым люди приобретают знания с самого начала.

Можно ли теперь вычислять вмешательства? Для этого есть простое правило, которое Перл назвал хирургией. Сделать так, чтобы событие произошло, значит освободить его от всех обычных причин и подчинить одной-единственной. На диаграмме это означает: удалить все стрелки, входящие в ту величину, на которую мы воздействуем. Пример с расстрельной командой делает это очевидным. Суд выносит приговор, капитан отдаёт приказ, солдаты А и Б стреляют, заключённый умирает. На уровне наблюдения: если выстрелил А, значит, был приказ, значит, выстрелил и Б. Теперь вмешаемся. Солдат А стреляет по собственной воле, никакого приказа не было. Что мы теперь можем сказать? Заключённый мёртв, а Б, скорее всего, не стрелял. Стрелка от капитана к солдату А вычеркнута, поэтому его выстрел больше ничего не говорит о приказе. Для чистой логики такая ситуация противоречива, поскольку правила запрещают стрелять без приказа, а для любого ребёнка нет.

Следующий шаг позволяет ответить на вопрос, мучивший статистиков и упомянутый в предыдущей главе. Какие переменные следует учитывать в анализе, а какие нет? Оказывается, любая диаграмма складывается всего из трёх видов соединений.

Цепь: огонь вызывает дым, дым вызывает срабатывание сигнализации. Если знать, есть ли дым, то огонь больше ничего не добавляет. Это экранирование Райхенбаха. Вилка: возраст ребёнка влияет и на размер обуви, и на умение читать. Отсюда ложная связь между обувью и чтением, которая исчезает среди семилетних. Большие ботинки ребёнка читать не научат. Эту вилку и называют смешением. Если общая причина не учтена, мы приписываем связь не тому, что её вызывает.

А третье соединение совсем неочевидно, и оно объясняет многое. Коллайдер, или столкновение: две стрелки ведут в одну точку. Талант и красота вместе влияют на то, становится ли человек известным. В населении в целом талант и красота между собой не связаны. Но если взять только известных людей, то среди них талант и красота окажутся связаны обратно: кто не очень красив, тот, скорее всего, очень талантлив, иначе бы его никто не узнал. Связь появилась из-за того, что мы отобрали людей по следствию. Американский врач Джозеф Берксон заметил в 1946 году то же самое в больницах: две болезни, не связанные в населении, оказываются связанными среди госпитализированных, поскольку больной с любой из них попадает в больницу с большей вероятностью. Позже канадский врач Дэвид Сакетт подтвердил это на конкретных данных: болезни костей встречались у 7,5 процента населения независимо от болезней лёгких, но у лежащих в больнице с болезнями лёгких у 25 процентов. Этой же логикой объясняется и задача о трёх дверях из первой главы.

Американский математик Джордан Элленберг приводит пример, который многие сразу примеряют к себе. Вам кажется, что красивые люди часто бывают неприятны? Вероятно, вы встречаетесь с людьми, которые либо красивы, либо приятны, либо и то и другое, а неприятного и некрасивого вы никогда не увидите рядом с собой за столиком. Отсюда обратная связь, которой в населении нет. По Перлу, такая корреляция есть иллюзия в самом точном смысле слова, причём нередко мы создаём её сами, выбирая, что учитывать, а что нет. То же самое с успешными предпринимателями, которые советуют бросить университет: мы видим только тех, кто прошёл отбор. Кто бросил и не преуспел, книг не пишет.

Вот почему совет учитывать всё подряд так вреден. Контроль над средним звеном вилки устраняет ложную связь. Но контроль над коллайдером её создаёт. Перл разбирает случай, когда поправка на такую переменную, казалось бы совершенно безобидную, порождает смещение там, где без неё ничего подобного не было: привычка пристёгиваться ремнём безопасности в анализе курения и болезней лёгких, включавшаяся в анализ как всего лишь ещё один показатель осторожности. Набор переменных для поправки определяется не свойствами данных и не их статистикой, а только структурой причинных связей. Для этого существует простой кри-

терий, названный критерием чёрного хода. Перекрой все некаузальные пути, начинающиеся со стрелки, входящей в изучаемую причину, и не трогай её следствий. Многие споры поколений статистиков о том, что считать смешивающим фактором, сводятся теперь к небольшой игре на бумаге, которая решается за доли секунды на компьютере. Определение смешения, замечает Перл, никогда не удавалось дать в статистических понятиях, потому что оно не статистическое.

Насколько это было неочевидно, показывает парадокс, поставивший в тупик медиков на сорок лет. В 1959 году американский биостатистик Якоб Ерушалми, изучая более пятнадцати тысяч детей, заметил, что дети курящих матерей в среднем весят меньше, а маловесные дети умирают гораздо чаще. Но если взять только маловесных, то дети курящих матерей выживают лучше. Значит ли это, что курение защищает младенца? Сам Ерушалми оставался скептиком насчёт вреда табака, и этот результат выглядел как довод в его пользу. Ответ был получен только в начале XXI века. Малый вес при рождении вызывается разными причинами: и курением, и серьёзными врождёнными пороками. Если взять только маловесных, то вес оказывается коллайдером. Ребёнок лёгок либо из-за курения, либо, чаще, из-за чего-то более опасного. Среди маловесных детей курящих матерей меньше детей с тяжёлыми пороками просто потому, что у их малого веса уже есть другое объяснение. Поэтому и выживают они лучше, а курение тут ни при чём. Перл заявляет, что загадка держалась так долго именно потому, что с 1960 по 1990 год не было языка, на котором её можно было выразить. И прибавляет тревожное замечание: этот случай заметили только благодаря тому, что вывод выглядел нелепо. А сколько таких ошибок прошло незамеченными, потому что они не противоречили ничьим ожиданиям?

Эта работа не осталась только за информатиками. Для нашей темы важно, что очень большой вклад внесли люди с кафедры философии. Кларк Глимур, основатель кафедры философии в Университете Карнеги-Меллона, и его коллеги Питер Спиртес и Ричард Шайнс в книге «Причинность, предсказание и поиск» (1993) поставили обратный вопрос: можно ли, зная только статистические независимости, определить возможные причинные структуры? Они построили программу TETRAD, которая при определённых условиях находит семейство диаграмм, согласных с данными. Это редкий случай, когда кафедра философии стала источником практического инструмента, которым затем пользуются биологи и специалисты по машинному обучению. Сюда же относится американский философ Джеймс Вудворд, чья книга «Как сделать так, чтобы что-то случилось» (2003) превратила эти идеи в теорию научного объяснения. Причина, по Вудворду, есть рукоятка: X причина Y, если есть такое воздействие на X, которое изменяет Y, и эта зависимость устойчива при других воздействиях. Это обобщение и на те случаи, когда воздействие практически невозможно: мы не можем подвинуть Луну, но вполне осмысленно спросить, что случилось бы с приливами, если бы её подвинули. Как показывает обзор Кристофера Хичкока, вмешательство можно определить без ссылки на людей, просто как независимую причину, которая перекрывает обычные механизмы.

Теперь становится понятным и то, что делает случайное распределение. Со времён Фишера рандомизированное испытание считают золотым стандартом, причём часто как нечто само собой разумеющееся. Что в нём происходит? В обычном случае лечение получают по множеству разных причин: кто-то богаче, кто-то осторожнее, у кого-то такая же болезнь была у родителей. Все эти неизвестные причины образуют стрелки, ведущие к лечению и к исходу. Если врач распределяет лечение по жребию, у величины «лечение» остаётся единственная причина: жребий. Все остальные стрелки разом отсекаются, в том числе и те, о которых никто не догадывается. Рандомизация просто и есть хирургия, проведённая не на бумаге, а в реальной жизни. Растения, как заметил Перл, карт не читают. В ней нет ничего мистического. Но и никакого священного статуса она не имеет. Иногда испытание невозможно: нельзя заставить людей курить двадцать лет. И в таких случаях наблюдение может дать законный, хотя и условный причинный вывод, если диаграмма оправдана.

Здесь Перл спорит и с обратной ошибкой, совсем не той, какую мы обсуждали в предыдущей главе. Там учёные говорили слишком много, здесь слишком мало. В 1998 году в ведущем медицинском журнале вышла работа по данным о жителях Гонолулу японского происхождения: за двенадцать лет из тех, кто проходил меньше мили в день, умерли 43 процента, а из проходивших больше двух миль 21,5 процента. С учётом возраста, физического состояния и образа жизни разница сохранялась. Авторы заявили, что влияние намеренной ходьбы их работа оценить не может. Перл считает такую скромность чрезмерной и не вполне честной: если вы считаете, что все существенные смешивающие факторы учтены, скажите это и назовите свой вывод условным. Точный перечень допущений, стоящих за выводом, ценнее, чем дежурная оговорка, которая освобождает автора от ответственности, но читатель всё равно всё поймёт как причину.

Здесь классический пример связан с Джоном Сноу, но обычно его рассказывают неверно. В популярной версии лондонский врач в 1854 году нанёс на карту случаи холеры, снял рукоятку с водоразборной колонки на Брод-стрит, и эпидемия прекратилась. На деле эпидемия к тому времени уже спадала. Настоящее достижение Сноу было куда умнее. В Лондоне воду поставляли две компании, причём в некоторых кварталах их трубы шли вперемешку. Одна из них брала воду из Темзы ниже места сброса городских нечистот, другая выше. Дома получали воду той или другой компании без всякой связи с богатством и образом жизни жильцов, часто даже не зная об этом. Смертность от холеры у потребителей первой компании оказалась примерно в восемь раз выше. Сама ситуация поставила опыт на трёхстах тысячах человек, и, как писал Сноу, ни одного опыта нельзя было бы придумать лучше. Поставщик воды играет здесь роль жребия, такие величины теперь называют инструментальными переменными. Сноу напечатал брошюру за свой счёт, и было продано 56 экземпляров. Вывод, между прочим, можно было предъявить и тем, кто верил в болезнетворные испарения: для решения о переносе водозабора мнение о механизме не важно. Современный вариант того же приёма называют менделевской рандомизацией: при зачатии каждый получает гены случайно, и некоторые гены влияют, например, на уровень определённого холестерина всю жизнь. Сравнивая таких людей, можно проверить, действительно ли этот холестерин вызывает болезнь. Именно таким способом в 2012 году выяснилось, что «хороший» холестерин, вопреки широко распространённому убеждению, от инфаркта, по-видимому, не защищает, хотя корреляция с меньшей смертностью у него есть.

А как быть с курением, где нельзя поставить опыт, но и инструмента нет? В 1950-е годы Фишер и другие скептики отвечали: может быть, существует ген, вызывающий и тягу к сигаретам, и рак. Это чистая вилка с неизмеренной общей причиной, и опровергнуть её прямым способом было невозможно. Американский статистик Джером Корнфилд в 1959 году сделал поразительный для своего времени ход, не ссылаясь на лаборатории. Он посчитал. Если курильщики болеют в девять раз чаще, то гипотетический ген должен встречаться у них по меньшей мере в девять раз чаще, чем у некурящих. Если он есть у 11 процентов некурящих, он должен быть у 99 процентов курильщиков. Если у 12 процентов, полное объяснение становится математически невозможным. Перл предлагает ещё один путь, называемый критерием парадного входа. Если курение влияет на рак только через смолы в лёгких, а скрытый ген на сами смолы не влияет, то действие курения можно вычислить вовсе без знания о гене. Оба допущения в случае курения спорны, что признавал сам Перл после возражений американского статистика Дэвида Фридмана. Но на других данных, о программах профессиональной подготовки, этот приём дал результат, близкий к рандомизированному опыту. Появилось полное исчисление с тремя простыми правилами для устранения do из выражения. В 2006 году две группы независимо доказали его полноту: если правила не помогают, то никакая хитрость не поможет, и остаётся только ставить опыт. Вопрос «можно ли в принципе ответить на это по наблюдениям?» получил математически точный ответ.

У этой истории есть продолжение, в котором вся глава проверяется на деле. Фишер оказался отчасти прав: ген действительно есть. Методы поиска по всему геному обнаружили в 2008 году вариант на пятнадцатой хромосоме, связанный с никотиновыми рецепторами. Его носители курят немного больше и немного чаще заболевают. Но кому именно нужно верить теперь: Фишеру или Корнфилду? Численный ответ получается без спора. Две копии варианта, которые есть примерно у одного человека из девяти, увеличивают риск на 77 процентов, тогда как само курение увеличивает его более чем в десять раз. А когда американский эпидемиолог Тайлер Вандервил разделил действие гена на прямое и косвенное, то косвенное, через курение, оказалось ничтожным: один-три процента, примерно одна лишняя сигарета в день. Подозреваемый, которого Фишер назвал вслепую, был настоящим. Он не имел никакого отношения к тому преступлению, в котором его обвиняли. Такие количественные ответы выгодно отличаются от споров, которые велись в этой области сорок лет назад, зато всё ещё оставляют материал для возражений: одна из первооткрывательниц гена считает, что носители могут глубже затягиваться.

И ещё один пример, показывающий, зачем всё это нужно за пределами лаборатории. Бывает, что причина известна, а механизм нет. В 1747 году шотландский флотский врач Джеймс Линд показал в одном из первых контролируемых опытов, что цитрусовые излечивают цингу. Цинга, унёсшая, по оценкам, миллионы моряков, на британском флоте почти исчезла. Затем произошла удивительная вещь. Считалось, что помогает кислота. Поэтому испанские лимоны заменили вест-индскими лаймами, столь же кислыми, но содержащими вчетверо меньше витамина С, а сок для долгого хранения ещё и подвергали обработке, разрушавшей витамин. Когда в XX веке экспедиции ушли в Арктику и Антарктику, цинга вернулась, и врачи решили, что лимоны тут ни при чём, а всё дело в испорченном мясе. Роберт Скотт в 1911 году тщательно проверил мясо и не взял никаких цитрусовых. Что бы ни погубило его отряд в итоге, по Перлу среди причин была цинга. Механизм «цитрусовые, кислота, цинга» был неверен. Истинной была цепочка «цитрусовые, витамин С, цинга». В Венгрии в 1930-е годы эту аскорбиновую кислоту выделил Альберт Сент-Дьёрдьи. Такое промежуточное звено называют посредником, и верное знание о нём позволяет предсказать, что случится при замене одного средства другим. Заметьте, что простая проверка «работают ли лимоны» здесь ничего бы не изменила: они работали всегда. Проверить нужно было не связь, а модель связи.

Обратная ошибка случается, когда посредника учитывают там, где это недопустимо, и последствия могут быть очень конкретными. Американский военный хирург Джон Крэг, работавший в Багдаде в 2006 году, настойчиво отстаивал жгут, который долго считался крайней мерой. По некоторым оценкам, жгуты спасли за десятилетие около двух тысяч жизней. Но работа 2015 года, сравнившая раненых с наложенным жгутом и без него, никакой разницы в выживаемости не нашла. Перл указывает, что собирали данные только о тех, кто успел живым добраться до госпиталя. Между тем жгут действует прежде всего тем, что позволяет дожить до госпиталя. Анализ тем самым заранее исключил основной путь, которым средство могло помочь. Статистически всё правильно. По существу это вопрос о том, как устроена цепь, и на него сами данные не отвечают.

Та же история повторяется и с дискриминацией. В 1973 году Калифорнийский университет в Беркли обнаружил, что в аспирантуру приняли 44 процента мужчин и 35 процентов женщин. Но в отдельных отделениях женщин принимали не реже, а часто чаще. Американский статистик Питер Бикел показал, что женщины чаще подавали на те отделения, куда труднее попасть. Значит ли это, что дискриминации нет? Другой американский статистик, Уильям Краскал, построил контрпример, показав, что без знания структуры на это ответить нельзя. Это и называют парадоксом Симпсона, и суть его вовсе не в числах, а в том, что одни и те же числа могут требовать разного прочтения. Если лекарство хуже в каждой группе, а в сумме лучше, то при одной истории правы отдельные группы, а при другой сумма. В таблице об этом

не написано ничего. Перл настаивает, что таблица без истории о том, как получены данные, не отвечает ни на один вопрос о действии. Сноска Юла вернулась с полным счётом за столетие молчания.

Перл даёт условный числовой пример. В опыте с новым лекарством инфаркт случается у женщин в 5 процентах случаев без лекарства и в 7,5 с ним; у мужчин в 30 и 40 процентах. В каждой группе лекарство вредит. Но если сложить, то без лекарства инфаркт у 22 процентов, а с ним у 18. Вредно для мужчин, вредно для женщин, полезно для всех? Лекарства с такими свойствами не существует, и наша интуиция протестует совершенно правильно. Но какому результату верить? Если женщины реже соглашаются принимать лекарство и реже страдают инфарктами, то пол есть смешивающий фактор, и надо смотреть по группам: лекарство вредно. А теперь возьмём те же самые числа, только вместо пола поставим давление. Лекарство снижает давление, а уже от него зависит инфаркт. Давление есть посредник, и правильно считать по сумме: лекарство полезно. Цифры одни. Разница только в том, что произошло раньше и что на что влияет. Об этом не может сказать никакая таблица, в какой бы большой базе данных она ни находилась.

Миры, которых нет

Вернёмся теперь к второму определению Юма и к третьей ступени лестницы. Вопрос «что было бы, если бы?» кажется либо детской игрой, либо чистой метафизикой. Данные по определению суть факты. Они говорят о том, что случилось, и не могут сказать о том, что случилось бы. Человек принял лекарство и умер. Никакая таблица не покажет того же самого человека, в той же самой жизни, без лекарства. Для статистиков это было основанием держаться от таких высказываний подальше: контрфактуальные утверждения казались им слишком зыбкими, как и для большинства философов науки XX века. Гемпель и Гудмен потратили на них немало сил, потому что законы природы, в отличие от случайных обобщений, поддерживают контрфактуалы. Если бы эта спичка была чиркнута, она бы загорелась. А если бы эта монета лежала в моём кармане, она бы не обязательно была серебряной, даже если все монеты в кармане серебряные. Они ясно видели, что такой разницы между законом и случайностью нельзя определить без такого словаря.

В 1973 году американский философ Дэвид Льюис предложил решение, поражающее смелостью. Высказывание «если бы A, то B» истинно, если в ближайшем к нашему из возможных миров, где есть A, есть и B. Выпил Джо аспирин, и голова прошла. Мир, в котором он его не выпил, а всё остальное то же, ближе к нашему, чем мир, где он вдобавок ударился о дверь ванной, поэтому нужно смотреть, болит ли голова в первом. Причина теперь определялась просто: A вызвало B, если при отсутствии A не было бы и B. Всё прекрасно, кроме одного: что это за миры и как установить их близость? Льюис не уклонялся. Он считал, что все возможные миры существуют столь же реально, как наш, и в одном из них действительно есть говорящие ослы. В некрологе 2001 года газета «Нью-Йорк таймс» приводила определение, данное его взглядам: бешеный модальный реалист. Логика у него безупречна, но физик, спросивший, как проверить это на опыте, не получит ответа.

Перл делает ход, который для нашей серии заслуживает особого внимания. Не нужно спорить, пишет он, существуют ли такие миры как физические или хотя бы метафизические сущности. Если мы стремимся объяснить, что люди имеют в виду, говоря «A вызвало B», достаточно предположить, что люди умеют порождать альтернативные миры в голове, согласованно решать, какой из них ближе к нашему, и приходить к одному и тому же. А согласие возникает потому, что у них общая модель. Контрфактуал, как он пишет в другом месте, сводится к величинам, полученным безобидными действиями над уравнениями или диаграммами. Близость мира измеряется просто: тем, сколько стрелок пришлось удалить из модели. Логика, лишённая

способа представления, говорит он, есть метафизика. Контрфактуалы от этого перестают быть мистикой и становятся вычислением. Двое, разделяющие одну причинную модель, разделят и все свои контрфактуальные суждения. В том же духе он сравнивает их с мнимыми числами, которые тоже долго казались подозрительными, пока их полезность не сняла все сомнения.

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.