

СТЮАРТ РИЧИ

НАУЖО-
ОБРАЗ-
НАЯ
ЧУШЬ



Corvus

РАЗОБЛАЧЕНИЕ
МОШЕННИЧЕСТВА,
ПРЕДВЗЯТОСТИ,
НЕДОБРОСОВЕСТНОСТИ
И ХАЙПА В НАУКЕ



Книжные проекты
Дмитрия Зимина

Стюарт Ричи
Наукообразная
чушь. Разоблачение
мошенничества, предвзятости,
недобросовестности
и хайпа в науке

Текст предоставлен правообладателем

http://www.litres.ru/pages/biblio_book/?art=70284787

С. Ричи. Наукообразная чушь. Разоблачение мошенничества, предвзятости, недобросовестности и хайпа в науке: ООО

«Издательство Аст»; Москва; 2024

ISBN 978-5-17-136104-4

Аннотация

Чтобы определиться со своим мировоззрением и принимать взвешенные решения, влияющие на нашу жизнь, нам среди прочего необходимо сформировать свое отношение к высказываниям экспертов и результатам научных исследований. “Ученые врут и все скрывают” или “ученые доказали, значит, так и есть”? Между двумя этими крайностями лежит огромное поле неоднозначности, и книга Стюарта Ричи помогает нам сориентироваться на нем, найти свою позицию.

Приводя потрясающие воображение примеры, Ричи описывает системные проблемы в публичном освещении научных работ и устройстве самой научной деятельности, позволяя критически переосмыслить все, что мы читали, смотрели и слушали о науке до сих пор. Однако надежды на прогресс он у нас вовсе не отнимает – напротив, мы много узнаем о том, что делается и что еще можно сделать, чтобы доверие к работе ученых крепло на более надежных и устойчивых основаниях.

В формате PDF A4 сохранен издательский макет.

Содержание

Предисловие	7
Часть I. Как должно быть – и как на самом деле	26
Глава 1. Как работает наука	26
Глава 2. Кризис воспроизводимости	49
Конец ознакомительного фрагмента.	85

Стюарт Ричи
Научнообразная
чушь. Разоблачение
мошенничества,
предвзятости,
недобросовестности
и хайпа в науке

Stuart Ritchie

Science Fictions. Exposing Fraud, Bias, Negligence and Hype
in Science

© Stuart Ritchie 2020

© А. Якименко, перевод на русский язык, 2024

© А. Бондаренко, художественное оформление, макет,
2024

© ООО “Издательство Аст ”, 2024

Издательство CORPUS ®

* * *

Посвящается Кэтрин

*Вот это – научный факт. Реальных
доказательств тому нет, но это научный факт.
Из сериала Brass Eye¹*

¹ Morris C. et al. *Paedogeddon!* // *Brass Eye*. Series 2. Episode 1. 26 July 2001.

Предисловие

...Все же уму человеческому постоянно свойственно заблуждение, что он более поддается положительным доводам, чем отрицательным...²
Фрэнсис Бэкон “Новый Органон” (1620)³

В январе 2011 года мир узнал, что студенты обладают экстрасенсорными способностями.

Результаты из новой научной статьи произвели сенсацию: в серии лабораторных экспериментов с участием более чем тысячи человек было доказано существование экстрасенсорного предвидения – умения заглянуть в будущее с помощью сверхчувственного восприятия. Статью написал не какой-то неизвестный безумец, а ведущий профессор психологии Дэрил Бем из Корнеллского университета, входящего в Лигу плюща. И опубликована она была не абы где, а в одном из самых главных научных журналов по психологии, уважаемом и рецензируемом⁴. Казалось, наука официально одобрила феномен, до того считавшийся абсолютно невозможным.

² Перевод С. Краси́льщикова.

³ Bacon F. *Novum Organum*. New York: P. F. Collier & Son, 1620/1902. [Бэкон Ф. *Новый Органон* // Соч. в 2-х т. Т. 2. М.: Мысль, 1972.]

⁴ Bem D. J. *Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect*. *Journal of Personality and Social Psychology*. 100, no. 3 (2011): 407–25.

Я тогда был аспирантом, занимался психологией в Эдинбургском университете. И прилежно ознакомился со статьей Бема. Вот в чем заключался один из экспериментов. Студенты смотрели на экран компьютера, где появлялись два изображения занавеса. Им сказали, что за одним из изображений есть еще картинка и нужно предположить, за каким именно, нажав на него. Поскольку никакой дополнительной информации предоставлено не было, студенты могли лишь гадать. Когда они делали выбор, занавес исчезал – и они видели, правы оказались или нет. После тридцати шести повторений эксперимент завершался. Получилось нечто странноватое. Когда за занавесом прятался какой-нибудь нейтральный, скучный объект вроде стула, результат был почти идеально случайным: студенты делали правильный выбор в 49,8 % случаев, то есть вероятность угадать была, по сути, пятьдесят на пятьдесят. Однако – и в этом-то вся странность – когда спрятанная картинка была порнографической, студенты выбирали соответствующее изображение занавеса чуточку чаще, чем это предписывается простым угадыванием: в 53,1 % случаев, если быть точным. Разница превысила порог так называемой статистической значимости. В статье Бем предположил, что некое подсознательное, экстрасенсорное половое влечение самую малость подталкивало студентов к эротической картинке еще до того, как та появлялась на экране⁵.

⁵ Студенты также обладали обратной способностью: когда за одним из зана-

Некоторые другие эксперименты Бема были менее откровенными, но не менее обескураживающими. В одном из них на экране поочередно появлялись сорок не связанных между собой слов. Затем проводилась проверка памяти – студенты должны были напечатать столько слов, сколько им удалось запомнить. Далее компьютер случайным образом выбирал двадцать слов из списка и снова высвечивал их на экране. На этом эксперимент завершался. Бем сообщил, что во время теста на запоминание студенты чаще припоминали те двадцать слов, *которые вскоре снова увидят*, хоть и не могли знать – разве что благодаря некой сверхъестественной интуиции, – какие именно слова им покажут. Как если бы студент готовился к экзамену, сдавал его, затем занимался еще уже после сдачи – и вот эта постэкзаменационная зубрежка каким-то образом прокладывала себе путь назад во времени, улучшая отметку. Времени положено идти лишь в одном направлении, если, конечно, законы физики внезапно не утрачивают своей силы. Причинам положено предшествовать своим следствиям, а не наступать после них. Однако публикация статьи Бема сделала эти странные результаты частью научной литературы.

Что важно, эксперименты Бема совершенно нехитрые, они не требуют ничего сложнее стационарного компьютера.

весов скрывалась жестокая картинка, они психически дистанцировались от нее, выбирая соответствующее изображение лишь в 48,3 % случаев – отличие от результата, который должен был бы получиться при чистом угадывании, опять-таки оказалось статистически значимым.

Если Бем прав, любой исследователь мог получить доказательство существования паранормального, просто следуя его описанию того, как проводились эксперименты, – даже аспирант с примерно нулевыми ресурсами, каким был я. Именно это я и сделал: связался с двумя другими психологами, тоже сомневавшимися в объявленных результатах, Ричардом Вайзманом из Университета Хартфордшира и Крисом Френчем из Голдсмитского университета Лондона, и мы договорились трижды повторить эксперимент Бема со списком слов, по разу в каждом из трех наших солидных университетов. Несколько недель спустя, в течение которых мы набирали участников, ждали, пока они пройдут тест на запоминание, и смотрели, как на их лица наползает недоумевающее выражение, когда в конце мы объясняли, что именно проверяем, – итак, несколько недель спустя мы получили результаты. И они... не показали ничего. Наши студенты не были ясновидцами: слова, которые после теста снова высвечивались на экране, они припоминали не лучше и не хуже, чем остальные. Пожалуй, законы физики все-таки были вне опасности.

Мы надлежащим образом изложили свои результаты и отослали статью в тот же научный журнал, где была опубликована работа Бема, – *Journal of Personality and Social Psychology*. И почти моментально нам дали от ворот поворот. Редактор отклонил статью в течение нескольких дней, аргументировав это тем, что политика журнала – никогда не публиковать статьи, посвященные повторению каких-то преды-

дущих экспериментов, вне зависимости от того, воспроизвелись прежние результаты или нет⁶.

Разве неправы мы были, почувствовав себя оскорбленными? Журнал опубликовал статью, в которой были сделаны чрезвычайно смелые заявления – заявления, не просто представляющие интерес для психологов, а такие, что в корне изменили бы науку, окажись они правдивы. Результаты этой статьи попали в публичное пространство и получили широкую огласку в популярных медиа, в частности, Бема пригласили на вечернюю телепрограмму *The Colbert Report*, где ведущий употребил эффектное выражение “порно, путешествующее во времени”⁷. И однако же редакторы журнала отказались даже рассматривать возможность публикации статьи, рассказывающей о повторении исследования и ставящей сенсационные результаты под сомнение⁸.

⁶ Aldhous P. *Journal Rejects Studies Contradicting Precognition*. New Scientist. 5 May 2011.

⁷ The Colbert Report. *Time Travelling Porn – Daryl Bem*. 2011.

⁸ После еще нескольких отказов мы в конце концов опубликовали статью в другом научном журнале: Ritchie S. J. et al. *Failing the Future: Three Unsuccessful Attempts to Replicate Bem’s “Retroactive Facilitation of Recall” Effect*. PLOS ONE. 7, no. 3 (2012): e33423. Примечательно, что отклонивший нашу статью журнал опубликовал критические замечания, касающиеся статистического анализа в исследовании Бема (Wagenmakers E.-J. et al. *Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011)*. Journal of Personality and Social Psychology. 100, no. 3 (2011): 426–32), а также ответ Бема и его коллег (Bem D. J. et al. *Must psychologists change the way they analyze their data?* Journal of Personality and Social Psychology. 101, no. 4 (2011): 716–9). И все равно редколлегия журнала не собиралась рассматривать возможность публикации статьи о

А тем временем разворачивался и другой эпизод, тоже поднимающий тревожные вопросы о современном состоянии научной практики. Журнал *Science*, считающийся одним из самых престижных научных журналов в мире (уступающий разве что *Nature*), опубликовал статью Дидерика Стапела, социального психолога из Тилбургского университета в Нидерландах. В статье под заглавием “Преодолевая хаос” описывалось несколько исследований, проведенных в лаборатории и на улице, которые продемонстрировали, что люди сильнее подвержены предубеждениям – и расовым стереотипам – в более беспорядочной и загрязненной среде⁹. Эта статья и некоторые другие из десятков работ Стапела произвели фурор во всем мире. “Хаос способствует стереотипам” – написала новостная служба журнала *Nature*; “Где мусор – там расизм” – грянула *Sydney Morning Herald*¹⁰. Это был пример того типа исследований в области социальной психологии, что дают легкие для понимания результаты с “ясными практическими выводами”, как написал сам Стапел, – в данном случае следовало “заблаговременно выявлять беспорядок в окружающей среде и немедленно вмешиваться”¹¹.

попытке повторить исходное исследование. Далее мы увидим, что редакторы журнала с тех пор пересмотрели свое отношение к этой важнейшей проблеме.

⁹ Stapel D. A., Lindenberg S. *Coping with Chaos: How Disordered Contexts Promote Stereotyping and Discrimination*. *Science*. 332, no. 6026 (2011): 251–3.

¹⁰ Ball P. *Chaos Promotes Stereotyping*. *Nature*. 7 April 2011; Phillips N. *Where There's Rubbish There's Racism*. *Sunday Morning Herald*. 11 April 2011.

¹¹ Stapel D. A., Lindenberg S. *Coping with Chaos*.

Проблема заключалась в том, что все это было неправдой. Некоторые коллеги Стапела насторожились, заметив, что результаты его экспериментов *слишком уж* идеальны. Мало того, обычно серьезные ученые настолько загружены, что доверяют черную работу вроде сбора данных своим студентам, однако Стапел будто бы сам выходил и собирал все данные собственноручно. После того как его коллеги в сентябре 2011 года поделились своей озабоченностью с университетским руководством, Стапел был отстранен от работы. За этим последовали многочисленные разбирательства¹².

В автобиографии-исповеди, которую он потом написал, Стапел признался: вместо того чтобы собирать данные для своих исследований, он допоздна засиживался один в своем кабинете или за письменным столом у себя дома и вбивал в таблицу числа, необходимые для его воображаемых результатов, просто-напросто их выдумывая. “Я делал ужасные вещи, возможно даже отвратительные, – писал он. – Я подделывал научные данные и сочинял исследования, которых никогда не было. Работал я один, отчетливо осознавая, что делаю... И ничего не чувствовал: ни отвращения, ни стыда, ни сожалений”¹³. Поразительно, но его мошенничество было сложноорганизованным и проработанным до мелочей. “Я

¹² Levelt Committee et al. *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel [English Translation]*. 28 Nov. 2012.

¹³ Stapel D. A. *Faking Science: A True Story of Academic Fraud*. Tr. Brown N. J. L. Strasbourg, France, 2014, 2016. <http://nick.brown.free.fr/stapel>

выдумывал целые школы, где якобы проводил свои исследования, учителей, с которыми якобы обсуждал эксперименты, лекции, которые якобы читал, уроки по социальным наукам, проведению которых якобы способствовал, подарки, которые якобы дарил людям в благодарность за их участие в проекте”¹⁴.

Стапел описал, как распечатывал пустые бланки будто бы для того, чтобы раздать участникам исследований, показывал их коллегам и студентам, заявляя, что вот-вот запустит новый проект, а потом вдали от чужих глаз... выбрасывал всю эту макулатуру в бак для переработки. Это не могло больше продолжаться. Результаты инициированных разбирательств были однозначны – Стапела уволили вскоре после временного отстранения от работы. С тех пор не менее пятидесяти восьми его статей были отозваны – изъяты из научной летописи – из-за фальсификации данных.

Случаи с Бемом и Стапелом, когда уважаемые ученые публиковали на первый взгляд невозможные (Бем) или откровенно жульнические (Стапел) результаты, хорошо встряхнули дисциплину психологических исследований и вообще всю научную область. Как авторитетные научные журналы допустили эти работы до публикации? И скольким еще опубликованным ранее статьям нельзя верить? Как выяснилось, эти случаи послужили прекрасным примером гораздо более глубоких проблем в том, как мы занимаемся на-

¹⁴ Там же.

укой.

В обоих случаях ключевой вопрос имел отношение к *воспроизведению результатов*. Научное открытие заслуживает того, чтобы его воспринимали всерьез, если это не что-то возникшее по случайности, или из-за неисправностей оборудования, или по милости ученого, мухлюющего или замалчивающего какие-то нестыковки. Это открытие должно действительно произойти. И если оно произошло, то я теоретически могу пойти и получить примерно те же результаты, что и вы. Во многом это и есть суть науки, именно это отличает ее от других способов познания мира: если какой-то результат не воспроизводится, трудно вообще назвать его научным.

Итак, тревогу вызвало не столько то, что эксперименты Бема были сомнительными, а Стапела – плодом его воображения (какие-то просчеты и ложные результаты неизбежны, они будут всегда¹⁵, а стало быть, увы, всегда будут и мошенники), настоящая проблема заключалась в том, как научное сообщество повело себя в обеих ситуациях. Наша с коллегами статья, посвященная воспроизведению экспериментов Бема, была бесцеремонно отвергнута журналом, опублико-

¹⁵ Действительно, научный прогресс зависит от нашей способности находить допущенные ранее ошибки и недоработки. Например, в начале XX века физики осознали, что классическая механика Ньютона, долгое время считавшаяся универсальной, не может объяснить поведение очень маленьких и очень быстрых частиц, и тогда разработали квантовую механику. Обсуждение этого вопроса с точки зрения измеримых величин вроде скорости света и постоянной Планка см. здесь: Milton M. J. T., Possolo A. *Trustworthy Data Underpin Reproducible Research*. *Nature Physics*. 16, no. 2 (2020): 117–9.

вавшим исходную работу, а в случае Стапела почти никто даже никогда не *пытался* воспроизвести его результаты¹⁶. Иными словами, научное сообщество продемонстрировало, что согласно принимать на веру громкие заявления, звучащие в этих статьях, не проверяя, насколько на самом деле излагаемые результаты достоверны. А если воспроизводимость результатов никак не проверяется, откуда нам вообще знать, что они не просто случайны или сфальсифицированы?

Пожалуй, лучше всего отношение многих ученых к воспроизведению экспериментов выразил сам Бем в одном интервью через несколько лет после его печально известного исследования. “Я обеими руками за научную строгость, – сказал он, – но мне самому не хватает на нее терпения... Если бы вы ознакомились с предыдущими моими экспериментами, вы бы увидели, что все они служили своего рода риторическими приемами. Я собирал данные для того, чтобы изложить свою точку зрения. Я использовал данные как средство убеждения и, в сущности, никогда не переживал о том, воспроизведутся мои результаты или нет”¹⁷.

¹⁶ Я нашел единственный пример – вот эту статью, которая была опубликована спустя долгое время после разоблачения мошенничества Стапела, хотя само исследование проводилось до того: IJzerman H. et al. *Perceptual Effects of Linguistic Category Priming: The Stapel and Semin (2007) Paradigm Revisited in Twelve Experiments*. Acta Psychologica. 157 (2015): 23–9. Больше о затаянных попытках опубликовать это исследование (что кое в чем напоминает мой собственный опыт с работой Бема) написано здесь: Zwaan R. *When Replicating Stapel is not an Exercise in Futility*. Zeitgeist. 18 Jan. 2015.

¹⁷ Цитируется по: Engber D. *Daryl Bem Proved ESP Is Real: Which Means Science*

Беспокоиться, воспроизведутся ли результаты, – отнюдь не что-то опциональное. В этом основной смысл науки, смысл, который должен явно обнаруживаться в системе рецензирования и журнальных публикаций, играющей роль бастиона на пути ошибочных открытий, ложно понятых экспериментов и сомнительных данных. Однако, как будет показано дальше, эта система серьезно поломана. Ученые получают важные результаты, но они расцениваются как не слишком интересные для публикации, так что их изменяют или прячут, искажая научную летопись и вредя нашей медицине, технологиям, методикам обучения и государственной политике. Огромные ресурсы, вливаемые в науку в надежде на полезную отдачу, растрачиваются на исследования, абсолютно не несущие новых знаний. Легко предотвратимые ошибки и ляпы регулярно прорывают линию Мажино системы рецензирования. Книжки, сообщения средств массовой информации и наши головы наполняются “фактами”, которые неверны, преувеличены или заведомо ложны. И в самых худших случаях, особенно когда дело касается медицинской науки, гибнут люди.

В других книгах ученые выводятся как борцы с целой галереей аферистов, представителей псевдонауки: с креационистами, гомеопатами, плоскоземельцами, астрологами и иже с ними, которые своей деятельностью олицетворяют полное непонимание науки и совершают над ней надруга-

тельство – обычно неосознанно, иногда злонамеренно и всегда безответственно¹⁸. Моя же книга иная. В ней раскрывается глубинное разложение самой науки: разложение, влияющее на среду, в которой проводятся исследования и публикуются их результаты. Наука – практика, которая должна отличаться строжайшим скептицизмом, самым острым рационализмом и трезвым эмпиризмом, – стала вместилищем ошеломляющего количества заблуждений, некомпетентности, лжи и самообмана. И при этом основная цель науки – пробиваться как можно ближе к истине – подрывается.

В первой части этой книги будет показано, что занятие наукой подразумевает куда больше, чем просто проведение экспериментов и проверку гипотез. Наука по своей сути *социальна*, ведь вы должны убеждать других людей – остальных ученых – в том, что обнаружили. А поскольку наука – это еще и *человеческая* сфера, мы понимаем, что любой ученый будет подвержен человеческим особенностям, таким как иррациональность, предвзятость, сбой внимания, внутригрупповой фаворитизм и откровенное жульничество для получения желаемого. Чтобы ученые могли убеждать друг друга, стараясь при этом преодолеть неотъемлемые ограничения человеческой природы, наука создала систему сдер-

¹⁸ Классический пример такой книги: Sagan C. *The Demon-Haunted World: Science as a Candle in the Dark*. New York: Ballantine Books, 1997. [Саган К. *Мир, полный демонов. Наука – как свеча во тьме*. М.: АНФ, 2021.]

жек и противовесов, которая – в теории – отделяет научные зерна от плевел. Такой процесс тщательного изучения и проверки на достоверность, приводящий к предполагаемому золотому стандарту публикаций в рецензируемых научных журналах, описывается в первой главе. Однако во второй главе показывается, что этот процесс катастрофически нарушен: в различных областях науки существуют бесчисленные опубликованные результаты, которые не воспроизводятся и достоверность которых крайне сомнительна.

Затем, во второй части книги, мы зададимся вопросом, почему так происходит. Мы обнаружим, что наша публикационная система, вместо того чтобы устранять или преодолевать всяческие человеческие факторы, позволяет им накладывать свой отпечаток на научную летопись – и именно потому, что *считает* себя объективной и беспристрастной. Возобладала своеобразная самоуверенность, странная самоуверенность, когда само существование системы рецензирования словно бы мешает нам распознавать ее недостатки. Казалось бы, рецензируемые статьи должны настолько приближаться к объективной фактографии устройства нашего мира, насколько это вообще возможно. Однако в ходе обсуждения многих десятков таких статей мы увидим, что нельзя надеяться, будто рецензирование гарантирует, что ученые будут честно (глава третья), непредвзято (глава четвертая), добросовестно (глава пятая) обращаться со своими результатами и трезво их оценивать (глава шестая).

В третьей части книги мы глубже погрузимся в научную практику. В седьмой главе будет показано: дело не только в том, что система не справляется со всеми обсуждаемыми видами злоупотреблений, – в действительности современное устройство научно-исследовательской деятельности *стимулирует* возникновение этих проблем, подталкивая ученых к заикленности на авторитете, известности, финансировании и репутации в ущерб строгости и надежности результатов. Наконец, после того как мы диагностируем проблему, в восьмой главе будет описан ряд зачастую радикальных реформ научной практики, которые могли бы помочь переориентировать ее в направлении исходной цели – открывать факты о нашем мире.

Чтобы обосновать слабости научно-исследовательского процесса, по ходу повествования я буду рассказывать поучительные истории из самых разнообразных областей науки. Отчасти потому, что я психолог, преобладать будут примеры из психологии¹⁹. Но в этой книге будет много говорить-

¹⁹ В этой книге я рассматриваю множество сучков в глазах других ученых, поэтому – если вы простите мне минутку самоанализа – правильно будет проверить наличие любых потенциальных бревен в собственном глазу. За годы, прошедшие с моих попыток воспроизвести результаты Бема, у меня вышло много статей на разные темы, хотя преимущественно все-таки про человеческий интеллект, интересующий меня больше всего. В первую очередь нужно сказать, что я никогда умышленно не подделывал результаты. Однако глупо было бы думать, что я неуязвим для предубеждений. Они часто – или, возможно, обычно – бессознательны, и история некоего исследования легко переписывается так, что начинает казаться, будто исходно вы именно таким его и задумывали. Хорошо, что я

ся о психологии не только из-за моего образования и опыта работы. Причина также в том, что после инцидентов с Бемом и Стапелом (в числе многих других) психологи начали заниматься глубоким самокритическим анализом. Пожалуй, активнее, чем коллеги в любой другой области науки, мы, психологи, стали выявлять свои укоренившиеся недостатки и разрабатывать систематические подходы для борьбы с ни-

опубликовал немало отрицательных результатов, то есть статей, в которых не было найдено подтверждений основной гипотезе. См., например: Ritchie S. J. et al. *Polygenic Predictors of Age-Related Decline in Cognitive Ability*. *Molecular Psychiatry*. 25 (2020): 2584–98; а также мою самую первую научную публикацию: Ritchie S. J. et al. *Irlen Colored Overlays Do Not Alleviate Reading Difficulties*. *Pediatrics*. 128, no. 4 (2011): e932–8. Опять-таки мне легко можно возразить, что в этой моей дебютной статье с отрицательными результатами изучалась слишком маленькая выборка и потому реальные эффекты могли быть упущены (см. обсуждение статистической мощности в пятой главе). Некоторые мои статьи подвергались справедливой критике других ученых, например, когда я вляпался в переподгонку (об этом феномене мы поговорим в четвертой главе). См. Bailey D. H., Littlefield A. K. *Does Reading Cause Later Intelligence? Accounting for Stability in Models of Change*. *Child Development*. 88, no. 6 (2017): 1913–21. Я даже опубликовал статью о гене-кандидате, используя метод, по которому мы пройдемся в пятой главе. См. Ritchie S. J. et al. *Alcohol Consumption and Lifetime Change in Cognitive Ability: A Gene × Environment Interaction Study*. *AGE*. 36, no. 3 (2014): 9638. И почти наверняка я виновен в подверженности хайпу: в некоторых беседах с журналистами о науке я позволял себе слишком вольные формулировки или позже – все мы крепки задним умом – сожалел, что не добавил какие-то ценные замечания и не сделал важные оговорки. И я совершил ошибку, утверждая, что “опубликованы сотни рецензированных статей на эту тему”, как будто это служит индикатором правды. И что касается рецензирования: уж точно были случаи, когда я не посвящал статье, которую рецензировал, должного количества времени и мог нечаянно пропустить какие-то ошибки. Нисколько не сомневаюсь, что в будущем вскроются еще какие-нибудь мои оплошности или упущения.

ми – подходы, которые многие другие дисциплины начинают заимствовать.

Первый шаг в починке нашей сломанной научной системы – научиться распознавать и исправлять те ошибки, что могут сбить ее с пути. А сделать это возможно только с помощью опять же науки. В книге я постоянно буду прибегать к *метанауке* – сравнительно новому типу научного исследования, в котором изучается само научное исследование. Если наука – это процесс выявления и устранения ошибок, то метанаука представляет собой тот же процесс, но нацеленный на внутренние проблемы.

Ошибки многому могут научить. Музыкант Тодд Рандгрэн во вступительном треке одного из своих альбомов предлагает слушателям сыграть в игру под названием “Студийные звуки”. Он описывает всевозможные огрехи, которые могут проявиться при записи музыки: гудение, шипение, щелчки, когда в микрофон пропевают слова, содержащие букву “п”, прерывистый монтаж и тому подобное. Рандгрэн предлагает слушателям считывать такие шероховатости при прослушивании его песен в этом альбоме и любых других записей. И как понимание огрехов звукозаписывающих студий может по-новому раскрыть для вас процесс создания музыки, точно так же понимание того, как наука отклоняется от верного пути, может многое сказать о процессе, благодаря которому мы постигаем новое знание.

Узнав о серьезных проблемах с тем, как мы занимаемся наукой, вы придете в замешательство. Сколько любопытных результатов, о которых вы читали в новостях и научно-популярных книгах или узнавали из документальных фильмов (открытий, настолько вас восхитивших, что вы поделились ими с друзьями, или заставивших вас переосмыслить устройство нашего мира), основаны на слабых исследованиях, которые нельзя воспроизвести? Сколько раз врач прописывал вам лекарство или метод лечения, сведения об эффективности которого зиждутся на ущербных данных? Сколько раз вы меняли свой рацион, покупательские привычки или какой-то другой аспект образа жизни, руководствуясь научным исследованием, только чтобы убедиться, что его результаты через несколько месяцев будут полностью опровергнуты в новом исследовании? Сколько раз государственные деятели устанавливали законы или выработывали политику, напрямую влияющую на жизнь людей, ссылаясь на научные результаты, не выдерживающие критики? Ответ на каждый из этих вопросов один: это происходило гораздо чаще, чем вам хотелось бы думать.

Наивно рассчитывать, что абсолютно каждое научное исследование будет истинным – неким отчетом о незыблемых фактах, который никогда не подвергнется пересмотру в будущих исследованиях. Мир для этого чересчур беспорядочен. Надеяться можно только на то, что наши научные исследования заслуживают доверия, то есть честно сообщают, что

произошло в экспериментах. Если хваленая система рецензирования не способна оправдать это доверие, наука утрачивает одно из своих основных и наиболее ценных качеств, наряду со способностью делать то, что она делает лучше всего: коренным образом менять наш мир благодаря постепенному появлению и развитию новых знаний, технологий, методик и средств лечения.

Моя цель – восхвалять науку, а не хоронить ее. В этой книге я отнюдь не нападаю на саму науку или ее методы. Напротив, я защищаю научные методы и вообще научные принципы от того, как сейчас занимаются наукой. Беды, о которых мы будем говорить, так тревожат именно из-за важности науки: позволяя ей померкнуть, а прогрессу страшно забуксовать, мы подвергаем себя риску разрушить одно из величайших достижений нашего вида.

Однако нанесенный ущерб поправим. Теоретически – а возможно, и в практическом плане – наука все еще способна быть функциональной и надежной системой знаний, какой она нам и нужна. Пока мы будем обсуждать в этой книге череду научных провалов, стоит держать в голове позитивную мысль (дрожащий проблеск надежды и утешения, что вырывается из ящика Пандоры, наполненного мошенничеством, предвзятостью, недобросовестностью и хайпом, который мы вскроем в следующих главах): почти все описываемые проблемы были обнаружены *другими учеными*. Умные идеи метанауки, предложенные для борьбы с этими проблемами и

наведенным беспорядком, в значительной мере зародились внутри научного сообщества. Пусть во многих областях он и погребен глубоко, но дух самокритики, вдыхающий жизнь в подлинную науку, еще теплится.

И хорошо, что так, поскольку, как мы вскоре увидим, в науке царит изрядный бардак.

Часть I. Как должно быть – и как на самом деле

Глава 1. Как работает наука

Для того, чтобы ум мог заниматься такими предметами мышления надлежащим образом, недостаточно уединения, но требуется компания наших ближних и беседа с ними²⁰.

Дэвид Юм “О написании эссе” (1777)²¹

Наука – это социальный конструкт.

И прежде чем из-за этой фразы вы отшвырнете книгу, позвольте объяснить, что я имею в виду. Я употребляю это выражение не в том смысле, в каком его используют крайние релятивисты, постмодернисты, поборники антинауки и прочие, кто полагает, что реального мира не существует, что наука – лишь один не так чтобы особенный способ узнать об этом или даже что наука – это просто один из “мифов” сре-

²⁰ Перевод О. Артемьевой. Цит. по: *Своеволие философии: собрание философских эссе*. М.: ЯСК, 2019.

²¹ Hume D. *Of Essay-Writing // Essays: Moral, Political, and Literary*. Indianapolis: Liberty Fund, 1777.

ди многих других, в которые мы могли решить верить²². Наука вылечивала болезни, картировала мозг, прогнозировала климатические условия, расщепила атом – это лучший из имеющихся у нас методов для выяснения того, как работает вселенная и как подчинить ее нашей воле. Другими словами, наука – это лучший способ приближения к истине. Разумеется, мы никогда не достигнем конца пути – достаточно беглого взгляда на историю, чтобы увидеть, сколь высокомерно было бы утверждать, будто какие-либо факты неопровержимы и неизменны. И все же для постепенного продвижения к большему знанию о мире лучше научных методов нет ничего.

Однако одними этими методами нам прогресса не достичь. Недостаточно в одиночку сделать какое-то наблюдение в своей лаборатории, вы должны еще убедить других ученых в том, что открыли нечто реальное. И вот тут-то и начинается социальная часть. Философы уже долго обсуждают, насколько важно для ученых *показывать* коллегам-исследователям, как они пришли к своим выводам. Вот как это сформулировал Джон Стюарт Милль²³:

Даже при изучении природы, и здесь всегда возможны

²² Sokal A., Bricmont J. *Intellectual Impostures*. London: Profile Books, 1998, 2003.

²³ Mill J. S. *On Liberty*. London: Dover Press, 1859. [Милль Дж. С. *О свободе // Антология мировой либеральной мысли (1 половина XX века)*. М.: Прогресс-Традиция, 2000.]

различные объяснения одних и тех же фактов, возможна теория геоцентрическая и теория гелиоцентрическая, возможна и теория флогистона, и теория кислорода, – и чтобы признать какую-нибудь из них истинной, надо доказать, что другая не есть истина, а пока это не доказано или пока мы не знаем, как это доказывается, то, признавая одну из них истинной, не знаем, значит, оснований мнения, которого держимся²⁴.

Поэтому ученые работают в командах, путешествуют по миру с лекциями и докладами на конференциях, дискутируют друг с другом на семинарах, организуют научные общества для обсуждения результатов исследований и – что, пожалуй, самое важное – публикуют свои результаты в рецензируемых журналах. Эти социальные аспекты – не преимущества профессии и не просто товарищество. Они представляют собой научный процесс в действии – непрерывную череду коллективных актов: тщательно изучить, поставить под сомнение, проверить, уточнить и выработать *консенсус*. Хотя это и кажется на первый взгляд парадоксальным, но именно субъективный процесс науки и порождает ее беспримерную степень объективности²⁵.

²⁴ Перевод А. Неведомского.

²⁵ Longino H. E. *Science as Social Knowledge*. Princeton: Princeton University Press, 1990. См. также Longino H. *The Social Dimensions of Scientific Knowledge*. The Stanford Encyclopedia of Philosophy. Summer 2019; Reiss J., Sprenger J. *Scientific Objectivity*. The Stanford Encyclopedia of Philosophy. Winter 2017.

Вот в каком смысле наука является социальным конструктом. Всякое утверждение о нашем мире может быть названо научным знанием только после того, как оно пройдет эту общественную проработку, которая призвана отсеивать ошибки и упущения и давать возможность другим ученым судить, достойно ли новое открытие звания надежного, достоверного и важного. Благодаря тому, что каждое открытие прогоняют через такую полосу препятствий, конечные продукты научного процесса – публикуемые рецензируемые статьи – наливаются порядочной силой для общества. Мы говорим: это *наука*, а не просто шаблонная фраза, или болтовня, или чье-то мнение.

Социальная природа науки вместе с тем имеет и слабые места. Поскольку ученые так сильно сосредоточены на попытках убедить своих коллег, что подразумевает прохождение результатов исследования через этап рецензирования и дальнейшие шаги вплоть до публикации, им очень просто позабыть о настоящей цели науки – приближать нас к истине. А поскольку ученые еще и человеческие существа, способы, которыми они стараются убеждать друг друга, не всегда в полной мере рациональны или беспристрастны²⁶. Если

²⁶ Это утверждение я делаю под влиянием идеи специалистов по эволюционной теории Уго Мерсье и Дэна Спербера, согласно которой основная функция человеческого мышления как такового состоит в том, чтобы придумать, как лучше всего убедить других людей. См. Mercier H., Sperber D. *Why Do Humans Reason? Arguments for an Argumentative Theory*. Behavioral and Brain Sciences. 34, no. 2 (2011): 57–74.

мы не будем особенно внимательны, научный процесс может пропитаться несовершенствами самого человека.

Эта книга посвящена тому, как мы уделяли недостаточно внимания научному процессу. Тому, как мы в итоге очутились в ситуации, когда научная система не просто не учитывает наши человеческие слабости, но еще и усиливает их. В последние годы стало донельзя очевидно, что рецензирование отнюдь не гарантирует корректности и надежности, как ему положено, а публикационный процесс, который должен быть оплотом науки, превратился в ее ахиллесову пяту.

Однако чтобы понять, как же научная публикационная система так испортилась, сначала нам нужно узнать, как ей надлежит работать в отлаженном состоянии.

Представим, что вы хотите заняться наукой. Первым делом вам придется проштудировать научную литературу. Она представляет собой обширнейшую коллекцию журналов для специалистов, изданий, которые служат главными источниками нового научного знания. Идея периодических изданий, где ученые могли бы делиться результатами своей работы, восходит к 1665 году, когда Генри Ольденбург из Королевского общества Великобритании опубликовал первый номер журнала, приведем его полное заглавие, *Philosophical Transactions: Giving Some Account of the Present Undertakings, Studies, and Labours of the Ingenious in Many Considerable Parts of the World* (“Философские труды, дающие некоторый отчет

о нынешних начинаниях, исследованиях и трудах гениальных людей во многих значительных частях мира”)²⁷. Идея состояла в том, чтобы эти гениальные ученые могли присылать письма с описанием своих деяний для ознакомления заинтересованных читателей. До того ученые либо трудились в одиночку в угоду богатым правителям или частным покровителям и организациям (и их научные изыскания часто воспринимались скорее как некие салонные фокусы, чем как попытка найти истину) и издавали отдельные книги, либо формировали кружки единомышленников, которые обменивались письмами. Из такого вот клуба по переписке и выросли организации вроде Королевского общества²⁸.

Первые номера журнала Ольденбурга больше напоминали

²⁷ McDougall-Waters J. et al. *Philosophical Transactions: 350 Years of Publishing at the Royal Society (1665–2015)*. London: Royal Society, 2015. Некоторые историки возразили бы, что первым научным журналом следует считать французское издание *Le Journal des sçavans*, вышедшее в 1665 году всего за два месяца до *Philosophical Transactions*. Однако в нем публиковались статьи по огромному числу разных ученых тем и поначалу – в основном книжные рецензии и отрывки. Тогда как *Philosophical Transactions* исходно был посвящен главным образом научным новостям и наблюдениям. Наверное, справедливее было бы признать *Le Journal des sçavans* первым академическим изданием, а *Philosophical Transactions* – первым научным. См. McCutcheon R. P. *The “Journal Des Scavans” and the “Philosophical Transactions of the Royal Society”*. *Studies in Philology*. 21, no. 4 (1924): 626–8; Banks D. *Thoughts on Publishing the Research Article over the Centuries*. *Publications*. 6, no. 1 (2018): 10.

²⁸ David P. A. *The Historical Origins of “Open Science”: An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution*. *Capitalism and Society*. 3, no. 2 (2008): 5.

своеобразный вестник, где описывались недавние эксперименты и открытия. Скажем, в первом выпуске первого тома натурфилософ и энциклопедист Роберт Гук рассказал о самом первом наблюдении предположительно Большого красного пятна на Юпитере. Заметка вся целиком была такой:

Гениальный мистер Гук несколькими месяцами ранее поставил своего друга в известность о том, что несколько дней назад, как он сказал (а именно 9 мая 1664 года, около девяти часов вечера), он наблюдал в великолепный двенадцатифутовый телескоп маленькое пятно на наибольшем из трех более темных поясов Юпитера и что, время от времени поглядывая на него, обнаружил, что за два часа оно переместилось с востока на запад примерно на половину диаметра Юпитера²⁹.

Журнал существует и по сей день, только под несколько более удобоваримым названием *Philosophical Transactions of the Royal Society* (“Философские труды Королевского общества”)³⁰. Время шло, и на смену коротким новостным заметкам пришли статьи подлиннее, содержащие подробные описания экспериментов и исследований. Сейчас это изда-

²⁹ Hooke R. *A Spot in One of the Belts of Jupiter*. *Philosophical Transactions*. 1 (1665): 3.

³⁰ В 1900 году он был разделен на два журнала: один по математике и физическим наукам, другой по наукам биологическим. См. <https://royalsocietypublishing.org/journal/rstl>.

ние входит в глобальную экосистему из более чем тридцати тысяч журналов, от самых общих (как высокопрестижные *Nature* и *Science*, которые публикуют результаты самых значительных исследований со всего мира, относящихся к любой научной области) до узкоспециальных (вроде *American Journal of Potato Research*, интересующегося статьями исключительно на картофельную тему)³¹. Некоторые журналы издаются научными обществами, как – до сих пор – *Philosophical Transactions*, но большинство принадлежат коммерческим компаниям, таким как *Elsevier*, *Wiley* и *Springer Nature*³². Благодаря недавнему прогрессу все научные журналы были переведены в электронный вид, так что любой человек, который может заплатить взносы, взимаемые издателем с подписчиков, либо препоручить это библиотеке своего университета, овладевает мировым научным знанием, находящимся теперь всегда у него под рукой³³.

После прочтения журналов, относящихся к вашей области интересов, вы можете определиться с предметом иссле-

³¹ Ware M., Mabe M. *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*. The Hague, Netherlands: International Association of Scientific, Technical and Medical Publishers, 2015.

³² Обратите внимание, что до середины XVIII века, когда журнал *Philosophical Transactions* стал официальным изданием Королевского общества, он издавался разными отдельными исследователями и составителями.

³³ Большинство журнальных статей – они называются “экспериментальными” – сообщают о новых результатах исследований, но некоторые являются “обзорными” статьями, обобщающими все, что уже известно по какому-то конкретному научному вопросу.

дования. Например, есть научная теория, в рамках которой делается некое предсказание – формулируется гипотеза, а вы можете каким-нибудь хитрым способом ее проверить. Или вы знаете, что в имеющихся у ученых знаниях есть пробел, а вы понимаете, как его заполнить. Или в порыве вдохновения вы придумали эксперимент, который позволит открыть нечто совершенно новое. Правда, прежде чем приняться за воплощение своей идеи, обычно требуется раздобыть денег на финансирование исследования: чтобы, к примеру, купить оборудование или материалы, привлечь участников или оплатить труд ученых, нанятых для сбора данных. Если вы, скажем, не фармацевтическая компания, которая может позволить себе держать собственные лаборатории, основной способ получить жизненно важное финансирование – это подать заявку на грант. Гранты предоставляются правительством страны, бизнес-компаниями, дотационными фондами, некоммерческими организациями, благотворительными фондами или даже просто состоятельными физическими лицами. Вы можете подать заявку в Национальные институты здравоохранения (НИИ) или в Национальный научный фонд (это организации в США, финансируемые из государственного бюджета) либо в благотворительный фонд, поддерживающий науку, например *Wellcome Trust* или Фонд Билла и Мелинды Гейтс³⁴.

³⁴ www.nih.gov и www.nsf.gov. Похожие организации в других странах: Государственное агентство исследований и инноваций Великобритании (*UK*

Финансирование отнюдь не гарантировано, и любой ученый подтвердит, что один из самых выматывающих аспектов научной работы – это поиск источников финансирования для своих новых исследований, а вероятность получить отказ жестоко высока. Вечные попытки найти деньги оказывают важное косвенное влияние на саму науку, и мы еще к этому вернемся. Однако сейчас давайте представим, что вам повезло получить грант. Тогда вы приступаете к работе. На этапе сбора данных может потребоваться сталкивать между собой частицы в подземном коллайдере, искать окаменелые остатки в арктической зоне Канады, воссоздавать в чашке Петри точные условия для роста бактерий, организовывать посещение лаборатории и заполнение опросников для сотен людей, прогонять сложную математическую модель на компьютере. Этот этап может занять дни, месяцы, десятилетия.

Когда данные уже получены, вы, как это обычно бывает, располагаете набором чисел, который можете – сами или с помощью более подкованного в математике коллеги – проанализировать различными статистическими методами (это еще одно опасное место, куда мы позже вернемся). Затем вы должны все это описать в форме научной статьи. Типичная статья начинается с раздела “Введение”, где вы кратко изла-

Research and Innovation; www.ukri.org), Национальный фонд естественных наук Китая (National Natural Science Foundation of China; www.nsf.gov.cn/english/site_1/index.html), Японское общество содействия науке (Japan Society for the Promotion of Science; www.jsps.go.jp/english). См. также <https://wellcome.ac.uk/> и www.gatesfoundation.org.

гаете, что уже известно по этой теме и что добавляет ваше исследование. Далее следует раздел “Методы”, где вы описываете, что конкретно делалось в ходе исследования, причем настолько подробно, чтобы любой человек мог – теоретически – снова провести точно такой же эксперимент. Потом вы переходите к разделу “Результаты”, где в подтверждение своих открытий демонстрируете числа, таблицы, графики и статистический анализ. А венчает статью раздел “Обсуждение”, где вы свободно строите спекулятивные теории – то есть вдумчиво и на основе фактов рассуждаете – о том, что все это значит. Статью вы предваряете разделом “Аннотация” – кратким, обычно около ста пятидесяти слов, описанием исследования и его результатов. Аннотация всегда доступна для прочтения кем угодно, даже если доступ к самой статье ограничен подпиской на журнал, поэтому имеет смысл подходить к написанию аннотации со всем тщанием, чтобы полученные вами результаты звучали убедительно. Научные статьи бывают самой разной длины и объема, и иногда описанная очередность разделов нарушается, но в целом ваша статья будет выстроена по этому принципу³⁵.

Когда статья написана, вы вступаете в мир научных журналов и борьбы за публикацию. Еще недавно для того, чтобы подать статью в журнал, нужно было напечатать несколь-

³⁵ Например, в некоторых научных журналах раздел “Методы” помещается в самый конец статьи, будто бы эта важнейшая информация – всего лишь второстепенное дополнение.

ко экземпляров и отправить по почте редактору, но теперь все осуществляется через интернет (хотя у многих журналов до сих пор столь допотопные и ненадежные веб-страницы, что вы с таким же успехом могли бы отправить им статью с почтовым голубем). Редактор журнала – зачастую это солидный ученый – прочитает вашу статью (или, что не исключено, будем уж честны, только аннотацию) и решит, достойна ли она того, чтобы ее опубликовать. Бóльшая часть журналов, особенно самые престижные, кичатся своей исключительностью, а следовательно, и высоким порогом попадания к ним на страницы (*Science*, например, принимает менее 7 % статей, поданных на рассмотрение), так что большинство статей на этой стадии будет отклонено без направления на рецензирование³⁶. Это первый этап контроля качества: редактор отделяет статьи, соответствующие темам журнала и перспективные с точки зрения их научного значения и качества, от тех, которые не стоят дальнейшего рассмотрения. Для доли статей, что приглянулись редактору, наступает этап рецензирования. Редактор находит двух-трех ученых, являющихся экспертами в вашей области исследования, и спрашивает, не согласятся ли они оценить вашу статью. Они могут отказаться, сославшись на занятость, так что редактор продолжит идти по списку потенциальных рецензентов до тех пор, пока несколько из них не дадут согласие. И вот тут наступает этап нервного ожидания: одобряют ли они вашу ра-

³⁶ www.sciencemag.org/site/feature/contribinfo/faq/index.xhtml#pct_faq

боту.

Большинство людей, включая ученых, полагают, что рецензирование всегда было ключевым моментом для научной публикации, однако его история сложнее. Хотя в XVII веке Королевское общество стремилось спрашивать кого-то из своих членов, считают ли они некую статью достаточно интересной для публикации в *Philosophical Transactions*, по крайней мере до 1831 года не требовалось давать письменную оценку каждой работе³⁷. И даже тогда формальная система рецензирования, какой мы ее знаем сегодня, не была универсальной, она стала таковой лишь в XX веке, да и то далеко не сразу (как можно понять из письма Альберта Эйнштейна, которое он послал в 1936 году редакторам журнала *Physical Review*, раздраженно заявляя, что снимает свою статью с рассмотрения в их журнале, поскольку они посмели отправить ее другому физику для оценки)³⁸. Только в 1970-х годах уже все журналы начали применять современную практику отправки поданных на рассмотрение статей независимым экспертам на рецензирование, присвоив им роль фильтра, какую они и играют сегодня³⁹.

Обычно рецензенты выполняют свою работу анонимно,

³⁷ Csiszar A. *Peer Review: Troubled from the Start*. Nature. 532, no. 7599 (2016): 306–8.

³⁸ Цитируется по: Baldwin M. *Scientific Autonomy, Public Accountability and the Rise of "Peer Review" in the Cold War United States*. Isis. 109, no. 3 (2018): 538–58.

³⁹ Там же.

что одновременно и благословение, и проклятие. Благословение – потому что анонимность позволяет им без обиняков высказывать свою точку зрения, не волнуясь насчет ответной реакции ученых, чью работу они критикуют (зеленый научный сотрудник может действительно откровенно высказаться о недостатках работы именитого профессора), а проклятие – поскольку анонимность, что уж тут, позволяет им без обиняков высказывать свою точку зрения, не волнуясь насчет ответной реакции ученых, чью работу они критикуют. Вот несколько выдержек из настоящих рецензий:

- Некоторые статьи читать одно удовольствие. Эта явно не из таких.
- Результаты слабые, какая-то размазня.
- Боюсь, эта статья послужит не столько прогрессу в данной области, сколько ее окончательному упадку.
- В рукописи делается три утверждения: первое известно уже годы, второе – десятки лет, третье – столетия.
- Когда вы писали это предложение, у вас не случилось припадка? Я вот бился в конвульсиях, пока читал его⁴⁰.

Если оценки рецензентов выглядят подобным образом, редактор, вероятно, отклонит вашу статью. Тогда вам захочется либо отступить, либо начать весь процесс заново, подав статью на рассмотрение в другой журнал, а если и туда

⁴⁰ <https://shitmyreviewerssay.tumblr.com>

не примут, еще в один, а если снова не примут, в следующий и так далее – нередко статьи проходят через полдюжины или больше журналов, обычно со все более низким рейтингом, прежде чем их примут к публикации. Если же рецензенты впечатлятся больше, вы получите шанс доработать свою статью, учтя их замечания, например заново проанализировав данные, проведя дополнительные эксперименты или перепишав какие-то разделы, и снова отправить ее редактору. Такая пересылка статьи туда-сюда в ходе ее доработки может иметь несколько итераций и часто длится месяцами. В конце концов, если рецензенты оказываются удовлетворены, редактор дает добро – и статья публикуется. Если журнал до сих пор выходит и в бумажной форме, вам доведется увидеть свою драгоценную статью в напечатанном виде, а иначе придется довольствоваться тем, чтобы с трепетом ждать ее появления на официальном сайте журнала. Вот и все. Вы оставили свой след в научной литературе, и у вас теперь есть публикация, которую вы можете добавить в свое резюме и на которую другие ученые могут ссылаться. Поздравляю, до конца дня можете отдыхать.

Описание всего процесса, данное мной выше, разумеется, слишком краткое и общее, но по существу такой процесс в той или иной форме присущ каждой научной области. Мы должны спросить себя: после того, как статья вышла из-под катка рецензирования, в окончательном своем виде она все еще дает правдивое представление о том, что делалось в ис-

следовании? Мы вернемся к этому вопросу позже. А пока нужно обсудить кое-что другое. Где гарантии, что участники описанного процесса – исследователь, подающий статью на рассмотрение, редактор журнала, рецензенты, оценивающие работу, – все проявляют честность и беспристрастность, каких требует истинная наука? Нет такого закона, который требовал бы, чтобы все руководствовались объективностью и рациональностью, когда имеют дело с наукой, поэтому требуется коллективный моральный устав, свод правил, направляющих поведение ученых⁴¹. Самая известная попытка записать эти неписанные правила была предпринята социологом Робертом Мертоном.

В 1942 году Мертон выдвинул четыре научных ценности, сегодня известных как “мертоновские нормы”. Называются они не слишком оригинально, зато служат правильными ориентирами для ученых. Первая – *универсализм*. Научное знание есть научное знание независимо от того, кто его формулирует, – при условии, что методы обретения этого знания надежны. Раса, пол, возраст, гендер, сексуальная ориентация, уровень доходов, социальное происхождение, национальность, известность или любая другая характеристика

⁴¹ Стоит отметить, что эти правила касаются научного поиска и анализа, они не имеют отношения к этическим нормам, которые все ученые также должны принимать во внимание. Эти правила, пожалуй, особенно важны для тех, кто проводит исследования на людях (или других животных), а еще для тех, кто работает с потенциально опасными технологиями или чьи эксперименты способны нанести экологический либо иной вред.

ученого не должна отражаться на том, как оцениваются его фактологические утверждения. Еще нельзя судить о чем-либо исследовании на основании того, приятен или неприятен этот человек (к облегчению некоторых из моих менее приветливых коллег). Вторая похожая норма – *бескорыстность*. Ученые не должны заниматься наукой из-за денег, по политическим или идеологическим причинам, ради славы (или репутации своего университета, страны и чего-либо еще) или чтобы потешить свое самолюбие. Они должны заниматься наукой во имя лучшего понимания вселенной, делая открытия и созидая, – и точка⁴². Как написал однажды Чарльз Дарвин, у ученого “не должно быть ни желаний, ни привязанностей – лишь каменное сердце”⁴³.

Оставшиеся две нормы напоминают нам о социальной природе науки. Третья – *коллективизм*. Она предписывает ученым делиться знаниями друг с другом⁴⁴. Этот принцип лежит в основе самой идеи публикации ваших результатов в журнале, с тем чтобы о них узнали другие, – мы занимаем-

⁴² Merton R. K. *The Normative Structure of Science* (1942) // *The Sociology of Science: Empirical and Theoretical Investigations*. Chicago and London: University of Chicago Press, 1973.

⁴³ Darwin Correspondence Project, Letter no. 2122, 9 July 1857. www.darwinproject.ac.uk/letter/DCP-LETT-2122.xml

⁴⁴ Вообще Мертон назвал коллективизм “коммунизмом”, но этот термин, скажем так, имеет иные коннотации. В последующих работах название сменилось на “коллективизм”, им я здесь и пользуюсь. См., например: Anderson M. S. et al. *Extending the Mertonian Norms: Scientists' Subscription to Norms of Research*. Journal of Higher Education. 81, no. 3 (2010): 366–93.

ся наукой сообща, мы должны знать детали работы других ученых, чтобы иметь возможность проанализировать и развить ее⁴⁵. И наконец, последняя норма: *организованный скептицизм*. Ничто не свято, и научное утверждение никогда не должно приниматься на веру. Нам следует воздерживаться от вынесения оценки любому открытию, пока мы тщательнейшим образом не проверили все данные и методы исследования. Самое очевидное воплощение нормы организованного скептицизма – рецензирование как таковое.

В теории все выглядит неплохо: придерживаясь четырех мертоновских норм, мы должны прийти к научной литературе, которой можно доверять, – к плечам гигантов, стоя на которых, по знаменитому выражению Ньютона, мы видим дальше. Разумеется, эти гиганты часто ошибались, взять хотя бы два примера Джона Стюарта Милля, процитированных выше: когда-то мы верили, что Солнце обращается вокруг Земли и что горючие вещества наполнены особой субстанцией – флогистоном, – высвобождающейся при их горении⁴⁶. Но в итоге мы отбросили эти теории, когда получили новые данные. И действительно, для ученого готовность изменить

⁴⁵ Мертон упоминает нелюдимого Генри Кавендиша, физика и химика, жившего в XVIII веке, как исторического нарушителя этой нормы: тот скрыл от мира множество своих важных экспериментов и теорий из чистой застенчивости, и лишь много позже его смерти эти результаты были получены заново.

⁴⁶ Best N. W. *Lavoisier's "Reflections on Phlogiston" I: Against Phlogiston Theory*. *Foundations of Chemistry*. 17, no. 2 (2015): 137–51.

свое мнение – это достоинство. Биолог Ричард Докинз описал случай, которому стал свидетелем лично, это произошло с “уважаемым престарелым мужем” с факультета зоологии Оксфордского университета⁴⁷.

Многие годы он страстно верил и учил студентов, что аппарат Гольджи (микроскопическая внутриклеточная структура) на самом деле не существует, что это – погрешность наблюдения, иллюзия. Каждый понедельник после обеда на факультете было заведено слушать научный доклад какого-нибудь заезжего лектора. В один из понедельников лектором оказался американский специалист по биологии клетки, представивший неотразимо убедительные свидетельства реальности аппарата Гольджи. В конце его выступления старик пробрался к подиуму и, пожимая американцу руку, с чувством провозгласил: “Дорогой коллега, позвольте выразить вам мою благодарность. Все эти пятнадцать лет я заблуждался”. Мы тогда аплодировали до боли в ладонях. <...> Не каждый ученый в реальной жизни [смог бы произнести такое]. Но для всех ученых подобные поступки являются эталоном – в отличие, скажем, от политиков, которые могли бы счесть старика беспринципным. У меня до сих пор комок к горлу подступает при воспоминании о том вечере⁴⁸.

⁴⁷ Dawkins R. *The God Delusion*. London: Bantam Books, 2006. [Докинз Р. *Бог как иллюзия*. М.: КоЛибри, 2018.]

⁴⁸ Перевод Н. Смелковой.

Вот что имеется в виду, когда говорят, что наука – это процесс “самокорректирующийся”. В конце концов, даже если это занимает много лет или десятилетий, старые неверные идеи ниспровергаются благодаря новым данным (а иногда, по довольно мрачному выражению физика Макса Планка, благодаря тому, что непреклонные поборники этих старых идей вымирают, оставляя науку следующему поколению⁴⁹). И опять: так дело обстоит в теории. На практике же публикационная система, описанная выше, опирается на мертоновские нормы неуклюже, во многом препятствуя процессу самокорректировки. Характерные особенности этого противоречия – между конкуренцией за гранты и шумихой вокруг престижных публикаций с одной стороны и открытой, бесстрастной, скептической оценкой науки с другой – будут становиться все очевиднее в последующих главах.

А пока отметьте, что именно заставило престарелого мужа из рассказа Докинза изменить свое мнение – “неотразимо убедительные свидетельства”. Едва ли имеет смысл пытаться скорректировать или уточнить наши научные теории в соответствии с новыми данными, если сами данные неубедительны, а еще хуже – если они вообще неверны. Это возвращает нас к идее, которую мы обсуждали в предисловии:

⁴⁹ Planck M. *Scientific Autobiography and Other Papers*. London: Williams & Norgate, Ltd., 1949. [См., например: Планк М. *Научная автобиография*. УФН. 64, № 4 (1958): 625–37.]

чтобы результаты заслуживали нашего доверия, они должны быть воспроизводимы. Вот как это сформулировал философ науки сэр Карл Поппер⁵⁰:

Только тогда, когда некоторые события повторяются в соответствии с некоторыми правилами и регулярностями (как в случае воспроизводимых экспериментов), наши наблюдения в принципе могут быть проверены каждым человеком. Даже наши собственные наблюдения мы не принимаем всерьез и не приписываем им статус научных наблюдений до тех пор, пока не повторим и тем самым не проверим их. Только в результате подобных повторений мы можем убедить себя в том, что имеем дело не с простым “совпадением”...⁵¹

Эта идея не то чтобы революционна – или была нова для Поппера, написавшего эти слова в 1950-х годах. Вернувшись в XVII век, когда журнал *Philosophical Transactions* только зародился, мы обнаружим, что сооснователь Королевского общества, химик Роберт Бойль, делал все возможное и невозможное, чтобы убедиться в воспроизводимости своих результатов. Он многократно демонстрировал эксперименты по установлению различных свойств воздуха и вакуума с ис-

⁵⁰ Popper K. *The Logic of Scientific Discovery*. London & New York: Routledge Classics, 1959/2002. [Поппер К. *Логика научного исследования // Поппер К. Логика и рост научного знания*. М.: Прогресс, 1983.]

⁵¹ Перевод В. Брюшинкина.

пользованием своего знаменитого воздушного насоса перед толпами зевак, а после заставлял их подписывать свидетельские показания о том, что они действительно наблюдали демонстрировавшийся процесс⁵². Он следил, чтобы его труды были написаны достаточно обстоятельно, дабы “человек, кому я их адресовал, сумел без ошибок и особых помех повторить столь необычные эксперименты”⁵³. И хотя соорудить сложный аппарат было весьма непросто, он призывал других натурфилософов повторить его эксперименты с воздушным насосом в других частях Англии и Европы и оказывал им в том содействие⁵⁴.

Таким образом, воспроизведение экспериментов долгое время было ключевым ингредиентом рецепта, по которому должна была работать наука (и кстати, это еще один ее социальный аспект: результаты воспринимаются всерьез только после того, как окажутся подтверждены несколькими исследователями). Но с тех пор, где-то между Бойлем и современным академическим сообществом, громадное число

⁵² Подобное повторение опыта, хоть и проведенное лет на сто позже, изображено – в драматическом свете – на картине Джозефа Райта (Райта из Дерби) “Эксперимент с птицей в воздушном насосе”, сейчас находящейся в Лондонской национальной галерее.

⁵³ Boyle R. *The New Experiments Physico-Mechanical, Touching the Spring of the Air and Its Effects*. London: Miles Flesher, 1682. Цит. по: Shapin S., Schaffer S. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton: Princeton University Press, 1985.

⁵⁴ Shapin S., Schaffer S. *Leviathan*.

ученых позабыли о важности повторений. В столкновении наших мертоновских идеалов с реалиями научной публикационной системы – не говоря уже о реалиях человеческой природы – идеалы оказались более хрупкими, оставив нас с научной литературой, полной ненадежных, недостоверных и невоспроизводимых результатов, которые зачастую больше путают, чем обогащают новым знанием.

В следующей главе мы узнаем, *насколько* научная литература запружена ненадежными, недостоверными и невоспроизводимыми результатами.

Глава 2. Кризис воспроизводимости

*Взовется на дыбы желание власти
И валится, споткнувшись, в тот же миг⁵⁵.
Уильям Шекспир “Макбет”, акт 1, сцена 7*

*“Опубликовано” и “достоверно” – не синонимы.
Райан Носек, Джефффри Спайс и Мэтт Мотел⁵⁶*

Несомненно, самая популярная книга по психологии за минувшее десятилетие – это “Думай медленно... решай быстро” Даниэля Канемана. Не так много существует руководств по человеческому разуму, превосходящих канемановское. Он получил Нобелевскую премию по экономике 2002 года за свои исследования человеческой (ир)рациональности и опубликовал результаты десятков искусных экспериментов, демонстрирующих ограничения нашей способности мыслить. Книга “Думай медленно... решай быстро” стала сенсацией, суммарные продажи исчислялись миллионами экземпляров, и она до сих пор недурно расходуется. И на то есть причины: это живо и доступно написанный путеводитель по всем ошибкам и искажениям человеческого

⁵⁵ Перевод М. Лозинского.

⁵⁶ Nosek B. A. et al. *Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability*. *Perspectives on Psychological Science*. 7, no. 6 (2012): 615–31.

мышления⁵⁷. Помимо многих других тем Канеман осветил работы по “праймингу”, как называют это явление психологи. Некоторые примеры прайминга (эффекта предшествования) имеют отношение к языку. Например, установлено, что если я показываю вам на экране компьютера набор слов по одному и прошу нажимать на клавишу всякий раз, как появляется слово “ложка”, то вы будете реагировать чуточку быстрее, если перед тем высвечивалось слово “вилка” (или какой-то еще из столовых приборов), чем когда предшествующим словом было “дерево” (или что-то еще не относящееся к принадлежностям для приема пищи). Восприятие слова “вилка” психологически “настраивает” вас быстрее реагировать на близкое по смыслу слово⁵⁸.

Канеман, однако, описал нечто более неожиданное. Он осветил результаты исследования по социальной психологии, демонстрирующие, что формирование установок, связанных с определенными *понятиями*, – обычно неосознанное – может ощутимо влиять на наше поведение. Один пример известен как “эффект леди Макбет”. В 2006 году в журнале *Science* вышла статья о том, что, когда исследователи просили участников эксперимента переписать текст об аморальных поступках, те потом с большей вероятностью хоте-

⁵⁷ Kahneman D. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011. [Канеман Д. *Думай медленно... решай быстро*. М.: АСТ, 2013.]

⁵⁸ Neely J. *Semantic Priming Effects in Visual Word Recognition: A Selective Review of Current Findings and Theories // Basic Processes in Reading: Visual Word Recognition*. Abingdon: Routledge, 2012.

ли купить мыло, а просьба вспомнить что-то неприглядное из собственных деяний заставляла их чаще брать при выходе из лаборатории дезинфицирующую салфетку (“Проклятое пятно!”). Здесь воздействие оказывалось без *словесных* установок: получалось, что мозг работает куда более связно и целостно, чем мы полагали, что между понятиями и концепциями, вроде бы связанными очень условно, образуются прочные сшивки. В данном случае это казалось свидетельством некоего глубинного наложения понятий морали и чистоты. Авторы статьи утверждали, что эти результаты могут даже объяснить, почему омовение рук является частью столь многих религиозных ритуалов по всему миру⁵⁹.

Канеман также сделал обзор исследований “денежного прайминга”. В другой статье, вышедшей в *Science* в том же 2006 году, рассказывалось, как социальные психологи обнаружили, что, ненавязчиво напоминая людям о деньгах – скажем, сажая их за стол, где по случайности стоит компьютер с плавающими банкнотами на экранной заставке, – можно побудить их чувствовать и вести себя так, словно они более независимы, и меньше беспокоиться о других⁶⁰. Авторы писали, что участники исследования, подвергшись денежному праймингу, предпочитали “играть в одиночку, работать в

⁵⁹ Zhong C. B., Liljenquist K. *Washing Away Your Sins: Threatened Morality and Physical Cleansing*. *Science*. 313, no. 5792 (2006): 1451–2.

⁶⁰ Vohs K. D. et al. *The Psychological Consequences of Money*. *Science*. 314, no. 5802 (2006): 1154–6.

одинокую и держать большую физическую дистанцию между собой и новым окружением”⁶¹. И действительно, в ответ на просьбу расставить в комнате стулья для личной беседы с незнакомым человеком участники, подвергшиеся денежному праймингу, ставили стулья почти на тридцать сантиметров дальше один от другого по сравнению с участниками эксперимента, которые видели пустой экран компьютера. Нехилый эффект для простой экранной заставки, вероятно, подумали вы. И такой сценарий повторялся в большинстве значимых исследований прайминга: очень тонкие установки вызывали заметные изменения в поведении людей.

Канеман заключил, что подобные исследования прайминга “ставят под угрозу наше восприятие себя как сознательных и независимых творцов своих суждений и выборов”⁶². Он не сомневался в их надежности. И писал: “Не доверять этим данным нельзя. Они не выдумки и не статистические отклонения. Вам придется принять основные выводы исследований за истину. А еще важнее – принять их истинность применительно к *вам*”⁶³.

Но Канеман напрасно столь безоговорочно доверял этим эффектам предшествования, хоть они и были опубликованы в одном из самых уважаемых научных журналов. Оказывается, наряду с разоблачением мошенничества Дидерика Ста-

⁶¹ Там же.

⁶² Перевод “Школы перевода Баканова”.

⁶³ Kahneman D. *Thinking, Fast and Slow*.

пела и публикацией “сверхъестественных” результатов Дэрила Бема именно исследование, посвященное праймингу, – или скорее неудачная попытка его повторить – стало еще одним из исходных стимулов, подхлестнувших изучение явления, известного сегодня как “кризис воспроизводимости”⁶⁴.

В том исследовании прайминга участников просили найти в списке разрозненных слов одно лишнее, так чтобы остальные слова можно было сложить в связное предложение. У одной половины участников лишние слова были случайными и нейтральными, а у другой – имеющими отношение к пожилым людям, например “старый”, “седой”, “мудрый”, “вяжет” и “Флорида” (Флорида известна в Америке как штат, где живет много пенсионеров). Выполнив задание, участники исследования могли уходить – но в тайне от них экспериментаторы замеряли, как быстро те шли по коридору к выходу из здания. Демонстрируя опять-таки ментальную связь между идеями и действиями, те участники, кого подвергли праймингу связанными со старостью словами, *уходили из лаборатории медленнее* по сравнению с контрольной

⁶⁴ Насколько мне известно, термин происходит из статьи Пашлера и Вагенмэйкера, которые не использовали словосочетание “кризис воспроизводимости” напрямую, но говорили о “кризисе доверия” в психологических исследованиях после серии неудавшихся повторений. Нельсон, Симмонс и Саймонсон обсуждали причины возникновения кризиса. См. Pashler H., Wagenmakers E.-J. *Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? Perspectives on Psychological Science*. 7, no. 6 (2012): 528–30; Nelson L. D. et al. *Psychology's Renaissance*. *Annual Review of Psychology*. 69, no. 1 (2018): 511–34.

группой⁶⁵.

Опубликованная в 1996 году, статья об этом эксперименте с тех пор была процитирована другими исследователями более пяти тысяч раз и вошла во все учебники по психологии – помню, как сам студентом изучал ее⁶⁶. Однако в 2012 году независимая группа попыталась провести точно такой же эксперимент, только с большей выборкой и совершеннее с технической точки зрения, – и не выявила никакой разницы в скорости ухода участников. Исследователи предположили, что в исходной работе получились такие результаты, поскольку сотрудники лаборатории, замерявшие время по секундомерам, знали, кто из участников как должен был бы себя вести, и это, вероятно, отражалось на замеряемых величинах. Измерение скорости ухода с помощью инфракрасных лучей, как было сделано в исследовании-повторении, свело к нулю предполагаемый эффект прайминга⁶⁷. В течение нескольких лет другие лаборатории пытались воспроизвести как “эффект леди Макбет”, так и эффект денежного прайминга, тоже с гораздо более внушительной и репрезен-

⁶⁵ Bargh J. A. et al. *Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action*. Journal of Personality and Social Psychology. 71, no. 2 (1996): 230–44.

⁶⁶ Количество цитирований (точное число – 5208) дано по состоянию на январь 2020 года в соответствии с *Google Scholar*.

⁶⁷ Doyen S. et al. *Behavioral Priming: It's All in the Mind, but Whose Mind?* PLOS ONE. 7, no. 1 (2012): e29081.

тативной выборкой⁶⁸. Эти попытки также очевидным образом провалились. Нет причин думать, что разнообразные результаты по праймингу были, процитируем Канемана, “выдумками”, – нужно исходить из предположения, что получены они были добросовестным образом. Но вот “статистические отклонения”? Пожалуй, именно они.

Другие исследования эффекта предшествования были не лучше. В одном утверждалось, что участники, подвергшиеся праймингу “расстоянием” – их попросили нарисовать на листе миллиметровки две удаленные друг от друга точки, – чаще чувствовали себя дистанцированными от друзей и родных; попытка повторить это исследование провалилась в 2012 году⁶⁹. В другом исследовании утверждалось, что когда примеры моральных дилемм распечатывались на листах с окантовкой в шахматную клетку, то участники эксперимента выносили более поляризованные суждения, поскольку этот узор заставлял их думать о концепции “белое и черное”; попытка повторить это исследование провалилась в 2018 го-

⁶⁸ Earp B. D. et al. *Out, Damned Spot: Can the “Macbeth Effect” Be Replicated?* Basic and Applied Social Psychology. 36, no. 1 (2014): 91–8. Эффект денежного прайминга: Klein R. A. et al. *Investigating Variation in Replicability: A “Many Labs” Replication Project.* Social Psychology. 45, no. 3 (2014): 142–52.

⁶⁹ Исходное исследование: Williams L. E., Bargh J. A. *Keeping One’s Distance: The Influence of Spatial Distance Cues on Affect and Evaluation.* Psychological Science. 19, no. 3 (2008): 302–8. Повторение: Pashler H. et al. *Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments.* PLOS ONE. 7, no. 8 (2012): e42510.

ду⁷⁰. Сходное направление исследований, согласно которому возможно сделать людей более категоричными, подвергнув их праймингу “отвращением”, было поставлено под сомнение в обзоре 2015 года⁷¹.

Надо отдать Канеману должное – позже он признал, что был неправ, переоценив научную достоверность эффектов прайминга. “Экспериментальные доказательства идей, представленных мною в той главе, были значительно слабее, чем я полагал, когда писал ее, – сказал он через шесть лет после выхода книги “Думай медленно... решай быстро”. – Это было попросту ошибкой: все, что я должен был знать для того, чтобы сдерживать свой энтузиазм, я знал... но не обду-

⁷⁰ Исходное исследование: Zarkadi T., Schnall S. “Black and White” Thinking: Visual Contrast Polarizes Moral Judgment. *Journal of Experimental Social Psychology*. 49, no. 3 (2013): 355–9. Повторение: IJzerman H., Laine P.-J. Does Background Color Affect Moral Judgment? Three Pre-Registered Replications of Zarkadi and Schnall’s (2012) Study 1. Preprint, PsyArXiv (30 July 2018).

⁷¹ Для прайминга “отвращением” экспериментаторы часто делали так, чтобы в помещении дурно пахло. Поэтому исследования на эту тему особенно примечательны: во многих статьях психологам приходилось с непроницаемым видом разглагольствовать об эффектах “спрея с запахом кишечных газов”, а в одной работе авторы невозмутимо обсуждали “патентованный одорант под названием «Жидкая задница»”. О средстве “Жидкая задница” см. Adams T. G. et al. *The Effects of Cognitive and Affective Priming on Law of Contagion Appraisals*. *Journal of Experimental Psychopathology*. 3, no. 3 (2012): 473. Обзор этого направления исследований: Landy J. F., Goodwin G. P. Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental Evidence. *Perspectives on Psychological Science*. 10, no. 4 (2015): 518–36.

мал”⁷². Однако вред уже был нанесен: нобелевский лауреат объявил миллионам людей, что “нельзя не доверять” этим исследованиям.

Прайминг – не единственный психологический эффект, о котором узнали миллионы. Гарвардский психолог Эми Кадди прославилась в 2012 году, после того как выступила на конференции TED, восхваляя “позы силы”. Непосредственно перед тем, как вы окажетесь в стрессовой ситуации, скажем на собеседовании, вам нужно потратить две минуты, советовала она, чтобы в каком-нибудь укромном уголке (вроде туалетной кабинки) встать в открытую, экспансивную позу, например широко расставив ноги и уперев руки в бедра. Эта властная поза стимулирует вас психологически и гормонально. В эксперименте, проведенном Кадди и ее коллегами в 2010 году, обнаружилось, что по сравнению с людьми, которым предписано было сидеть скрестив руки или ссутулившись, те, кому велели принять позу силы, не только чувствовали себя более властными, но и шли на больший риск в азартной игре и имели повышенный уровень тестостерона и пониженный – кортизола, гормона стресса⁷³.

⁷² McCook A. *“I Placed Too Much Faith in Underpowered Studies”*: Nobel Prize Winner Admits Mistakes. Retraction Watch. 20 Feb. 2017. Еще Канеман написал открытое письмо, адресованное социальным психологам, в котором сообщил, что узрел “угрозу катастрофы”, и призвал их изменить свой подход к исследованиям; см. тут: <https://go.nature.com/2T7A2NV>.

⁷³ Carney D. R. et al. *Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance*. Psychological Science. 21, no. 10 (2010): 1363–8.

Идея Кадди, будто люди, на две минуты принимающие позу силы, могут “существенно менять исход жизненных ситуаций”, вызвала глубокий отклик: ее выступление на конференции TED стало вторым по количеству просмотров из всех – его посмотрели больше семидесяти трех с половиной миллионов раз⁷⁴. Далее, в 2015 году, вышла книга Кадди по саморазвитию под названием “Присутствие [духа]”, бестселлер по версии газеты *The New York Times*, и издатель сообщил нам, что там представлена “захватывающая наука”, способная “освободить [нас] от страха в напряженные моменты жизни”⁷⁵. Консервативная партия Великобритании, похоже, прониклась идеей Кадди, поскольку в тот самый год появилась череда фотографий, на которых представители этой партии на разных конференциях и выступлениях принимали позы с широко расставленными ногами, что вызвало немало насмешек⁷⁶. В том же 2015 году другая команда ученых попыталась воспроизвести эффекты поз силы. И хотя те,

⁷⁴ На момент написания этой книги в феврале 2020 года общее количество просмотров на сайте TED составляло 56 миллионов, а на *YouTube* – еще 17,6 миллиона. Выступление исходно называлось “Язык тела формирует вашу личность”, но затем, когда грянул кризис воспроизводимости, оно было переименовано и стало называться “Язык тела может формировать вашу личность”. Cuddy A. *Your Body Language May Shape Who You Are*. TEDGlobal 2012. June 2012.

⁷⁵ Cuddy A. J. C. *Presence: Bringing Your Boldest Self to Your Biggest Challenges*. New York: Little, Brown and Company, 2015. Цитаты взяты с издательского сайта: www.littlebrown.com/titles/amy-cuddy/presence/9780316256575.

⁷⁶ Khaleeli H. *A Body Language Lesson Gone Wrong: Why is George Osborne Standing like Beyoncé?* The Guardian. 7 Oct. 2015.

кто принимал такие позы, действительно сообщали о большей уверенности в себе, исследование, увы, “не подтвердило влияние поз силы на тестостерон, кортизол и финансовый риск”⁷⁷.

Кризис воспроизводимости направил прожектор критики также и на более ранние работы по психологии – со сходными тревожными результатами. Вероятно, самое известное исследование по психологии за всю историю – это стэнфордский тюремный эксперимент 1971 года, когда психолог Филип Зимбардо разделил группу молодых мужчин на “охранников” и “заключенных” и велел им неделю оставаться в импровизированной тюрьме в подвале факультета психологии Стэнфордского университета. Настораживающе быстро, по

⁷⁷ Ranehill E. et al. *Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women*. *Psychological Science*. 26, no. 5 (2015): 653–6. С тех пор дебаты по поводу поз силы только набирали обороты. Авторы обзора 2017 года заключили, что эффекты поз силы – это “предположения, на данный момент не имеющие эмпирических обоснований”. См. Simmons J. P., Simonsohn U. *Power Posing: P-Curving the Evidence*. *Psychological Science*. 28, no. 5 (2017): 687–93. Кадди парировала своим собственным обзором, где отмечался-таки общий эффект, хотя позже и было показано, что – наряду с другими проблемами подобных исследований – большинство результатов в статьях, на которые она ссылалась, объяснялись, вероятно, отрицательным эффектом ссутуливания, нежели чем положительным воздействием поз силы. См. Cuddy A. J. C. et al. *P-Curving a More Comprehensive Body of Research on Postural Feedback Reveals Clear Evidential Value for Power-Posing Effects: Reply to Simmons and Simonsohn* (2017). *Psychological Science*. 29, no. 4 (2018): 656–66. Про ссутуливание см. Credé M. *A Negative Effect of a Contractive Pose is not Evidence for the Positive Effect of an Expansive Pose: Commentary on Cuddy, Schultz, and Fosse* (2018). SSRN. 2018.

словом Зимбардо, “охранники” принялись наказывать “заключенных”, мучая их столь садистски, что Зимбардо пришлось досрочно прекратить эксперимент⁷⁸. Наряду с исследованиями подчинения, проведенными в 1960-х годах Стэнли Милгрэмом, в которых обнаружилось, что многие участники готовы воздействовать сильными электрическими разрядами на незадачливых “учеников” (удары током и ученики были ненастоящими, но участники об этом не знали), эксперимент Зимбардо приводится как одно из основных доказательств власти ситуации над человеческим поведением⁷⁹. Поставьте, как говорится, хорошего человека в плохую ситуацию – и дела могут очень быстро пойти очень скверно. О стэнфордском тюремном эксперименте рассказывают фактически каждому студенту на планете, изучающему психологию, и Зимбардо благодаря ему стал одним из самых известных и уважаемых современных психологов. Он использовал результаты своего эксперимента, чтобы, например, выступать в качестве свидетеля-эксперта на суде над амери-

⁷⁸ Zimbardo P. *The Lucifer Effect: How Good People Turn Evil*. London: Rider, 2007. [Зимбардо Ф. *Эффект Люцифера. Почему хорошие люди превращаются в злодеев*. М.: АНФ, 2013.]

⁷⁹ Milgram S. *Behavioral Study of Obedience*. *Journal of Abnormal and Social Psychology*. 67, no. 4 (1963): 371–8. Эксперименты Милгрэма тоже подвергались справедливой критике. Чем сильнее участники верили, что действительно бьют “учеников” током, тем с меньшей вероятностью повышали разряд. Об этом см., например: Perry G. et al. *Credibility and Incredulity in Milgram’s Obedience Experiments: A Reanalysis of an Unpublished Test*. *Social Psychology Quarterly*. 83, no. 1 (2020): 88–106.

канскими военными, служившими охранниками в иракской тюрьме Абу-Грейб. Зимбардо утверждал, что ситуация, в которой оказались охранники, и роли, которые их заставили на себя взять, – вот причины их шокирующих издевательств и пыток над заключенными⁸⁰.

Хотя выводы из стэнфордского тюремного эксперимента всегда были противоречивы, лишь недавно мы начали понимать, насколько низкокачественным было то исследование⁸¹. В 2019 году ученый и кинорежиссер Тибо Ле Тексье опубликовал статью под названием “Разоблачение стэнфордского тюремного эксперимента”. Он представил доселе неизвестную расшифровку видеозаписей, на которых Зимбардо вмешивается непосредственно в эксперимент, раздавая своим “охранникам” очень точные инструкции, как себя вести, – вплоть до того, что предлагает конкретные способы обезчеловечивания заключенных, например отказывать им в использовании туалетов⁸². Очевидно, столь основательно срежиссированная постановка эксперимента была далека от естественного примера того, что происходит, когда обычные люди оказываются в специфических социальных ролях. Как

⁸⁰ Zimbardo P. *Our inner heroes could stop another Abu Ghraib*. The Guardian. 29 Feb. 2008.

⁸¹ Fromm E. *The Anatomy of Human Destructiveness*. New York: Holt, Rinehart and Winston, 1975. [Фромм Э. *Анатомия человеческой деструктивности*. М.: АСТ, 2004.]

⁸² Le Texier T. *Debunking the Stanford Prison Experiment*. American Psychologist. 74, no. 7 (2019): 823–39.

бы то ни было, несмотря на колоссальное внимание, которое долгие годы привлекал к себе стэнфордский тюремный эксперимент, его “результаты” с научной точки зрения не имеют смысла⁸³.

Как вы, наверное, догадались, психологов напугала совокупность неудавшихся попыток повторить эксперименты (как в исследованиях прайминга) и странных результатов (вроде паранормальных открытий Бема) наряду с разоблачением ложных представлений (как в эксперименте Зимбардо) и мошенничества (поддельные данные Стапела). Сколько же исследованиям в области психологии, недоумевали они, можно доверять? Чтобы получить представление о том, насколько дела плохи, они начали объединяться в команды для проведения крупномасштабных повторов значимых исследований в разных лабораториях. Самым заметным стало крупное сообщество ученых, отобравшее сто работ из трех топовых журналов по психологии и попытавшееся их воспроизвести. Читать о результатах, опубликованных в 2015 году в *Science*, было горько: в конечном счете лишь 39 % работ были признаны успешно воспроизведенными⁸⁴. В дру-

⁸³ Дебаты продолжаются, и Зимбардо ответил на критику. См., например: Zimbardo P. *Philip Zimbardo's Response to Recent Criticisms of the Stanford Prison Experiment*. 23 June 2018. См. также ответ Ле Тексье на более позднюю версию заявления Зимбардо (на момент написания этой книги – еще не опубликованную): Le Texier T. *The SPE Remains Debunked: A Reply to Zimbardo and Haney* (2020). Preprint, PsyArXiv (24 Jan. 2020).

⁸⁴ Open Science Collaboration. *Estimating the Reproducibility of Psychological*

гом подобном предприятии 2018 года ученые пытались повторить двадцать одно исследование по социальным наукам из двух самых авторитетных многопрофильных журналов в мире – *Nature* и *Science*. На сей раз воспроизвелось 62 %⁸⁵. В последующих масштабных повторах исследований, касающихся разнообразных психологических феноменов, воспроизвелось 77, 54 и 38 % результатов⁸⁶. Почти все повторы, даже успешные, продемонстрировали, что в исходных статьях эффекты были преувеличены. В целом кризис воспроизводимости легким движением руки стер, похоже, около половины всех исследований по психологии⁸⁷.

Science. *Science*. 349, no. 6251 (2015): aac4716.

⁸⁵ Camerer C. F. et al. *Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015*. *Nature Human Behaviour*. 2, no. 9 (2018): 637–44.

⁸⁶ Последнее число соответствует шести удачным попыткам воспроизвести шестнадцать исследований. Ebersole C. R. et al. *Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester via Replication*. *Journal of Experimental Social Psychology*. 67 (2016): 68–82.

⁸⁷ Тут некоторые критики могут возразить, что я сам попал в вырытую другим яму. Я подчеркивал важность надежных результатов, однако, заявляя, что разразился кризис воспроизводимости, полагаюсь на попытки воспроизвести разные исследования, которые не являются репрезентативной выборкой из всей научной литературы. Вывод о том, что лишь “около половины” опубликованных результатов воспроизводятся, возможно, нельзя обобщать на всю науку. Такой аргумент был приведен в критической заметке к одному из исследований, посвященных повторению других работ: Gilbert D. T. et al. *Comment on “Estimating the Reproducibility of Psychological Science”*. *Science*. 351, no. 6277 (2016): 1037. Хотя я не согласен со многими приведенными там аргументами (некоторые причины скепсиса описаны здесь: Lakens D. *The Statistical Conclusions in Gilbert et al* (2016)

Возможно, все не столь уж плохо – по двум причинам. Во-первых, следует ожидать, что некоторые результаты, на самом деле надежные, иногда не получается воспроизвести просто по невезению⁸⁸. Во-вторых, какие-то повторы могли провалиться из-за того, что их проводили с небольшими изменениями в методологии по сравнению с исходным исследованием (правда, если результат настолько нестабилен, что исчезает при малейших изменениях в постановке эксперимента, то возникает вопрос, имеет ли он вообще какой-то смысл и применение)⁸⁹. По этим причинам иногда трудно понять, является результат “воспроизводимым” или нет, на основании только одной-двух попыток его повторить. Кро-

Are Completely Invalid. The 20 % Statistician. 6 March 2016), замечание насчет репрезентативности справедливо. Мы все еще плохо понимаем, какая именно часть результатов по всем научным направлениям воспроизводима, даже в областях вроде психологии, где были предприняты эти масштабные попытки повторить прежние результаты, – правда может оказаться пригляднее, чем показывают те исследования, или же наоборот. Но сам факт, что мы этого не знаем, – а также то, что столь много громких, нашумевших открытий рассыпалось в прах при более внимательном рассмотрении, – является, я убежден, поводом для немалого беспокойства. Ответы на другие критические замечания, оспаривающие идею, что разразился кризис, см. тут: Pashler H., Harris C. R. *Is the Replicability Crisis Overblown? Three Arguments Examined*. Perspectives on Psychological Science. 7, no. 6 (2012): 531–6.

⁸⁸ Bird A. *Understanding the Replication Crisis as a Base Rate Fallacy*. British Journal for the Philosophy of Science. 13 Aug. 2018.

⁸⁹ Разумеется, авторы исходной работы (те, чьи результаты не воспроизвелись) зачастую заявляли, что изменения были на самом деле значительными и серьезно испортили эксперимент. Каждый случай следует разбирать отдельно, однако подобный аргумент наводит на мысль о некоей предвзятости.

ме того, доля воспроизводимых исследований для разных областей психологии, похоже, отличается: например, в статье 2015 года, вышедшей в *Science*, когнитивная психология (изучение памяти, восприятия, языка и так далее) проявила себя лучше, чем социальная (к коей относятся всевозможные исследования прайминга, обсуждавшиеся выше)⁹⁰.

Однако в целом на психологию все это подействовало опустошительно. Дело было не только в том, что разоблачались такие легковесные, эффектные исследования, как посвященные праймингу или позам силы, – огромное количество куда более “серьезных” работ по психологии (стэнфордский тюремный эксперимент и многие другие) тоже оказалось поставлено под сомнение. И проблема была не в том, что откопали какое-то ненужное старье и наглядно показали, что оно никуда не годится, – как когда папа Стефан VI в 897 году эксгумировал труп одного из своих предшественников, папы Формоза, и отдал под суд (тот был признан виновным). Нет, на работы, воспроизвести которые не удалось, продол-

⁹⁰ Другая область, где все неплохо, – психология личности. Психолог Кристофер Сото провел масштабную работу по воспроизведению результатов исследований личности – корреляций личностных качеств, оцениваемых по опросникам, с такими показателями, как удовлетворенность жизнью и романтическими отношениями, религиозные и политические взгляды и карьерный успех. Доля успешно воспроизведенных результатов получилась 87 %, что весьма достойно по сравнению с другими областями, которые мы обсуждали. Soto C. J. *How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project*. *Psychological Science*. 30, no. 5 (2019): 711–27.

жали как ни в чем не бывало ссылаться как ученые, так и писатели: целые направления исследований и пользующиеся хорошим спросом научно-популярные книги строились на таком шатком фундаменте. Слово “кризис” кажется весьма точным описанием ситуации.

Мы можем попробовать утешиться тем, что в психологии как дисциплине есть нечто уникальное, что и вызвало ее кризис воспроизводимости. У психологов незавидная работа: они пытаются разобраться в крайне изменчивых и чрезвычайно сложных человеческих существах, со всеми их разными личностями, знаниями, опытом, настроениями и особенностями. Изучаемые психологами объекты, такие как мысли, эмоции, внимание, способности, восприятие, обычно неуловимы – их трудно, если вообще возможно, зафиксировать в лабораторном эксперименте. А в социальной психологии ученым приходится изучать, как все эти хитросплетенные люди друг с другом взаимодействуют. Не могла ли невероятная сложность задачи сделать открытия в психологии особенно ненадежными по сравнению с другими науками?

В этом аргументе есть кое-что справедливое: во многих исследованиях по психологии интересующее ученых явление едва затрагивается, тогда как другие, более “точные” науки, скажем физика, характеризуются лучше разработанными теориями и более точными и по-настоящему объективными измерениями. Однако нельзя сказать, что только в пси-

хологии есть проблемы с воспроизводимостью: хотя ни в какой другой области науки столь систематически и детально еще не изучалась доля успешно воспроизводящихся результатов, есть намеки на однотипные проблемы в огромном количестве разных направлений.

- Экономика: в исследовании 2016 года, повторяющем восемнадцать работ по микроэкономике (когда люди приходят в лабораторию и принимают участие в экспериментах, посвященных их экономическому поведению, – что не слишком отличается от исследований по психологии), доля воспроизводимости равнялась лишь 61 %⁹¹.

- Нейронауки: в исследовании 2018 года обнаружилось, что стандартные работы по функциональной нейровизуализации, когда с помощью магнитно-резонансной томографии регистрируется активность мозга, пока человек выполняет какие-то задания (или просто лежит внутри МРТ-сканера), отличались лишь “незначительной воспроизводимостью”⁹². Еще мир функциональной нейровизуализации сотрясла статья, в которой вскрылось, что дефолтные настройки пакета программ, широко используемого для анализа данных визуализации, содержат статистическую ошибку. Это привело к

⁹¹ Camerer C. F. et al. *Evaluating Replicability of Laboratory Experiments in Economics*. Science. 351, no. 6280 (2016): 1433–6.

⁹² Turner B. O. et al. *Small Sample Sizes Reduce the Replicability of Task-Based fMRI Studies*. Communications Biology. 1, no. 1 (2018): 62.

громадному числу случайных нескорректированных ложноположительных результатов и скомпрометировало примерно 10 % всех статей, когда-либо опубликованных по этой теме⁹³.

• Эволюционная биология и экология: на целый ряд классических результатов, давно попавших в учебники и вызубриваемых поколениями студентов, посыпались критические обзоры после попыток их воспроизвести. Так, выяснилось, что заявления о знаменитом “синдроме одомашнивания”, когда лисицы в СССР, отбираемые по признаку дружелюбности, начинали приобретать внешний облик одомашненных видов (например, висячие уши и укороченные, широкие морды), были сильно преувеличены, причем большинство признаков “приручения” существовало еще до начала процесса селекции⁹⁴. И многое из того, что, как мы думали, нам известно о половом отборе у птиц, было развенчано при получении более надежных данных. Скажем, в противоположность тому, что мы якобы знали, красная повязка на лапках у самцов зебровых амадин, похоже, *не* делает их сверхпривлекательными для самок; самцы воробьев с более

⁹³ Eklund A. et al. *Cluster Failure: Why fMRI Inferences for Spatial Extent Have Inflated False-Positive Rates*. Proceedings of the National Academy of Sciences. 113, no. 28 (2016): 7900–5; Eklund A. et al. *Cluster Failure Revisited: Impact of First Level Design and Physiological Noise on Cluster False Positive Rates*. Human Brain Mapping. 40, no. 7 (2019): 2017–32.

⁹⁴ Lord K. A. et al. *The History of Farm Foxes Undermines the Animal Domestication Syndrome*. Trends in Ecology & Evolution. 35, no. 2 (2020): 125–36.

крупным пятном черных перьев на горле (так называемым нагрудником), похоже, не доминируют в стае; а доказательства, что самок обыкновенных лазоревок больше привлекают определенные цвета оперения у самцов, неубедительны⁹⁵.

• Биология моря: в масштабном исследовании 2020 года, повторяющем другие работы, выяснилось, что закисление океана (как и изменение климата, это одно из последствий повышения уровня диоксида углерода в атмосфере) не влияет на поведение рыб⁹⁶. Таким образом, не удалось воспроизвести несколько исследований предыдущего десятилетия, получивших широкую огласку, которые явно показывали, что в закисленной среде рыбы становятся дезориенти-

⁹⁵ Зебровые амадины: Wang D. et al. *Irreproducible Text-Book “Knowledge”: The Effects of Color Bands on Zebra Finch Fitness*. *Evolution*. 72, no. 4 (2018): 961–76. См. также Law Y.-H. *Replication Failures Highlight Biases in Ecology and Evolution Science*. *The Scientist*. 31 July 2018. Воробы: Sánchez-Tójar A. et al. *Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias*. *eLife*. 7 (2008): e37385. Обыкновенные лазоревки: Parker T. H. *What Do We Really Know about the Signalling Role of Plumage Colour in Blue Tits? A Case Study of Impediments to Progress in Evolutionary Biology*. *Biological Reviews*. 88, no. 3 (2013): 511–36.

⁹⁶ Clark T. D. et al. *Ocean Acidification Does Not Impair the Behaviour of Coral Reef Fishes*. *Nature*. 577, no. 7790 (2020): 370–5. См. также Enserink M. *Analysis Challenges Slew of Studies Claiming Ocean Acidification Alters Fish Behavior*. *Science*. 8 Jan. 2020. Как отмечается в этой второй статье, из того, что поведение рыб, похоже, не меняется, не следует, что нам нужно перестать беспокоиться о закислении океана, которое вызывает много других пагубных явлений. Позднее авторов исходных статей (двадцати двух штук) о закислении океана и поведении рыб обвинили в научном мошенничестве. См. Enserink M. *Does Ocean Acidification Alter Fish Behavior? Fraud Allegations Create a Sea of Doubt*. *Science*. 6 May 2021.

рованными и иногда плывут *по направлению* к химическим сигналам хищников, а не от них.

- Органическая химия: журнал *Organic Syntheses*, придерживающийся необычной политики – член редколлегии пробует повторить в собственной лаборатории результаты каждой подаваемой на рассмотрение статьи, – сообщил, что отказывает авторам 7,5 % работ из-за провалившихся попыток воспроизвести исследование⁹⁷.

Есть бесчисленное множество и других примеров: почти каждый случай, что я буду описывать в этой книге, содержит научное “открытие”, при ближайшем рассмотрении оказавшееся либо менее надежным, чем казалось, либо и вовсе недостоверным. Однако еще тревожнее то, что эти примеры порождены исследованиями, которые подверглись столь тщательному изучению, – получается, это лишь те примеры, о которых мы знаем. Сколько еще результатов, должны задаться мы вопросом, окажутся невозпроизводимыми, если кому-то случится попробовать их повторить?

Одна из причин, почему мы живем в такой неопределенности, заключается в том, что, как говорилось в предисловии, почти никто не проводит исследований, повторяющих прежние работы. Хотя в нашем распоряжении для большинства областей нет количественных данных, анализ специа-

⁹⁷ www.orgsyn.org/instructions.aspx. См. также Chawla D. S. *Taking on Chemistry's Reproducibility Problem*. Chemistry World. 20 March 2017.

лизированной литературы для некоторых из них позволяет сделать мрачный вывод. В экономике жалкие 0,1 % всех опубликованных статей посвящены попыткам воспроизвести предыдущие исследования; в психологии этот показатель выше, но все равно весьма прискорбный – чуть больше 1 %⁹⁸. Если все неустанно рвутся вперед к новым открытиям, не делая остановок, чтобы проверить, надежны ли уже имеющиеся знания, так ли уж удивителен приведенный выше список провалившихся попыток что-то воспроизвести?

А вот что вызывает, пожалуй, еще большую озабоченность. Казалось бы, если вы получили точно такой же набор данных, как и в опубликованной ранее статье, вы сможете прийти к абсолютно тем же результатам, что описаны авторами. К сожалению, во многих областях исследователи стал-

⁹⁸ Поиск нужных статей проводился таким образом, что исследования, которые открыто не объявляли себя попытками воспроизвести предыдущие работы, могли оказаться неучтенными, поэтому итоговый процент, вероятно, на самом деле чуточку больше. Экономика: Mueller-Langer F. et al. *Replication Studies in Economics – How Many and Which Papers Are Chosen for Replication and Why?* Research Policy. 48, no. 1 (2019): 62–83. Психология: Makel M. C. et al. *Replications in Psychology Research: How Often Do They Really Occur?* Perspectives on Psychological Science. 7, no. 6 (2012): 537–42. Также обращаю ваше внимание: по поводу того, что считать попыткой воспроизвести исследование, ведутся споры. Некоторые ученые провели множество “содержательных” повторений, в целом похожих на исходное исследование, но в деталях иногда от него отличающихся. Это по-своему интересно, но это не “прямое” повторение, когда именно та же, насколько только возможно, работа проводится сызнова. Вот такого рода исследований-повторений и не хватает. См. Schmidt S. *Shall We Really Do It Again? The Powerful Concept of Replication is Neglected in the Social Sciences*. Review of General Psychology. 13, no. 2 (2009): 90–100.

кивались с невероятными трудностями при выполнении этой вроде бы нехитрой задачи. Иногда именно подобную проблему называют проблемой *воспроизводимости*, в противоположность проблеме *сходимости* результатов (последний термин обычно используется применительно к исследованиям, в которых ученые задаются теми же вопросами, но работают с другими данными)⁹⁹. Как это возможно, чтобы результаты в таких условиях не воспроизвелись? Иногда причина в ошибках исходного исследования. А бывает и так, что авторы исходной работы недостаточно четко описали свой анализ, например, прибегали ко всяким выкрутасам со статистикой, о которых в статье не доложили, и поэтому их конкретные шаги независимые исследователи воссоздать не могут. Когда другие ученые как-то по-своему проводят статистический анализ данных, результаты выходят иные. Такие статьи – словно кулинарная книга, где полно фотографий блюд, от которых просто слюнки текут, но мало внимания уделено описанию ингредиентов и рецептам, необходимым

⁹⁹ Во избежание путаницы заметим, что в англоязычной специализированной литературе существуют два разных термина: “проблема *сходимости* результатов” (*replicability* или *repeatability*) и “проблема *воспроизводимости* результатов” (*reproducibility*). В русскоязычной же литературе обычно не делается различий между этими случаями и используется единый термин – “проблема *воспроизводимости*”. Учитывая, что и в английском языке применение двух разных терминов не строгое (на что, в частности, указывает и сам автор в примечании 49 к этой главе), в русском переводе данной книги используется только термин “воспроизводимость”, тем более что необходимые детали соответствующих исследований там, где они важны, поясняются автором отдельно. (*Здесь и далее – прим. перев.*)

для создания этих шедевров.

В макроэкономике (изучающей, например, налоговую политику и ее влияние на экономическое развитие стран) при повторном анализе шестидесяти семи статей ученые, используя точно такие же наборы данных, сумели воспроизвести результаты лишь двадцати двух, и последующее привлечение к работе авторов тех статей помогло несильно¹⁰⁰. В науках о Земле исследователи испытывали как минимум небольшие трудности при получении тех же результатов в случае тридцати семи из тридцати девяти изучавшихся ими статей¹⁰¹. А когда исследователи машинного обучения проанализировали набор статей об “алгоритмах рекомендаций” (это тип компьютерных программ, которые используются сайтами вроде *Amazon* и *Netflix*, чтобы на основании того, что люди вроде вас выбирали раньше, предугадывать, какую покупку вам сейчас захотелось бы сделать или какой фильм посмотреть), то смогли воспроизвести только семь из восемнадцати работ на эту тему, незадолго до того представ-

¹⁰⁰ Chang A. C., Li P. *Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals say “Usually Not”*. Finance and Economics Discussion Series. 2015, no. 83 (2015): 1–26. Washington: Board of Governors of the Federal Reserve System. Подробный обзор проблемы воспроизводимости в экономике: Christensen G., Miguel E. *Transparency, Reproducibility, and the Credibility of Economics Research*. Working Paper no. 22989. National Bureau of Economic Research. 2016.

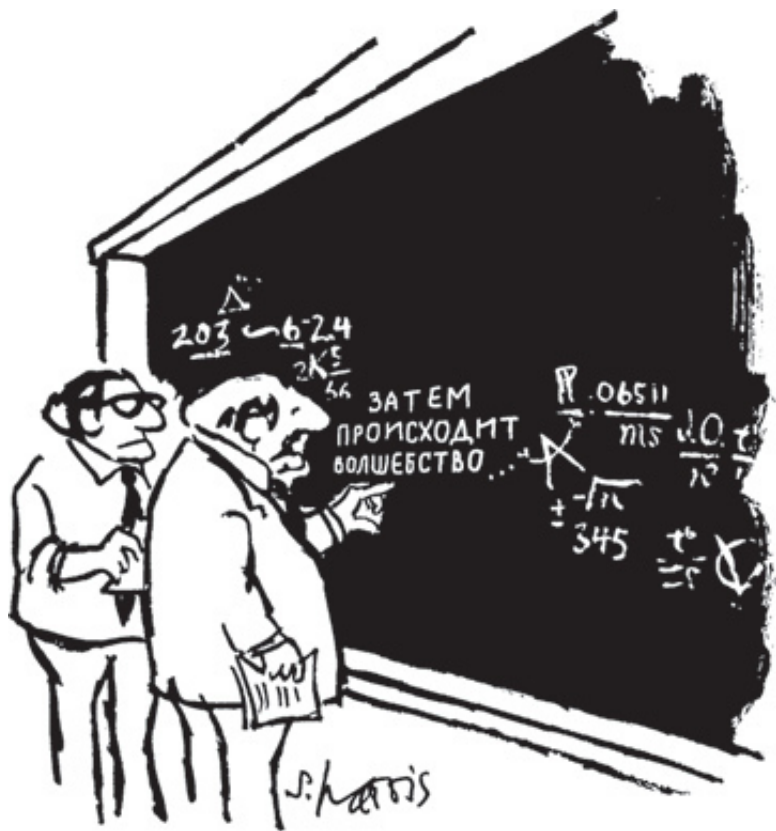
¹⁰¹ Konkol M. et al. *Computational Reproducibility in Geoscientific Papers: Insights from a Series of Studies with Geoscientists and a Reproduction Study*. International Journal of Geographical Information Science. 33, no. 2 (2019): 408–29.

ленных на престижных конференциях по компьютерным системам¹⁰². Те статьи – воплощение классической карикатуры Сидни Харриса.

Вы вправе удивиться, почему некоторые из перечисленных выше примеров вообще имеют значение. Хотя мы и наблюдали плохую воспроизводимость в кое-каких важных областях, например в экономической теории, каким образом наша жизнь может измениться, если кучка ученых в итоге разойдется во взглядах на то, работают ли позы силы и отличаются ли альфа-самцы воробьев более крупным пятном черных перьев? На этот вопрос есть два ответа. Первый заключается в том, что на чашу весов положен более общий принцип: наука критически важна для нашего общества, и мы не должны допускать появления низкокачественных, невозпроизводимых исследований, компрометирующих ее, ни в одной области. Если мы позволим стандартам в любой области просесть, мы рискуем испортить репу-

¹⁰² И даже хуже: из этих семи статей в целых шести методы избыточны по сравнению с гораздо более простыми методами, которые были известны за много лет до того, как создавались эти новые алгоритмы. Dacrema M. F. et al. *Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches*. RecSys 2019. Copenhagen, Denmark. См. также отчет по компьютерным наукам, который дает понять, что у новых исследователей не получается воспроизвести результаты применения нескольких классических алгоритмов – а это своего рода бомба замедленного действия, поскольку “молодые ученые не хотят выглядеть хулиателями старших коллег”, публикуя статьи о провалившихся попытках воспроизвести действие алгоритмов, которые были разработаны старшими коллегами и на которых держится их репутация. Hutson M. *Artificial Intelligence Faces Reproducibility Crisis*. Science. 359, no. 6377 (2018): 725–6.

тацию науки в целом. Второй ответ связан с научным направлением, которое мы еще не рассматривали, где отсутствие воспроизводимости имеет бесспорные прямые последствия. Это, конечно же, область медицинских исследований.



“Думаю, вам следует подробнее расписать второй шаг”

Примерно в то время, когда кризис воспроизводимости назревал в психологии, ученые из биотехнологической компании *Amgen* попробовали повторить пятьдесят три ключевых “доклинических” исследования рака, результаты которых были опубликованы в топовых научных журналах (доклинические исследования – это те, что проводятся на первых этапах разработки лекарства, обычно на мышах или на человеческих клетках *in vitro*¹⁰³)¹⁰⁴. Всего шесть из этих попыток (а это лишь 11 %) увенчались успехом. Итог сходных попыток другой фирмы, *Bayer*, оказался немногим лучше – около 20 %¹⁰⁵. Подобное отсутствие строгого подкрепления результатов в области доклинических исследований – вероятно, одна из причин, по которым испытания лекарств от рака так часто разочаровывают: согласно одной оценке, только 3,4 % таких лекарств проходят весь путь от первых доклинических исследований до применения на людях¹⁰⁶.

¹⁰³ То есть в пробирке.

¹⁰⁴ Begley C. G., Ellis L. M. *Raise Standards for Preclinical Cancer Research*. *Nature*. 483, no. 7391 (2012): 531–3.

¹⁰⁵ Prinz F. et al. *Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?* *Nature Reviews Drug Discovery*. 10 (2011): 712. Обратите внимание, что приведенная в статье сотрудниками *Bayer* диаграмма включает только 70 % исследований рака – остальные 30 % связаны с женским здоровьем или сердечно-сосудистой системой.

¹⁰⁶ Wong C. H. et al. *Estimation of Clinical Trial Success Rates and Related Parameters*. *Biostatistics*. 20, no. 2 (2019): 273–86. Из всех разнообразных лекарств

Подобные неприятные открытия заставили ученых, занимающихся раком, как и психологов, задуматься об общем состоянии их области. В 2013 году они объединились для совместной попытки повторить пятьдесят важных доклинических исследований рака в независимых лабораториях¹⁰⁷. В тех работах утверждалось, например, что конкретный вид бактерий может быть связан с ростом опухоли при колоректальном раке и что некоторые мутации при лейкемии связаны с активностью определенного фермента¹⁰⁸. Но еще до начала проведения исследований-повторений возникли трудности. В каждой без исключений исходной статье для всех до единого экспериментов авторы приводили недостаточно сведений для того, чтобы читатель мог понять, как провести точно такой же опыт¹⁰⁹. Технические детали экспериментов – например, плотность используемых клеток или особенно-

доля тех, что добиваются от доклинических испытаний до применения на людях, оценивается в данной работе в 13,8 %, так что с исследованиями рака все особенно плохо.

¹⁰⁷ Nosek B. A., Errington T. M. *Reproducibility in Cancer Biology: Making Sense of Replications*. eLife. 6 (2017): e23383. В названии “Проект по проверке воспроизводимости: биология рака” термин “воспроизводимость” используется в том же смысле, в каком я использую слово “сходимость” (то есть попытка получить те же результаты на другой выборке). Я выбрал определения для этой книги таким образом, чтобы отразить сложившийся консенсус, однако стоит понимать, что не все придерживаются той же терминологии.

¹⁰⁸ Repass J. et al. *Replication Study: Fusobacterium Nucleatum Infection is Prevalent in Human Colorectal Carcinoma*. eLife. 7 (2018): e25801.

¹⁰⁹ Errington T. *Reproducibility Project: Cancer Biology – Barriers to Replicability in the Process of Research*. 2019. <https://osf.io/x9p5s/>

сти измерений и анализа – попросту не упоминались. Работа по повторению исследований забуксовала, поскольку потребовала обширной переписки с авторами исходных статей, которым для отыскания конкретных деталей своих экспериментов зачастую приходилось откапывать старые лабораторные журналы и связываться с бывшими коллегами¹¹⁰. Кто-то сотрудничать не хотел: о 45 % авторов исходных статей исследователи отзывались как о “минимально полезных” или “вообще не оказавших помощи”¹¹¹. Возможно, они беспокоились, что ученые, повторяющие их работу, окажутся некомпетентными или что их будущие исследования не получат финансирования, если воспроизвести их прежние результаты не удастся¹¹².

Позже в более масштабном исследовании были случайным образом выбраны двести шестьдесят восемь биомедицинских статей, в том числе и из области клинических испытаний. Обнаружилось, что ни в одной из них, за *единственным* исключением, не содержался полный протокол исследования. А это означало опять-таки, что даже для того, чтобы предпринять *попытку* повторить эксперименты, требовалось знать дополнительные детали, не описанные в ста-

¹¹⁰ Baker M., Dolgin E. *Cancer Reproducibility Project Releases First Results*. Nature. 541, no. 7637 (2017): 269–70; Engber D. *Cancer Research Is Broken*. Slate. 19 April 2016.

¹¹¹ Errington T. *Reproducibility Project* (см. слайд 11).

¹¹² Kaiser J. *The Cancer Test*. Science. 348, no. 6242 (2015): 1411–3.

ть¹¹³. Другой анализ выявил, что в 54 % биомедицинских статей даже не описывалось толком, какие животные, реагенты или клетки использовались в экспериментах¹¹⁴. Давайте на минуту задумаемся, насколько же это странно. Если в статье исследование описывается лишь в общих чертах, а необходимые детали приходится месяцами выуживать из электронной переписки с авторами (а то они и вовсе оказываются навеки утеряны), зачем вообще было ее писать? Вернувшись мысленно хотя бы в XVII век к Роберту Бойлю, вспомним, что изначальная, глубинная цель ученых состояла в том, чтобы докладывать о каждой мелочи в своих штудиях, так чтобы другие могли досконально все изучить и попробовать повторить их исследования. Упомянутые статьи провалили эту фундаментальную проверку, равно как и опубликовавшие их журналы не сумели выполнить свою базовую, важнейшую функцию.

¹¹³ Iqbal S. A. et al. *Reproducible Research Practices and Transparency across the Biomedical Literature*. PLOS Biology. 14, no. 1 (2016): e1002333. Обратите внимание, что в полную выборку вошло 441 исследование, но только 268 из них содержали эмпирические данные.

¹¹⁴ Vasilevsky N. A. et al. *On the Reproducibility of Science: Unique Identification of Research Resources in the Biomedical Literature*. PeerJ. 1 (2013): e148. Проблемы, касающиеся недостаточно подробного написания статей, выходят за пределы биомедицины. О политологии, например, см. Wuttke A. *Why Too Many Political Science Findings Cannot Be Trusted and What We Can Do About It: A Review of Meta-Scientific Research and a Call for Academic Reform*. Politische Vierteljahresschrift. 60, no. 1 (2019): 1–19. Об экологии см. Parker T. H. et al. *Transparency in Ecology and Evolution: Real Problems, Real Solutions*. Trends in Ecology & Evolution. 31, no. 9 (2016): 711–9.

В проекте по воспроизведению исследований рака все эти проблемы с попытками повторить эксперименты вкупе с некоторыми финансовыми затруднениями привели к тому, что ученым пришлось постепенно сократить число работ, которые они намеревались повторить, – с пятидесяти до всего лишь восемнадцати¹¹⁵. К моменту написания этих строк отчет о повторении результатов четырнадцати из них уже опубликован, и картина получилась смешанная: для пяти исходных статей важные результаты (включая связь между лейкемией и ферментом) четко воспроизвелись, для четырех – воспроизвелись частично, для трех – совсем не воспроизвелись (в том числе связь между бактериями и колоректальным раком), а для двух результаты даже не получились интерпретировать¹¹⁶. Повторять, стоит отметить, непросто.

Проблемы с воспроизводимостью в медицине затронули не только лабораторные доклинические исследования – они могут напрямую влиять на средства лечения, прописываемые врачами своим пациентам. Оказывается, широко распространенные средства лечения часто основаны на низкокачественных исследованиях: вместо того чтобы прочно уxo-

¹¹⁵ Kaiser J. *Plan to Replicate 50 High-Impact Cancer Papers Shrinks to Just 18*. Science. 31 July 2018. Обратите внимание, что в примечании 51 к этой главе (Errington T. *Reproducibility Project*) обсуждается пятьдесят одно исследование, а не пятьдесят.

¹¹⁶ Все о “Проекте по проверке воспроизводимости: биология рака” см. здесь: <https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology>.

дить корнями в доказательства, общепризнанная медицинская мудрость регулярно вступает в противоречие с результатами новых исследований. Подобное происходит так часто, что ученые-медики Винай Прасад и Адам Сифу окрестили это явление “медицинской реверсией”¹¹⁷.

Один особенно поразительный пример медицинской реверсии связан с “интранаркозным пробуждением”. Это неброское название дали жуткому (но, благо, редкому) осложнению, когда человек пробуждается во время операции, иногда чувствуя невыносимую боль от рассечения собственных тканей, и не может ни двигаться, ни говорить, ни сделать что-либо еще. Исследования 1990-х годов обосновали использование прибора под названием “монитор биспектрального индекса” (или “BIS-монитор” – от английского словосочетания *bispectral index*). По сути это электрод, который крепится к коже головы и позволяет хирургам удостовериться, что пациент действительно без сознания. Исследования воплотились в широко распространенную практику: к 2007 году половина операционных в США обзавелась таким прибором, и с ним было проведено около сорока миллионов операций по всему миру¹¹⁸. Однако выяснилось, что с исходными работами не все в порядке. Когда в 2008 году провели более масштабное и более высококачественное исследо-

¹¹⁷ Prasad V. K., Cifu A. S. *Ending Medical Reversal: Improving Outcomes, Saving Lives*. Baltimore: Johns Hopkins University Press, 2015.

¹¹⁸ Lang J. *Awakening*. The Atlantic. Feb. 2013.

вание, обнаружилось, что BIS-монитор бесполезен: “интранаркозное пробуждение случалось, даже когда [приборные] показания... укладывались в допустимый диапазон”¹¹⁹.

В 2019 году Прасад, Сифу и их коллеги проанализировали более трех тысяч статей из трех самых престижных медицинских журналов и обнаружили, что не менее чем в трехстах девяноста шести из них опровергались устоявшиеся в медицинской практике мнения¹²⁰. Вот лишь несколько примеров:

• *Роды*. В некоторых предыдущих исследованиях было показано, что, когда женщина рождает двойню, плановое кесарево сечение – самый безопасный вариант для младенцев. В результате это стало общепринятой практикой (по крайней мере в Северной Америке). Но в крупном рандомизированном исследовании 2013 года ученые не выявили никакой разницы с точки зрения здоровья детей¹²¹.

¹¹⁹ Avidan M. S. et al. *Anesthesia Awareness and the Bispectral Index*. New England Journal of Medicine. 358, no. 11 (2008): 1097.

¹²⁰ Herrera-Perez D. et al. *A Comprehensive Review of Randomized Clinical Trials in Three Medical Journals Reveals 396 Medical Reversals*. eLife. 8 (2019): e45183. Это было продолжением похожего исследования тех же авторов, когда обнаружилось 146 “медицинских реверсий”: Prasad V. et al. *A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices*. Mayo Clinic Proceedings. 88, no. 8 (2013): 790–8.

¹²¹ Barrett J. F. R. et al. *A Randomized Trial of Planned Cesarean or Vaginal Delivery for Twin Pregnancy*. New England Journal of Medicine. 369, no. 14 (2013): 1295–1305.

• *Аллергия.* Аллергия на арахис бывает смертельной, и если у родителя она есть, то высок риск, что она разовьется и у его детей. Долгие годы рекомендации для младенцев из группы риска, основанные на результатах прежних исследований, предписывали не давать детям арахис как минимум до трех лет и кормящим матерям тоже его избегать. Оказывается, этот совет был неверным, все как раз наоборот: добротные рандомизированные испытания 2015 года показали, что лишь у примерно 2 % детей из группы риска, евших арахис в первые годы жизни, к пяти годам развилась на него аллергия, тогда как среди детей из группы риска, не употреблявших арахис, аллергия появилась почти у 14 %¹²².

• *Инфаркт миокарда.* В некоторых небольших исследованиях было показано, что шансы выжить при остановке сердца у человека повышаются, если на несколько градусов снизить его температуру тела. Рекомендацию, основанную на этом открытии, начали включать в руководства для парамедиков. Однако в крупном исследовании 2014 года выяснилось, что охлаждение не влияет на шансы выжить, а то и вообще, возможно, повышает вероятность второго инфаркта при транспортировке пациента в больницу¹²³.

• *Инсульт.* На основании исследований предполагалось,

¹²² Du Toit G. et al. *Randomized Trial of Peanut Consumption in Infants at Risk for Peanut Allergy.* New England Journal of Medicine. 372, no. 9 (2015): 803–13.

¹²³ Kim F. et al. *Effect of Prehospital Induction of Mild Hypothermia on Survival and Neurological Status Among Adults with Cardiac Arrest: A Randomized Clinical Trial.* JAMA. 311, no. 1 (2014): 45–52.

что человека, перенесшего инсульт, лучше всего как можно раньше заставить двигаться: садиться в кровати, стоять, ходить, если возможно. Концепция “ранней мобилизации” встречается во многих широко используемых больничных методичках. Однако в масштабном рандомизированном исследовании 2015 года было показано, что ранняя мобилизация на самом деле вела к более *неблагоприятным* исходам для пациентов с инсультом¹²⁴. Аналогично в исследовании 2016 года выяснилось, что принятая практика переливать пациентам с инсультом тромбоциты (это процедура, восполняющая запасы клеточных элементов, участвующих в процессе свертывания крови, что в теории помогает предотвратить дальнейшее кровотечение) в действительности все только ухудшала¹²⁵.

Вполне понятно, почему врачи и авторы клинических рекомендаций иногда невольно полагаются на мало достоверные доказательства. Зачастую альтернатива – это вообще отсутствие доказательств, а ведь их долг – помочь пациентам, нуждающимся в лечении, *прямо сейчас*

¹²⁴ AVERT Collaboration. *Efficacy and Safety of Very Early Mobilisation within 24 h of Stroke Onset: A Randomised Controlled Trial*. Lancet. 386, no. 9988 (2015): 46–55.

¹²⁵ Baharoglu M. I. et al. *Platelet Transfusion versus Standard Care after Acute Stroke Due to Spontaneous Cerebral Haemorrhage Associated with Antiplatelet Therapy (PATCH): A Randomised, Open-Label, Phase 3 Trial*. Lancet. 387, no. 10038 (2016): 2605–13.

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.