

ДЭНИЕЛ ДЕННЕТ

# Насосы интуиции

*и другие инструменты  
мышления*

CoRpus



Книжные проекты  
Дмитрия Зимина

*Лучшая книга, которую мне  
довелось читать.*

Ричард Докинз

## Annotation

Новая книга знаменитого, профессора Университета Тафтса и автора выдающихся работ по философии сознания *Breaking the Spell*, *Darwin's Dangerous Idea* и *Consciousness Explained*, американского философа Дэниела Деннета (р. 1942) – вдохновенное описание мастерской и рабочих инструментов мыслителя, которое одновременно является своего рода синтетическим изложением его философии. Особое внимание автор уделяет мысленным экспериментам, которые он называет “насосами интуиции”, коротким историям, “которые призваны стимулировать искреннее, прочувствованное интуитивное озарение”. Не ограничиваясь такими хорошо известными инструментами мышления, как бритва Оккама или *reductio ad absurdum*, автор предлагает множество собственных изобретений такого рода. Его цель, как всегда, состоит в том, чтобы научить нас “серьезно и не без изящества размышлять об очень сложных вещах”. Книга, явившаяся из личного опыта философа – настоящий путеводитель как по вселенной человеческого мышления, так и по философии сознания Д. Деннета, и она будет интересна не только профессиональным философам, но и любому человеку, стремящемуся лучше разобраться в природе собственной мысли.

- [Дэниел Деннет](#)
  - 
  - 
  - [Предисловие](#)
  - [I. Введение](#)
  - [II.](#)
    - 
    - [1. Совершая ошибки](#)
    - [2. “На основе пародии”: использование \*reductio ad absurdum\*](#)
    - [3. Правила Рапопорта](#)
    - [4. Закон Старджона](#)
    - [5. Бритва Оккама](#)
    - [6. Метла Оккама](#)
    - [7. Использование обывателей в качестве ложной аудитории](#)
    - [8. Выпрыгивание](#)
    - [9. Три вида гудлинга: скорейность, нагромождения и тустеп Гудда](#)
    - [10. Оператор “безусловно”: ментальный блок](#)
    - [11. Риторические вопросы](#)
    - [12. Что такое глубокость?](#)
    - [Резюме](#)
  - [III.](#)
    - 
    - [13. Убийство на Трафальгарской площади](#)
    - [14. Старший брат из Кливленда](#)
    - [15. “Мой папа – врач”](#)
    - [16. Манифестная и научная картина мира](#)
    - [17. Народная психология](#)
    - [18. Интенциональная установка](#)
    - [19. Личностное и публичностное](#)
    - [20. Каскад гомункулов](#)
    - [21. Оператор “вроде как”](#)
    - [22. Чудо-ткань](#)
    - [23. В диспетчерской работа](#)
  - [IV.](#)
    - 
    - [24. Семь секретов силы компьютеров](#)
    - [25. Виртуальные машины](#)
    - [26. Алгоритмы](#)

- [27. Автоматизация лифта](#)
- [Резюме](#)
- [V](#)
  - [28. Пунктик насчет рыжих](#)
  - [29. Странствующий четвертачник, Земля-Двойник и гигантский робот](#)
  - [30. Радикальный перевод и кроссворд Куайна](#)
  - [31. Семантические и синтаксические движки](#)
  - [32. Болотный человек и коровакула](#)
  - [33. Два черных ящика](#)
  - [Резюме](#)
- [VI](#)
  - [34. Универсальная кислота](#)
  - [35. Библиотека Менделя: чрезвычайно большое и исчезающе малое](#)
  - [36. Гены как слова и как подпрограммы](#)
  - [37. Древо жизни](#)
  - [38. Подъем в проектное пространство: небесные крюки и подъемные краны](#)
  - [39. Компетентность без понимания](#)
  - [40. Плавающие основания](#)
  - [41. Понимают ли цикады простые числа?](#)
  - [42. Как объяснить смотровые прыжки](#)
  - [43. Опасайтесь первого млекопитающего](#)
  - [44. Когда происходит видообразование?](#)
  - [45. Вдовделы, Митохондриальная Ева и ретроспективные коронации](#)
  - [46. Циклы](#)
  - [47. Что именно глаз лягушки сообщает ее мозгу?](#)
  - [48. Скачки в пространстве Вавилонской библиотеки](#)
  - [49. Кто автор “Спамлета”?](#)
  - [50. Шум в виртуальном отеле](#)
  - [51. Херб, Элис и малыш Хэл](#)
  - [52. Мемы](#)
  - [Резюме](#)
- [VII](#)
  - [53. Два контробраза](#)
  - [54. Чутье на зомби](#)
  - [55. Зомби и зимбо](#)
  - [56. Проклятие цветной капусты](#)
  - [57. Жизненная сила: сколько это в “настоящих деньгах”?](#)
  - [58. Печальная история мистера Клапгра](#)
  - [59. Настроенная колода](#)
  - [60. Китайская комната](#)
  - [61. Падение телеклона с Марса на Землю](#)
  - [62. Самость как центр нарративной гравитации](#)
  - [63. Гетерофеноменология](#)
  - [64. Исследовательница цвета Мария: разоблачение упорного башмака](#)
  - [Резюме](#)
- [VIII](#)
  - [65. Подлый нейрохирург](#)
  - [66. Детерминистская модель: игра “Жизнь”](#)
  - [67. Камень, ножницы, бумага](#)
  - [68. Две лотереи](#)
  - [69. Инертные исторические факты](#)
  - [70. Компьютерный шахматный марафон](#)
  - [71. Конечная ответственность](#)

- [72. Сфексовость](#)
  - [73. Мальчики из Бразилии: еще один упорный башмак](#)
  - [Резюме](#)
- [IX](#)
  - [74. Сделка с дьяволом](#)
  - [75. Философия как наивная автоантропология](#)
  - [76. Махматные истины высшего порядка](#)
  - [77. 10 процентов хорошего](#)
- [X](#)
- [XI](#)
- [Приложение](#)
- [Источники](#)
- [Библиография](#)
- [Цитаты и иллюстрации, защищенные авторским правом](#)
- [notes](#)
  - [1](#)
  - [2](#)
  - [3](#)
  - [4](#)
  - [5](#)
  - [6](#)
  - [7](#)
  - [8](#)
  - [9](#)
  - [10](#)
  - [11](#)
  - [12](#)
  - [13](#)
  - [14](#)
  - [15](#)
  - [16](#)
  - [17](#)
  - [18](#)
  - [19](#)
  - [20](#)
  - [21](#)
  - [22](#)
  - [23](#)
  - [24](#)
  - [25](#)
  - [26](#)
  - [27](#)
  - [28](#)
  - [29](#)
  - [30](#)
  - [31](#)
  - [32](#)
  - [33](#)
  - [34](#)
  - [35](#)
  - [36](#)
  - [37](#)
  - [38](#)
  - [39](#)
  - [40](#)

- [41](#)
  - [42](#)
  - [43](#)
  - [44](#)
  - [45](#)
  - [46](#)
  - [47](#)
  - [48](#)
  - [49](#)
  - [50](#)
  - [51](#)
  - [52](#)
  - [53](#)
  - [54](#)
  - [55](#)
  - [56](#)
  - [57](#)
  - [58](#)
  - [59](#)
  - [60](#)
  - [61](#)
  - [62](#)
  - [63](#)
  - [64](#)
  - [65](#)
  - [66](#)
  - [67](#)
  - [68](#)
  - [69](#)
  - [70](#)
  - [71](#)
  - [72](#)
  - [73](#)
  - [74](#)
  - [75](#)
  - [76](#)
  - [77](#)
  - [78](#)
  - [79](#)
  - [80](#)
  - [81](#)
  - [82](#)
  - [83](#)
  - [84](#)
-

**Дэниел Деннет**

**Насосы интуиции и другие инструменты мышления**

© Daniel C. Dennett 2013

© З. Мамедьяров; Е. Фоменко, перевод на русский язык 2019

© А. Бондаренко, художественное оформление, макет, 2019

© ООО “Издательство Аст”, 2019

\* \* \*

*Посвящается Университету Тафтса,  
моему академическому дому*

## Предисловие

Университет Тафтса более сорока лет остается моим академическим домом. Он всегда подходил мне в самый раз: не слишком большая, но и не слишком маленькая нагрузка, прекрасные коллеги, у которых можно многому научиться, минимум академических звезд, достаточно серьезные студенты, заслуживающие внимания, но не требующие опеки день и ночь, тихая обитель, где все привержены решению проблем реального мира. С момента основания Центра когнитивных исследований в 1986 г. Университет Тафтса поддерживал мои исследования, часто избавляя меня от необходимости проходить через запутанный процесс получения грантов и давая огромную свободу работать с людьми из разных сфер, либо посещая далекие семинары, лаборатории и конференции, либо приглашая в Центр коллег-ученых. В этой книге показано, чем я занимался все эти годы.

Весной 2012-го я опробовал первый вариант нескольких глав на семинаре, который провел на кафедре философии Университета Тафтса. Я поступал так много лет, но на этот раз мне хотелось, чтобы студенты помогли сделать книгу как можно более доступной для непосвященных, поэтому я исключил всех старшекурсников и студентов-философов и пригласил на семинар только дюжину бесстрашных первокурсников, двенадцать – а точнее, тринадцать из-за административной ошибки – студентов, которые вызвались первыми. Мы с таким азартом набрасывались на каждую тему, что студенты увидели: они могут на равных спорить с профессором; я же понял, что могу объяснять все гораздо подробнее и проще. Вот список моих юных помощников, которых я благодарю за смелость, изобретательность, энергию и энтузиазм: Том Эддисон, Ник Босуэлл, Тони Каннистра, Брендан Флейг-Голдштейн, Клэр Хиршберг, Кaleb Малчик, Картер Палмер, Амар Пател, Кумар Раманатхан, Ариэль Раскоу, Николай Ренедо, Микко Силлиман и Эрик Тондре.

Второй вариант рукописи, появившийся после того семинара, прочитали мои дорогие друзья Бо Дальбом, Сью Стаффорд и Дэйл Петерсон, которые дали честные оценки написанному и сделали ряд полезных замечаний. Большинство их я учел, после чего книгу прочитали мой редактор Дрейк МакФи и его способный ассистент Брендан Карри из издательства *W. W. Norton*, которые также помогли мне улучшить текст, за что я им очень благодарен. Отдельное спасибо программному координатору Центра когнитивных исследований Терезе Сальвато, которая всячески способствовала осуществлению этого проекта и, эффективно управляя Центром, косвенным образом помогала мне, составляя график моих командировок таким образом, чтобы я мог больше времени и сил уделять созданию и применению своих инструментов мышления.

И наконец, как всегда, спасибо моей жене Сьюзен. Мы с ней – одна команда целых пятьдесят лет, и всего мы достигли вместе.

*Дэниел Деннет*

*Блю-Хилл, Мэн. Август 2012 года*

# I. Введение

## Что такое насос интуиции?

*Столярным ремеслом не займешься голыми руками, а голый мозг не годен для мышления.*

*Бо Дальбом*

Мыслить трудно. Мыслить о проблемах так трудно, что голова может разболеться от одной мысли о мысли о них. По мнению моего коллеги, нейropsychолога Марселя Кинсборна, мы считаем, что мыслить трудно, поскольку на тяжком пути к истине нас соблазняют более легкие дороги, в итоге заводящие в тупик. Чтобы мыслить, нам в первую очередь необходимо противостоять этим соблазнам. Мы то и дело сбиваемся с пути, из-за чего нам приходится одергивать себя, чтобы сосредоточиться на задаче, которая стоит перед нами. *Ох.*

Есть одна история о Джоне фон Неймане, математике и физике, который превратил идею Алана Тьюринга (теперь она называется машиной Тьюринга) в реальный электронный компьютер (теперь он называется машиной фон Неймана, такой, как ваш ноутбук или смартфон). Фон Нейман был виртуозным мыслителем, славившимся потрясающей способностью мгновенно осуществлять сложнейшие вычисления в уме. История – имеющая, как и все подобные истории, великое множество вариантов – гласит, что однажды коллега предложил фон Нейману задачу, которую можно было решить двумя способами: прибегнув к сложному и трудоемкому вычислению или же применив изящное решение из тех, что каждого заставляют воскликнуть “Эврика!”. У этого коллеги была теория: он считал, что математики всегда будут выбирать трудоемкий путь, в то время как физики (которые ленивее, но умнее) не станут очертя голову бросаться в омут, а найдут быстрое и простое решение. Какой же способ выберет фон Нейман? Задача была вполне типична: два поезда, разделенные расстоянием в 100 километров, движутся навстречу друг другу по одной колее, причем скорость одного составляет 30 километров в час, а другого – 20 километров в час. Птица, летящая со скоростью 120 километров в час, вылетает от поезда А (когда между поездами 100 километров), долетает до поезда Б, разворачивается, летит обратно к приближающемуся поезду А – и так снова и снова, пока поезда не сталкиваются. Какое расстояние пролетит птица до столкновения? “Двести сорок километров”, – почти мгновенно ответил фон Нейман. “Проклятье, – ответил его коллега. – Я думал, вы пойдете сложным путем и будете суммировать бесконечный ряд”. “Ох! – смутившись, воскликнул фон Нейман и хлопнул себя по лбу. – Так был и легкий путь!” (Подсказка: сколько пройдет времени, прежде чем поезда столкнутся?)

Одни люди, подобно фон Нейману, от природы настолько гениальны, что им не составляет труда распутывать сложнейшие загадки, в то время как другие не обладают такой скоростью мышления, но при этом наделены огромным запасом “силы воли”, которая помогает им не сворачивать с пути упорного поиска истины. Мы же, все остальные, не гении вычислений, а обычные, немного ленивые люди, тем не менее тоже стремимся понять, что происходит вокруг. Что же нам делать? Использовать инструменты мышления во всем их многообразии. Эти инструменты расширяют границы нашего воображения и способствуют концентрации, благодаря чему мы получаем возможность серьезно и даже не без изящества размышлять об очень сложных вещах. В этой книге собраны мои любимые инструменты мышления. Я планирую не просто описать их, но и применить на практике, чтобы аккуратно провести ваш разум по неизведанной территории к принципиально новым представлениям о смысле, сознании и свободе воли. Мы начнем с простых и общедоступных инструментов, которые применяются во всевозможных областях. Некоторые из них вам знакомы, но другие известны не так широко. Затем я познакомлю вас с рядом инструментов, предназначенных для конкретных целей. Они были разработаны, чтобы изучить ту или иную любопытную идею и помочь мыслителям выбраться из топкого болота, которое по-прежнему затягивает и морочит экспертов. Мы также обсудим и раскритикуем неудовлетворительные инструменты мышления и неудачные механизмы убеждения, которые могут увести вас в сторону, если вы недостаточно осторожны. Даже если вы не сумеете добраться до конечной точки моего маршрута – или вам надоест и вы решите меня покинуть, – это путешествие снабдит вас новыми способами мышления на различные темы и мышления о мышлении.

Физик Ричард Фейнман – пожалуй, еще более прославленный гений, чем фон Нейман, – обладал перwokлассным мозгом, но при этом также любил повеселиться. Нам остается только порадоваться, что



ему нравилось делиться секретами ремесла, к которым он прибегал, чтобы облегчить себе жизнь. Как бы вы ни были умны, вы становитесь еще умнее, когда идете легким путем (если, разумеется, таковой существует). Автобиографические книги Фейнмана “Вы, конечно, шутите, мистер Фейнман!” и “Какое тебе дело до того, что думают другие?” должны входить в список обязательного чтения для любого начинающего мыслителя, поскольку в них содержится множество подсказок, как справляться со сложнейшими задачами – и даже как пустить окружающим пыль в глаза, когда ничего лучшего в голову не приходит. Вдохновившись богатством полезных наблюдений из его книг и его искренностью при описании работы собственного разума, я решил попробовать свои силы в подобном проекте, пускай и не столь автобиографичном, и задался целью убедить вас думать на эти темы *по-моему*. Я приложу немало усилий, чтобы заставить вас отказаться от ряда самых твердых убеждений; при этом я не буду ничего от вас утаивать. Одна из главных моих целей – объяснять по ходу дела, чем именно я занимаюсь и зачем.

Как и все ремесленники, кузнец не может работать без инструментов, но – согласно старому (почти забытому сегодня) наблюдению – кузнецы отличаются от остальных ремесленников тем, что способны сами изготовить себе инструменты для работы. Плотники не делают пилы и молотки, портные не делают ножницы и иголки, а слесари не делают гаечные ключи, но кузнецам вполне под силу выковать молоты, клещи, наковальни и чеканы из сырьевого материала – железа. А что насчет инструментов мышления? Кто их делает? Из чего? Лучшие мыслительные инструменты предложили нам философы, которые создали их из одних лишь идей, полезных конструкций информации. Рене Декарт дал нам *прямоугольную систему координат* с осями  $x$  и  $y$ , без которой фактически нельзя было бы и помыслить о *математическом анализе* – лучшем в своем роде инструменте мышления, изобретенном одновременно Исааком Ньютоном и философом Готфридом Вильгельмом Лейбницем. Блез Паскаль дал нам *теорию вероятности*, которая позволяет без проблем определять вероятность различных исходов. Будучи талантливым математиком, преподобный Томас Байес дал нам *теорему Байеса*, которая легла в основу байесовской статистики. Однако большинство описанных в этой книге инструментов гораздо проще – это не точные, систематические математические и научные машины, а прикладные инструменты разума. Среди них:

*Ярлыки.* Порой достаточно присвоить идее запоминающееся имя, чтобы не упустить ее из виду, пока вы обдумываете ее и пытаетесь понять. Полезнее прочих, как мы увидим, предупреждающие ярлыки и сигналы, которые указывают нам на вероятные источники ошибок.

*Примеры.* Некоторые философы полагают, что использование примеров в работе – это если и не прямое жульничество, то уж точно неуместная практика; подобным образом рассуждают и писатели, когда не соглашаются, чтобы их романы иллюстрировали. Писатели гордятся, что им под силу все описать словами, а философы гордятся, что им под силу орудовать искусно созданными абстрактными обобщениями, представленными в строгом порядке и как можно более напоминающими математические доказательства. Я рад за них, но им не стоит ожидать, что я порекомендую их работы кому-либо, кроме нескольких способных студентов. Эти работы сложнее, чем требуется.

*Аналогии и метафоры.* Переноса черты одной сложной вещи на другую сложную вещь, уже (казалось бы) понятную, вы прибегаете к невероятно действенному инструменту мышления, сила которого часто сбивает мыслителей с пути, когда их воображение оказывается в плену у предательской аналогии.

*Строительные леса.* Можно покрыть крышу, построить дом или починить трубу, используя одну лишь лестницу, которую вы будете снова и снова передвигать, спускаясь и поднимаясь по ней, чтобы получать доступ к небольшому участку работ, но часто гораздо проще в самом начале выделить время на создание грубых строительных лесов, которые позволят вам быстро и безопасно двигаться вокруг всего объекта работы. Ряд самых ценных инструментов мышления, описанных в этой книге, представляет собой примеры строительных лесов, на возведение которых уходит время, но которые затем позволяют одновременно решать несколько проблем, не передвигая лестницу.

И, наконец, специфический тип мысленных экспериментов, которые я назвал *насосами интуиции*.

Неудивительно, что философы любят мысленные эксперименты. Зачем нужна лаборатория, когда ответ на вопрос можно найти с помощью искусной дедукции? Мысленные эксперименты с успехом ставило множество ученых, от Галилея до Эйнштейна, поэтому их нельзя назвать исключительно инструментом философов. Некоторые мысленные эксперименты поддаются анализу, как радикальные аргументы, часто в форме *reductio ad absurdum*<sup>[1]</sup>, когда для положений оппонента выводится формальное противоречие (абсурдный результат), которое показывает, что все эти положения не могут быть верными. Один из моих любимых примеров – приписываемое Галилею доказательство, что тяжелые объекты падают не быстрее легких (если силой трения можно пренебречь). Если бы они действительно падали быстрее, сказал он, то, поскольку тяжелый камень А падал бы быстрее легкого камня Б, то, привяжи мы камень Б к камню А, камень Б замедлил бы падение камня А, подобно тормозу. Однако связка из камней А и Б тяжелее, чем один камень А, поэтому вместе два этих камня также должны падать быстрее, чем один камень А. Мы пришли к выводу, что, если привязать камень Б к камню А, получившаяся связка будет падать одновременно и быстрее, и медленнее, чем один камень А, то есть пришли к противоречию.

Другие мысленные эксперименты не столь прямолинейны, но часто не менее эффективны: это короткие истории, которые призваны стимулировать искреннее, прочувствованное интуитивное озарение – “Ну да, конечно, иначе и быть не может!” – о доказываемом тезисе. Их я назвал *насосами интуиции*. Я предложил этот термин, когда впервые публично раскритиковал знаменитый мысленный эксперимент философа Джона Сёрла “Китайская комната” (Searle 1980; Dennett 1980), и некоторые мыслители сочли, что я использовал термин в пренебрежительном или уничижительном смысле. Как раз наоборот! Я люблю насосы интуиции. При этом одни насосы интуиции великолепны, другие сомнительны и лишь некоторые способны ввести нас в заблуждение. Насосы интуиции на протяжении столетий остаются доминирующей силой философии. Для философов они сродни басням Эзопа, которые считались прекрасными инструментами мышления еще тогда, когда философов не было вовсе<sup>[2]</sup>. Если вы изучали философию в университете, вероятно, вам знакомы такие классические примеры, как приводимый Платоном в “Государстве” миф о пещере, где прикованные цепями люди видят лишь тени предметов на стене, а также его рассказ об обучении юноши-раба геометрии, изложенный в “Меноне”. Злой демон Декарта обманым путем заставлял Декарта верить, что мир полностью иллюзорен, и это можно считать первым мысленным экспериментом о виртуальной реальности, а в естественном состоянии Гоббса жизнь была скверна, жестока и коротка. Эти эксперименты не столь знамениты, как “Мальчик и волки” или “Кузнечик и муравей” Эзопа, но тоже широко известны и разработаны для стимуляции интуиции. Миф о пещере Платона информирует нас о природе восприятия и реальности, а юноша-раб служит иллюстрацией нашего врожденного знания; злой демон представляет собой величайший генератор скепсиса, а наше возвышение над естественным состоянием, которое происходит, когда мы вступаем в общественный договор, есть суть аллегории Гоббса. Все это – лейтмотивы философии, которые настолько живучи, что студенты запоминают их на долгие годы и могут воспроизводить с удивительной точностью, даже если они забыли тонкости связанных с ними доводов и анализа. Хороший насос интуиции надежнее любой из его вариаций. Мы рассмотрим всевозможные современные насосы интуиции, включая несовершенные, с целью понять, на что они годятся, как функционируют, как их использовать и даже как их создавать.

Вот короткий и простой пример: эксцентричный тюремщик. Каждую ночь он дожидается, пока все заключенные крепко заснут, а затем отпирает все двери и надолго оставляет их открытыми. Вопрос: свободны ли заключенные? Есть ли у них *возможность* уйти? Не то чтобы. Почему? Вот другой пример: драгоценности в урне. Однажды вечером вы проходите по тротуару мимо урны, в которую выбросили кучу драгоценностей. Может показаться, что вам выпал прекрасный шанс разбогатеть, вот только в этом шансе на самом деле нет ничего прекрасного, потому что он *минимален* – крайне маловероятно, что вы заметите свой шанс и реализуете его (или вообще примете его во внимание). Два этих простых сценария стимулируют интуитивные озарения, которые иначе могут на нас и не снизойти: они подчеркивают, как важно получать информацию о реальных шансах заблаговременно, чтобы мы успевали проанализировать ее и среагировать вовремя. В своем желании получить “свободу” выбора, не подверженного – как мы предпочитаем думать – влиянию “внешних сил”, мы часто забываем, что нам не стоит стремиться к изоляции от этих сил, ведь свобода воли не *исключает* нашего погружения в богатый каузальный контекст, а напротив, *требует* его.

Надеюсь, вы чувствуете, что на эту тему можно еще многое сказать! Эти крошечные насосы интуиции прекрасно поднимают вопрос, но ничего не решают – пока что. (Далее свободе воли будет посвящен целый раздел.) Нам нужно научиться правильно использовать подобные инструменты, внимательно смотреть себе под ноги и не попадаться в ловушки. Если считать насос интуиции искусно

сработанным инструментом убеждения, нам, вероятно, будет выгодно провести инженерный анализ этого инструмента и изучить все его движущиеся компоненты, чтобы определить, что именно они делают.

Когда в 1982 г. мы с Дагом Хофштадтером писали книгу “Глаз разума”, он дал прекрасный совет по этому поводу: считайте насос интуиции инструментом со множеством настроек и “крутите все регуляторы”, чтобы проверить, будут ли стимулироваться те же интуитивные озарения при использовании различных вариантов настройки.

Давайте определим и покрутим регуляторы “эксцентричного тюремщика”. Допустим – пока не доказано обратное, – что каждый элемент имеет собственную функцию, и посмотрим, какова эта функция, заменив один элемент другим или слегка его трансформировав.

1. Каждую ночь
2. он дожидается
3. пока все заключенные
4. крепко заснут
5. а затем отпирает
6. все двери
7. и надолго оставляет их открытыми.

Вот один из множества доступных нам вариантов:

Однажды ночью он приказал надзирателю дать снотворное одному из заключенных, а тот, выполнив приказ, случайно оставил на один час дверь в камеру этого заключенного незапертой.

Сценарий изменился достаточно сильно, не так ли? Каким образом? Он по-прежнему иллюстрирует основную мысль (так ведь?), но уже не столь эффективно. Особенно важна разница между естественным сном – когда проснуться *можно* в любую минуту – и сном под действием снотворного или коматозным состоянием. Другое отличие – “случайно” – подчеркивает роль намерения или недосмотра со стороны тюремщика или надзирателей. Повторение (“каждую ночь”), казалось бы, меняет вероятность в пользу заключенных. Когда и почему важна вероятность? Сколько вы готовы заплатить, чтобы *не участвовать* в лотерее, билеты которой имеет миллион человек, а “победителя” расстреливают? Сколько бы вы заплатили, чтобы *не играть* в русскую рулетку с шестизарядным револьвером? (Здесь мы используем один насос интуиции, чтобы проиллюстрировать другой. Этот трюк стоит запомнить на будущее.)

Другие регуляторы не столь очевидны: злобный хозяин тайно запирает двери в спальни своих гостей, пока те спят. Старшая медсестра, опасаясь пожара, ночью оставляет двери всех палат и смотровых незапертыми, но не сообщает об этом пациентам, полагая, что они будут спать крепче, если им об этом не говорить. А что если размеры тюрьмы несколько больше обычного – скажем, со всю Австралию? Невозможно запереть или отпереть все двери в Австралии. Что это меняет?

Настороженность, с которой мы должны подходить к каждому насосу интуиции, тоже представляет собой важный инструмент мышления, любимую тактику философов – “выход на метауровень”, то есть мышление о мышлении, разговоры о разговорах, рассуждения о рассуждениях. Метаязык – это язык, который мы используем, чтобы говорить о другом языке, а метаэтика – это отстраненный анализ этических теорий. Однажды я сказал Дагу: “Что бы ты ни сделал, я могу перевести это на метауровень”. Вся эта книга, само собой, является примером перехода на метауровень: в ней описывается, как обстоятельно мыслить о методах обстоятельного мышления (о методах обстоятельного мышления и т. д.)

[3]. Недавно (в 2007 г.) Даг составил список своих любимых подручных инструментов:

поиск ветра в поле  
безвкусица  
темные дела  
зелен виноград  
работа на износ  
слабые места  
пороховые бочки  
бредовые идеи  
пустые слова  
пара пустяков

## обратная связь

Если эти выражения вам знакомы, они для вас не “просто слова” – каждое из них представляет собой абстрактный когнитивный инструмент, подобный делению в столбик или определению среднего арифметического, и каждое играет свою роль в широком спектре контекстов, упрощая формулировку гипотез для проверки, распознавание неочевидных алгоритмов мира, поиск важных сходств и так далее. Каждое слово в вашем лексиконе – простой инструмент мышления, но некоторые слова полезнее других. Если какие-либо из этих выражений не входят в ваш словарь, вам стоит их добавить, поскольку таким образом вы сможете обдумывать мысли, сформулировать которые иначе было бы довольно сложно. Само собой, как гласит известный закон, когда у вас есть только молоток, вокруг вы видите только гвозди, поэтому ни один из этих инструментов не застрахован от чрезмерного использования.

Давайте разберем всего один: зелен виноград. Это выражение использовано в басне Эзопа “Лисица и виноград”, где описывается, как порой люди притворяются, что им нет дела до того, что они не могут получить, высказываясь об этом с пренебрежением. Можно очень многое сказать о только что услышанном от человека, просто спросив его: “Что, зелен виноград?” Собеседнику придется рассмотреть возможность, которая в ином случае, вероятно, осталась бы незамеченной, и это может сподвигнуть его мыслить иначе или взглянуть на проблему шире, а может и просто сильно его оскорбить. (Инструменты могут использоваться и в качестве оружия.) Мораль сей басни так хорошо знакома любому, что не обязательно помнить всю историю и держать в уме ее тонкости, которые порой и вовсе не важны.

Овладеть инструментами и использовать их с умом – это разные навыки, однако начать надо с овладения имеющимися инструментами или создания собственных. На страницах этой книги я опишу многие инструменты, которые изобрел сам, но упомяну и те, которые позаимствовал у других мыслителей, и каждому из них я отдам должное<sup>[4]</sup>. Даг не изобрел ни один из инструментов из собственного списка, но принял участие в разработке некоторых прекрасных образчиков из моего набора, таких как **выпрыгивание** и **сфексовость**.

Иные из наиболее действенных инструментов мышления имеют математическую природу, но я ограничусь лишь их упоминанием, поскольку эта книга отдает дань нематематическим инструментам, неформальным инструментам, если угодно, инструментам прозы и поэзии, силу которых ученые часто недооценивают. И вполне понятно, почему. Прежде всего, в научных журналах сложилась культура научного стиля, которая отдает предпочтение и даже настаивает на беспристрастном, неприукрашенном изложении проблем с минимумом словесных изыществ, риторики и аллюзий. Неумолимому однообразию на страницах наиболее серьезных научных журналов есть хорошее объяснение. Как в 1965 г. написал мне один из моих оппонентов, специалист по анатомии нервной системы Дж. З. Янг, раскритиковавший несколько замысловатый язык моей оксфордской диссертации (по философии, не по анатомии нервной системы), английский язык становится международным языком науки, а потому нам, носителям английского, следует писать работы, которые под силу прочесть “терпеливому китайцу, вооруженному хорошим словарем”. Результаты этой самодисциплины говорят за себя: неважно, из Китая вы, из Германии, из Бразилии – или даже из Франции, – свои наиболее важные работы вы все равно предпочитаете публиковать на английском, элементарном английском, который без особенных проблем поддается переводу и минимально опирается на культурные аллюзии, нюансы, игру слов и метафоры. Уровень взаимопонимания, достижимый в этой международной системе, не оценим, но платить за него все же приходится: некоторые аспекты мышления явно требуют неформального жонглирования метафорами и стимуляции воображения, хитроумной атаки на баррикады закрытого разума, и если что-то из этого невозможно перевести без проблем, то мне остается лишь надеяться на виртуозное мастерство переводчиков, с одной стороны, и растущую свободу владения английским в научной среде, с другой.

Вторая причина, по которой ученые часто с подозрением относятся к теоретическим дискуссиям, ведущимся “одними словами”, заключается в том, что они понимают: критиковать аргумент, не сформулированный на языке математических уравнений, гораздо сложнее, а результаты этой критики, как правило, не столь убедительны. Язык математики обеспечивает неопровержимость. Он подобен сетке на баскетбольном кольце, которая исключает споры насчет того, попал ли мяч в цель. (Любой, кто играл в баскетбол на площадке с кольцом без сетки, знает, как сложно бывает определить, влетел ли мяч в кольцо, если он не задел ни само кольцо, ни щит.) Но порой вопросы настолько сложны и неочевидны, что математики просто не могут с ними совладать.

Я всегда считал, что если я не могу объяснить, чем занимаюсь, группе способных студентов бакалавриата, то я сам это не до конца понимаю. Такое убеждение повлияло на все мои труды. Некоторые

профессора философии предпочитают проводить семинары продвинутого уровня только для студентов магистратуры. Но не я. Студенты магистратуры часто слишком хотят доказать друг другу и самим себе, что они подкованы в соответствующей области, а потому уверенно сыплют профессиональными жаргонизмами, сбивая с толку непосвященных (так они убеждают самих себя, что их занятие требует опыта) и хвастаясь своей способностью следить за нитью самых заумных (и мучительных) формальных доводов, не заходя при этом в тупик. Философия, написанная для продвинутых студентов магистратуры и коллег-экспертов, как правило, совершенно нечитаема – а потому ее почти никто и не читает.

Любопытным побочным эффектом моего стремления описывать аргументацию и объяснения языком, который без труда могут понять люди, не имеющие отношения к кафедрам философии, стало появление философов, “из принципа” не желающих принимать мои аргументы всерьез! Когда много лет назад я читал в переполненной оксфордской аудитории цикл лекций в рамках “Лекций Джона Локка”, уважаемый философ покинул одну из них, ворча, что будь он проклят, если сумеет узнать хоть что-то от человека, который привлек на “Лекции Локка” людей, далеких от философии! Насколько я могу судить, он остался верен своему слову и ничего от меня не узнал. Я не изменил свой стиль и ни разу не пожалел о необходимости за это платить. В философии есть время и место для строгих выкладок, где пронумерованы все предпосылки и озвучены правила логических выводов, однако часто все это не стоит демонстрировать широкой публике. Мы просим студентов магистратуры доказывать свои умения в диссертациях – и, к несчастью, некоторые из них так и не отказываются от этой привычки. При этом стоит заметить, что полярно противоположный грех злоупотребления высокопарной европейской риторикой, обильно сдобренной литературными украшениями и намеками на глубокомыслие, тоже не идет философии на пользу. Если бы мне пришлось выбирать, я бы всегда отдавал предпочтение упрямым аналитикам, готовым приводить заумные аргументы, а не склонным к словоблудию мудрецам. В разговоре с аналитиком обычно хотя бы можно понять, что он имеет в виду и что считает неверным.

Полагаю, чтобы внести максимально полезный вклад в науку, философу необходимо нащупать золотую середину между поэзией и математикой, предоставляя понятные объяснения головоломных задач. Для подобной работы не существует четких алгоритмов. Поскольку доступно *все*, каждый внимательно выбирает ориентиры. Нередко “невинное” допущение, неосмотрительно сделанное всеми сторонами, оказывается источником ошибки. Исследованию столь опасных концептуальных территорий существенно помогает использование инструментов мышления, разрабатываемых специально, чтобы объяснить альтернативные пути и пролить свет на их перспективы.

Такие инструменты мышления редко устанавливают *фиксированный* ориентир – твердую “аксиому” для всех будущих изысканий, – но гораздо чаще намечают *возможный* ориентир, вероятный ограничитель будущих изысканий, который сам может быть подвергнут пересмотру или вовсе отброшен, если кто-нибудь сумеет найти для этого повод. Неудивительно, что многие ученые не питают симпатии к философии: все доступно, ничего не определено на сто процентов, а сложнейшие сети аргументов, создаваемые, чтобы связать эти ориентиры, фактически висят в воздухе, лишены явного фундамента эмпирических доказательств или опровержений. В результате такие ученые поворачиваются к философии спиной и продолжают собственную работу, но при этом оставляют нерассмотренным целый ряд важнейших и любопытнейших вопросов. “Не спрашивай! Не говори! Преждевременно обсуждать проблему сознания, свободы воли, этики, смысла и креативности!” Но мало кому под силу практиковать такое воздержание, поэтому в последние годы ученые устремились в сферы, которые прежде ими игнорировались. Из чистого любопытства (а иногда, возможно, из жажды славы) они берутся за серьезные вопросы и вскоре выясняют, сколь сложно делать подвижки в этих областях. Должен признаться (хоть в такие моменты я и чувствую себя виноватым), что мне все же приятно наблюдать, как уважаемые ученые, которые всего несколько лет назад заявляли о своем презрении к философии<sup>[5]</sup>, то и дело оступаются, когда пытаются совершить переворот в этих сферах, прибегая к аргументированным экстраполяциям из собственных научных исследований. Еще забавнее получается, когда они обращаются за помощью к нам, философам, и впоследствии эту помощь принимают.

В первом разделе этой книги я опишу дюжину общих, универсальных инструментов, а в последующих разделах сгруппирую остальные главы не по типу инструмента, а по теме, в которой этот инструмент работает лучше всего. Начну я с фундаментальной философской проблемы – проблемы значения, или содержания, – а затем перейду к эволюции, сознанию и свободе воли. Некоторые из описываемых инструментов представляют собой программное обеспечение, удобные устройства, которые помогают нашему невооруженному воображению, подобно тому, как микроскопы и телескопы помогают невооруженному глазу.

На страницах этой книги я также познакомлю вас с несколькими ложными друзьями – инструментами, которые напускают тумана, вместо того чтобы проливать свет. Мне нужен был термин для обозначения этих опасных устройств, и я нашел подходящие слова, обратившись к своему опыту хождения под парусом. Многие моряки любят вернуть морские словечки, которые сбивают с толку сухопутных крыс: правый и левый борт, гужон и шкворень, ванты и отводы, люверсы, направляющие и все остальное. Однажды я шел на судне, где моряки то и дело придумывали шуточные определения этим терминам. *Нактоузом* мы стали называть наросты морских организмов на компасах, *бугель с гаком* превратился в цитрусовый напиток, который полагается пить на палубе, *канифас-блоком* мы назвали женский защитный маневр, а *упорным башмаком* – упрямое ортопедическое приспособление. С тех пор выражение “упорный башмак” – которым на самом деле обозначается подвижная подставка для гика, используемая при спускании паруса, – всегда ассоциируется у меня с беднягой в ортопедическом сапоге, не желающем идти куда следует. Именно поэтому я и решил назвать так инструменты мышления, которые оборачиваются против мыслителя: хотя и *кажется*, что они помогают достичь понимания, на самом деле они лишь заводят мыслителя во мрак, вместо того чтобы вести к свету. В последующих главах будет описано множество упорных башмаков, которым присвоены соответствующие предупреждающие ярлыки, и приведен целый ряд печальных примеров. В завершение я еще немного порассуждаю о том, что значит быть философом, если это кому-нибудь интересно, и дам несколько советов от дядюшки Дэна каждому из вас, кто почувствовал вкус к такому способу познания мира и гадает, сможет ли он построить карьеру в этой области.

## **II.**

### **Дюжина универсальных инструментов мышления**

Большинство инструментов мышления, описываемых в этой книге, довольно специфичны: они разработаны для применения к конкретной теме и даже к конкретной проблеме в рамках этой темы. Но прежде чем обратиться к этим насосам интуиции, позвольте познакомить вас с несколькими универсальными инструментами мышления, идеями и практиками, которые доказали свою состоятельность в широком спектре контекстов.

## 1. Совершая ошибки

*Тот, кто говорит: “Лучше вовсе ни во что не верить, чем верить в ложь!” – только показывает господствующий внутри него страх показаться глупцом... Он подобен генералу, который говорит солдатам, что лучше вообще не ввязываться в битву, чем рисковать получить рану. Не так одерживаются победы над нашими врагами и над нашей природой. Наши ошибки не так уж серьезны. В мире, где мы обречены совершать их, несмотря на все предосторожности, лучше не принимать их близко к сердцу и не проявлять излишнего беспокойства по их поводу.*

*Уильям Джеймс, “Воля к вере”*

*Если вы твердо решили проверить теорию или хотите объяснить какую-то идею, публикуйте свои результаты, какими бы они ни были. Публикуя результаты выборочно, мы подкрепляем свои аргументы. Однако публиковать следует все результаты без исключения.*

*Ричард Фейнман, “Вы, конечно, шутите, мистер Фейнман”*

Ученые часто спрашивают меня, почему философы тратят столько сил на изучение и преподавание истории своей науки. Химикам, как правило, достаточно лишь зачаточных знаний об истории химии, которые они подхватывают на ходу, а специалисты по молекулярной биологии, казалось бы, и вовсе не интересуются, что происходило в биологии года до 1950-го. Мой ответ таков: история философии по большей части представляет собой историю величайших ошибок очень умных людей, а не зная истории, ты обречен совершать эти проклятые ошибки снова и снова. Именно поэтому мы, философы, преподаем студентам историю философии, а ученые, которые беспечно отмахиваются от философии, делают это на свой страх и риск. Нет науки, свободной от философии, есть лишь наука, которую творят без оглядки на лежащие в ее основе философские допущения. Самые умные и самые везучие из ученых порой умело обходят все ловушки (возможно, они “прирожденные философы” или действительно так умны, как сами думают), но таких крайне мало. При этом нельзя сказать, что профессиональные философы не совершают старых ошибок и не отстаивают их правильность. Если бы вопросы не были столь сложны, о них не стоило бы и задумываться.

Порой хочется не просто *рискнуть* совершить ошибку, а по-настоящему совершить ее, даже если только ради того, чтобы стало понятно, что именно нужно исправить. Ошибки – ключ к прогрессу. Само собой, бывают моменты, когда важно не совершать ошибок: спросите любого хирурга или пилота. Но далеко не все понимают, что бывают и моменты, когда ошибки становятся единственным способом двигаться вперед. Многие студенты, которые поступают в престижные университеты, гордятся, что не совершают ошибок, ведь именно так они сумели продвинуться дальше своих одноклассников – во всяком случае, так им всегда говорили. Я часто замечаю, что мне приходится призывать их *выработать привычку* совершать ошибки, поскольку именно ошибки таят в себе наибольший потенциал для обучения. Студенты сталкиваются с “творческим кризисом” и часами в отчаянии бьются над первым предложением эссе. “Выкладывайте как есть!” – советую им я. Тогда у них появляется хоть какой-то материал для работы.

Мы, философы, специалисты по ошибкам. (Понимаю, это напоминает начало плохой шутки, но выслушайте меня до конца.) В то время как другие науки специализируются на получении правильных ответов на основополагающие вопросы, мы, философы, специализируемся на том, чтобы запутывать все на свете и переворачивать с ног на голову, в результате чего никто не может даже сформулировать верные *вопросы*, не то что уж дать на них ответы. При этом, задавая неверные вопросы, мы рискуем направить все изыскания по ложному следу. Когда такое случается, нужно обращаться к философам! В любой области, в любом исследовании приходится заниматься философией, чтобы выяснить, какие вопросы нужно задавать. Некоторые люди терпеть этого не могут. Они предпочитают находить готовые вопросы – ладно скроенные, безупречно отглаженные, готовые для ответа. Таким людям лучше заниматься физикой, математикой, историей или биологией. Работы всем хватит. Мы, философы, любим работать с вопросами,



которые приходится приводить в порядок, прежде чем на них отвечать. Это занятие не для всех. Но попробуйте – вдруг понравится?

В этой книге я буду ожесточенно набрасываться на то, что считаю ошибками других людей, но будьте уверены, опыта в совершении ошибок мне не занимать. Я попадался в ловушки и надеюсь, что попадусь еще не раз. Одна из целей этой книги – помочь вам совершать *хорошие* ошибки, которые освещают всем нам путь вперед.

Сначала теория, затем практика. Ошибки не просто дают нам возможность учиться; фактически они дают нам *единственную* возможность учиться и создавать нечто поистине новое. Чтобы учиться, нужны ученики. Ученики могут появиться лишь двумя способами, в которых нет места магии: они должны либо развить себя сами, либо вырасти под влиянием тех учеников, которые уже достигли нужного уровня развития. Биологическая эволюция предполагает великий, неизбежный путь проб и ошибок – и без *ошибок* от проб нет никакого толку. Гор Видал однажды сказал: “Добиться успеха недостаточно. Остальные должны потерпеть поражение”. Пробы могут быть *слепыми* или *прогнозными*. Зная многое, но не зная ответа на конкретный вопрос, вы можете совершать скачки – прогнозные скачки. Вы можете осмотреться перед скачком и далее *в некоторой степени* опираться на полученные ранее знания. Вам нет нужды гадать, но не стоит свысока смотреть на догадки, ведь среди их великолепных результатов... вы сами!

Эволюция стала одной из главных тем этой книги, как и всех остальных моих книг, по той простой причине, что она представляет собой важнейший, полезнейший процесс не только самой жизни, но и знания, обучения и понимания. Если вы попытаетесь понять мир идей и смыслов, свободы воли и морали, искусства, науки и даже самой философии, не имея основательных и достаточно подробных знаний об эволюции, вам придется работать одной рукой, держа вторую за спиной. Позже мы рассмотрим некоторые инструменты, призванные помочь вам размышлять над самыми сложными вопросами эволюции, но для этого сначала необходимо заложить крепкий фундамент. Для эволюции, которая ничего не знает, шаги к новому вслепую совершают мутации, представляющие собой случайные “ошибки” при копировании ДНК. Большинство таких ошибок не имеет никакого значения, ведь их никто не видит! Они столь же несущественны, как черновики ваших студенческих работ, которые вы не сдавали (или *не сдаете*) преподавателю на оценку. ДНК вида напоминает рецепт создания нового тела, и большая часть ДНК даже не играет роли в процессе формирования организма. (Именно поэтому ее часто называют “мусорной ДНК”.) Подавляющее большинство мутаций в тех цепочках ДНК, которые считываются и используются в процессе развития, наносит вред организму; многие из них и вовсе быстро становятся фатальными. Поскольку большая часть “проявляющихся” мутаций вредоносна, процесс естественного отбора работает таким образом, чтобы количество мутаций оставалось на минимальном уровне. В клетках каждого из вас работает очень, очень хороший копировальный механизм. К примеру, ваше тело состоит примерно из триллиона клеток и каждая клетка представляет собой либо идеальную, либо почти идеальную копию вашего генома длиной более трех миллиардов символов – тот самый рецепт, который впервые появился, когда во взаимодействие вступили яйцеклетка и сперматозоид ваших родителей. К счастью, копировальный механизм работает не безупречно, ведь в ином случае эволюция в конце концов остановилась бы, поскольку ее источники новизны оказались бы исчерпаны. Эти крошечные недостатки, эти “несовершенства” процесса ответственны за потрясающее устройство и сложную структуру мира живой природы. (Я не могу не добавить: если что-то и заслуживает именоваться первородным грехом, так это эти ошибки копирования.)

Чтобы совершать хорошие ошибки, главное не скрывать их – особенно от себя самого. Вместо того чтобы отрешиваться от ошибок, нужно стать их настоящим ценителем – нужно рассматривать их в качестве произведений искусства, ведь в некотором роде они и правда шедевральны. Основополагающей реакцией на любую ошибку должно быть восклицание: “Что ж, *такого* я больше не сделаю!” Естественный отбор не думает; он просто устраняет неудачные экземпляры, не давая им шанса на размножение; естественный отбор не повторяет *таких* ошибок, по крайней мере не так часто. У способных к обучению животных, которым под силу научиться издавать конкретный звук, касаться конкретного провода или есть конкретную пищу, в мозгу имеется *нечто*, обладающее сходной селективной силой. (Б. Ф. Скиннер и другие бихевиористы понимали необходимость этого и называли это обучением “с подкреплением”: *такой* ответ не получает подкрепления и “пресекается”.) Мы, люди, выводим все на новый уровень, повышая скорость и эффективность реакции. Мы можем *думать*, оценивать только что сделанное: “Что ж, *такого* я больше не сделаю!” Размышляя, мы сталкиваемся с проблемой, которую приходится решать любому, кто совершает ошибки: что именно подразумевается под “таким”? Какое именно мое действие привело к возникновению затруднений? Главное – изучить

конкретные детали возникшей по вашей вине путаницы, чтобы при следующей попытке учесть усвоенные уроки и не делать новый шаг в крошечной тьме.

Все мы слышали горькое восклицание: “Мне казалось, это сработает!” Эта фраза стала символом раскаяния идиота, признаком глупости, но на самом деле мы должны считать ее столпом мудрости. Любой человек, любой агент, который может искренне сказать: “Мне казалось, это сработает!” – стоит на пороге гениальности. Мы, люди, гордимся собственным интеллектом, а одна из его особенностей заключается в том, что мы *помним наши прошлые мысли* и можем давать им оценку: определять, какими они нам казались, почему они были столь привлекательны и что в итоге пошло не так. У меня нет свидетельств, которые позволили бы предположить, что еще хоть один вид на планете способен к такому же мышлению. Если бы это было так, этот вид был бы почти столь же умен, как мы сами.

В связи с этим вы должны научиться глубоко вздохнуть после ошибки, стиснуть зубы и *изучить* свои воспоминания о произошедшем как можно более внимательно и бесстрастно. Это непросто. Совершая ошибку, человек естественным образом чувствует смущение и гнев (сильнее всего мы злимся на самих себя), поэтому нужно усердно работать, чтобы преодолеть эти эмоциональные реакции. Попробуйте привить себе необычную привычку наслаждаться своими ошибками и с удовольствием изучать все странности, которые сбили вас с пути. В таком случае, выжав из них всю пользу, вы сможете радостно оставить их позади и перейти к следующей большой возможности. Но этого недостаточно: нужно активно искать шансы совершать великие ошибки, чтобы потом после них восстанавливаться.

Самую простую технику совершения ошибок мы все изучили в начальной школе. Помните, каким страшным и непонятным казалось нам сначала деление в столбик? Вам давали два невыразимо крупных числа, и нужно было понять, с чего начать. Сколько делителей содержится в делимом – шесть? семь? восемь? Как знать? Но знать и не нужно было – нужно было просто назвать число наугад и проверить результат. Помню, я был поражен, когда мне сказали, что начать нужно с “догадки”. И это *математика*? Разве в столь серьезных делах есть место угадкам? Но в конце концов мы все оценили красоту такой тактики. Если выбранное число оказывалось слишком маленьким, его нужно было увеличить и начать заново; если же оно было слишком большим, его приходилось уменьшить. В делении столбиком был огромный плюс: оно работало всегда, даже если сначала ты выбирал максимально дурацкое число – в таком случае просто приходилось повозиться подольше.

Эта техника совершения более или менее осознанного выбора, оценки его следствий и использования результата для внесения поправок нашла множество применений. Главное в этой тактике – сделать ошибку, которая достаточно понятна и точна, чтобы привести к определенным последствиям. Пока не появилась система GPS, штурманы определяли, где находится судно, высказывая предположение относительно его местоположения (они наугад называли *точные* значения широты и долготы), а затем рассчитывая, на какой именно высоте должно находиться солнце, если они – по невероятному совпадению – действительно находятся в этой точке. Используя этот метод, они не ожидали, что сразу правильно определяют свою позицию. В этом не было необходимости. Вместо этого они измеряли фактическую высоту солнца над горизонтом (точно) и сравнивали два значения. Дальнейшие нехитрые расчеты позволяли им определить, какие поправки необходимо внести в их изначальную догадку<sup>[6]</sup>. При использовании этого метода полезно изначальное высказывание достаточно точное предположение, однако вовсе неважно, что оно обречено быть ошибочным. Главное – сделать ошибку как можно точнее, чтобы затем ее можно было исправить. (Устройство GPS применяет ту же стратегию догадок и поправок, чтобы определять свое местоположение относительно спутников.)

Само собой, чем сложнее задача, тем сложнее и анализ. Исследователи искусственного интеллекта (ИИ) называют эту проблему “присваиванием коэффициентов доверия” (но ее вполне можно было бы назвать и “присваиванием коэффициентов осуждения”). Определить, чему доверять, а что осуждать, – одна из сложнейших задач, стоящих перед ИИ. С такой же проблемой сталкивается и естественный отбор. Каждый организм на земле рано или поздно умирает, когда к концу подходит та или иная сложная история жизни. Как вообще естественный отбор может смотреть сквозь туман всех этих деталей, чтобы понять, какие позитивные факторы “вознаградить” потомством, а какие негативные факторы “наказать” его отсутствием? Неужели некоторые братья и сестры наших предков действительно умирали бездетными, *потому что у них были веки неправильной формы*? Если нет, как процесс естественного отбора может объяснить, почему наши веки имеют итоговую превосходную форму? Один аспект ответа нам знаком: известная присказка “пока работает, не трожь” советует нам не вносить изменения в старые, консервативные дизайнерские решения и не рисковать без страховки. Естественный отбор автоматически сохраняет *все, что работало до этого момента*, и бесстрашно исследует крупные и мелкие инновации, причем крупные почти всегда ведут прямо к смерти. Такая стратегия очень расточительна, но никто

не обращает на это внимания. Естественный отбор по большей части определил форму наших век задолго до появления человека, приматов и даже млекопитающих. Чтобы сформировать веки такими, какие они сегодня, естественный отбор трудился более сотни миллионов лет, но в последние шесть миллионов лет в проект вносились лишь незначительные изменения, поскольку у нас были общие предки с шимпанзе и бонобо. Другой аспект ответа заключается в том, что естественный отбор работает с огромным числом случаев, где даже минимальные преимущества проявляются в статистике и могут накапливаться автоматически. (Остальные аспекты ответа на этот вопрос предполагают обсуждение технических тонкостей, избыточное для этой упрощенной дискуссии.)

Вот техника, которую блестяще используют фокусники, работающие с картами, – по крайней мере, лучшие из них. (Полагаю, я не навлеку на себя гнев фокусников за объяснение этого трюка, ведь я не буду раскрывать загадку *конкретного* фокуса, а изложу основополагающий общий принцип.) Хороший фокусник знает много фокусов, которые зависят от везения: они получаются не всегда и даже не часто. Некоторые приемы – их даже фокусами не назовешь – работают и вовсе только раз из тысячи! Вот как с этим справиться: сначала нужно сказать зрителям, что сейчас вы покажете какой-то фокус, и, не уточняя, какой именно фокус вы показываете, попробовать прием, срабатывающий в одном случае из тысячи. Само собой, он вряд ли сработает, поэтому вы спокойно перейдете ко второй попытке – и попробуете прием, который срабатывает один раз из сотни, – а когда вам не повезет и с ним (что произойдет почти наверняка), вы спокойно перейдете к приему номер 3, который срабатывает в одном случае из десяти, но при этом наготове будете держать прием номер 4, срабатывающий в половине случаев (к примеру). Если ничего не сработает (а к этому моменту *обычно* одна из страховок уже спасает вас от худшего), вы обратитесь к беспроигрышному приему, который не слишком впечатлит зрителей, но хотя бы гарантирует выполнение фокуса. В ходе целого представления вам вряд ли придется всегда полагаться на последнюю страховку – для этого вы должны быть невероятно невезучим, – а всякий раз, когда у вас получится один из более сложных трюков, зрители будут изумляться. “Невероятно! Как вы вообще могли узнать, какую я выбрал карту?” *Aga!* Вы этого и не знали, но нашли изящный способ ткнуть пальцем в небо – и не прогадали. Скрывая все “ошибки” из виду – не показывая пробы, которые не принесли желаемого результата, – вы творите “чудо”.

Таким же образом работает и эволюция: все глупые ошибки, как правило, остаются невидимыми, так что мы наблюдаем лишь поразительную серию триумфов. К примеру, подавляющее большинство – гораздо больше 90 процентов – всех существ, когда-либо существовавших на земле, умерли, не оставив потомства, но *ни одного из ваших предков* не постигла такая участь. Похоже, ваш род неуязвим!

Важное различие между дисциплиной науки и дисциплиной фокусов заключается в том, что фокусники всячески стараются скрыть свои ошибки от зрителей, а ученые совершают ошибки публично. В науке ошибки выставляют на обозрение, чтобы все могли на них учиться. Таким образом вы получаете возможность использовать чужой опыт, а не просто идти своим особым путем в пространстве ошибок. (Физик Вольфганг Паули однажды презрительно высказался о работе коллеги, назвав ее “даже не ошибочной”. Лучше поделиться с критиками явной ошибкой, чем несусветной бессмыслицей.) Кстати, это еще одна причина, объясняющая, почему люди настолько умнее всех остальных видов. Дело не столько в том, что у нас более крупный или более развитый мозг, и даже не в том, что мы умеем размышлять о своих прошлых ошибках, а в том, что мы *делимся преимуществами*, которые каждый из наших мозгов получает, когда проходит по собственному пути проб и ошибок<sup>[1]</sup>.

Меня удивляет, сколько поистине умных людей не понимает, что можно у всех на виду совершать серьезные ошибки, но при этом не терять своей репутации. Я знаю уважаемых ученых, которые готовы пойти на все, лишь бы только не признавать, что они в чем-то ошиблись. Очевидно, они никогда не замечали, что земля не разверзается под ногами у людей, когда те говорят: “Ой, вы правы. Похоже, я допустил ошибку”. На самом деле люди *любят*, когда кто-то признает свою ошибку; все обожают указывать на ошибки других. Великодушные люди ценят, когда вы даете им возможность помочь и благодарите их за помощь, если она оказывается кстати, а мелочные люди не упустят повода указать на ваш промах. И пусть! Как бы то ни было, все мы остаемся в выигрыше.

Само собой, как правило, люди не любят исправлять *глупые* ошибки других. Чтобы ошибку исправили, она должна быть *стоящей*: вам стоит найти оригинальную позицию, которая может оказаться как верной, так и неверной, построив целую пирамиду рискованных размышлений, подобную пирамиде рискованных фокусов. Положив в основу своей работы наработки других, вы можете продвинуться дальше. И вас ждет неожиданный бонус: если вы готовы на серьезный риск, люди примутся исправлять ваши случайные глупые ошибки, тем самым показывая, что вы не исключительны и тоже порой выдаете брак, как и все остальные. Я знаю невероятно осторожных философов, которые никогда – насколько это

известно — не совершали ошибок в своих работах. Как правило, они не слишком плодовиты, но немногочисленные выходящие из-под их пера работы обычно безупречны, хотя рискованными их не назовешь. Они специализируются на том, что указывают на ошибки других, и это довольно ценно, однако *им самим* никто не спускает с рук и малейшей оплошности. К несчастью, можно сказать, что их лучшие труды часто остаются незамеченными и непризнанными, поскольку они теряются в потоке громких работ более смелых мыслителей. В главе 76 мы увидим, что обычно полезная практика совершения смелых ошибок тоже имеет свои неблагоприятные побочные эффекты. Метасовет: не относитесь *ни к каким* советам с излишней серьезностью!

## 2. “На основе пародии”: использование *reductio ad absurdum*

Против лома нет приема, а ломом рационального анализа, великим рычагом, обеспечивающим непротиворечивость, служит *reductio ad absurdum* – доведение (суждения) до абсурда. При использовании этого приема вы берете утверждение или предположение и смотрите, можно ли обнаружить в нем какое-либо противоречие (или хотя бы противоречащее здравому смыслу следствие). Если у вас получается найти противоречие, это суждение отбрасывается или хотя бы отправляется на пересмотр. Мы постоянно прибегаем к этому методу, не задумываясь о лежащей в его основе логике: “Если это медведь, то у медведей есть рога!” или “Он не успеет к ужину, если только не прилетит сюда, как Супермен”. Когда дело касается сложных теоретических противоречий, рычаг непротиворечивости получает мощный энергетический заряд, но провести черту между справедливой критикой и карикатурным опровержением аргумента очень сложно. Неужели ваш оппонент действительно настолько глуп, чтобы считать истинным утверждение, которое вы только что довели до абсурда несколькими ловкими маневрами? Однажды я проверял работу, в которой студент допустил прелюбопытную опisku, заменив “аналогию” “пародией”, в результате чего получилась прекрасная фраза “на основе пародии”. На мой взгляд, это удачное название для презренных споров с использованием *reductio ad absurdum*, которые так часто вспыхивают в беспорядке научной и философской полемики.

Помню, несколько лет назад я присутствовал на семинаре по когнитивистике, который проводился в MIT лингвистом Ноамом Хомским и философом Джерри Фодором. В ходе семинара собравшихся щедро потчевали развенчанием работ других когнитивистов, не встречавших их одобрения. В тот день все копы были направлены против директора Лаборатории искусственного интеллекта Йельского университета Роджера Шенка, который – в описании Хомского – представлял полным идиотом. Я довольно хорошо знал Роджера и был знаком с его работой. Хотя я и не был полностью согласен с его точкой зрения, в изложении Ноама ее сложно было узнать, поэтому я поднял руку и предположил, что он, возможно, не разобрался во всех нюансах позиции Роджера. “О нет, – усмехнувшись, возразил Ноам. – Именно так он и считает!” После этого, к великому удовольствию собравшихся, он продолжил свой разбор. Через несколько минут я вмешался снова. “Должен признать, что подвергающиеся критике взгляды просто нелепы, – сказала я, и Ноам одобрительно усмехнулся, – но в таком случае зачем же вы тратите время, критикуя этот абсурд?” Это замечание возымело действие.

Что насчет меня самого? Довожу ли я до абсурда чужие идеи? Справедливее ли я при этом? Вот несколько примеров. Решать вам. Однажды на конференции в Венеции мы с французским нейробиологом Жан-Пьером Шанжэ сошлись с нейрофизиологом сэром Джоном Экклсом и философом сэром Карлом Поппером в споре о сознании и мозге. Мы с Шанжэ были материалистами (которые полагают, что сознание *и есть* мозг), а Поппер и Экклс – дуалистами (которые утверждают, что сознание нельзя считать материальной сущностью, подобной мозгу, поскольку оно представляет собой сущность второго порядка, взаимодействующую с мозгом). За много лет до этого Экклс получил Нобелевскую премию за открытие синапса – микроскопического промежутка между нейронами, который молекулы глутамата и другие нейротрансмиттеры и нейромодуляторы пересекают триллионы раз за день. Согласно Экклсу, мозг напоминает мощный орган, а триллионы синапсов представляют собой его клавиатуру. Нематериальное сознание – бессмертная душа в представлении пубоко верующего католика Экклса – играет с синапсами, каким-то образом подталкивая молекулы глутамата на квантовом уровне. “Забудьте все теоретические дискуссии о нейронных сетях и тому подобном, это полный вздор, – сказал он. – Сознание – в глутамате!” Когда настала моя очередь говорить, я сказал, что хочу убедиться, что правильно понял его точку зрения. Если сознание в глутамате, а я вылью чашку глутамата в раковину, будет ли это убийством? “Что ж, – ответил он несколько смущенно, – пожалуй, наверняка сказать будет сложно”<sup>[8]</sup>.

Вам может показаться, что между католиком-дуалистом сэром Джоном Экклсом и атеистом-материалистом Фрэнсисом Криком было мало общего, если не считать их Нобелевских премий. Однако по крайней мере некоторое время их представления о сознании характеризовались одинаковым сомнительным схематизмом. Далекие от науки люди часто не понимают, насколько полезен в науке схематизм, ведь он позволяет разобраться с ужасно сложными вещами, используя работающую модель, которая *почти* верна, и до поры до времени отбрасывая беспорядочные тонкости. Пожалуй, лучше всего в истории науки схематизм применили Крик и Джеймс Уотсон, которые расшифровали структуру ДНК, пока Лайнус Полинг и другие ученые пытались понять все тонкости генома. Крик выступал за смелые ходы, на случай если удастся решить проблему одним махом, но такой подход, конечно же, работает

не всегда. Однажды мне выпал шанс продемонстрировать это на одном из знаменитых чаепитий Крика в Ла-Холье. Эти чаепития представляли собой неформальные встречи коллег по лаборатории, на которых каждый мог задавать вопросы и участвовать в обсуждении. В тот раз Крик сделал важное заявление: недавно было доказано, что нейроны коркового поля V4 “были равнодушны” к цвету (то есть по-разному на него реагировали). Затем он выдвинул поразительно простую гипотезу: сознательный опыт красного, к примеру, *представлял собой* активность в соответствующих чувствительных к красному нейронах этой области сетчатки. Поразмыслив немного, я спросил: “Получается, если извлечь несколько чувствительных к красному нейронов, поместить их в чашку Петри и возбудить при помощи микроэлектрода, в этой чашке Петри возникнет *осознание красного*?” Когда оппонент доводит аргумент до абсурда, можно в ответ схватить быка за рога и подтвердить его вывод, то есть *переиграть* соперника. К такому приему, как известно, однажды прибегнул австралийский философ Дж. К. Смарт, который сказал, что да, согласно его теории морали, иногда *действительно* правомерно подставить и повесить невинного человека! Крик решил переиграть меня. “Да! Это будет единичный случай осознания красного!” Но чьего осознания красного? Этого он не сказал. Позже он пересмотрел свою позицию по этому вопросу, и все же в своем стремлении найти так называемые НКК (нейронные корреляты сознания) они с нейробиологом Кристофом Кохом так и не отказались от своей приверженности этой идее.

Возможно, другой пример лучше покажет, в чем проблема представления о толике сознания в чашке Петри. Физик и математик Роджер Пенроуз и анестезиолог Стюарт Хамерофф совместно разработали теорию сознания, которое определялось не плутамом, а квантовыми эффектами в микротрубочках нейронов. (Микротрубочки – это трубчатые белковые цепочки, которые служат в качестве опорных балок и магистралей в цитоплазме всех клеток, не только нейронов.) Когда Хамерофф объяснил свою позицию на второй международной конференции по науке о сознании “Тусон II”, я спросил из зала: “Стюарт, в вашей практике работы анестезиологом приходилось ли вам принимать участие в одной из сложнейших операций, в ходе которых пациентам пришивают оторванные руки?” Нет, такого опыта у него не было, но он слышал о подобных операциях. “Поправьте меня, если я что-то упускаю, Стюарт, но согласно вашей теории, если бы вы были анестезиологом на такой операции, вы бы чувствовали себя обязанным обезболить оторванную руку, лежащую на льду, верно? Как-никак микротрубочки в нервах руки будут работать, как и микротрубочки остальной нервной системы, а значит, рука будет мучиться от боли, так?” Судя по лицу Стюарта, эта мысль никогда не приходила ему в голову. Мысль о том, что осознание (красного, боли, чего угодно) – в некотором роде сетевое свойство, которое предполагает скоординированные действия бесчисленного множества нейронов, сначала может не понравиться, но доведение до абсурда может помочь людям понять, почему ее надо воспринимать всерьез.

### 3. Правила Рапопорта

Насколько снисходительным надо быть, критикуя взгляды оппонента? При наличии *очевидных* противоречий позиции оппонента на них, конечно же, стоит указать. Если же противоречия *неочевидны*, их необходимо аккуратно поставить на вид – а затем использовать для критики. Но поиск неочевидных противоречий часто сводится к придиркам, передергиваниям и – как мы видели – открытой пародии. Желание обличить оппонента и уверенность, что он *точно* в чем-то да ошибается, приводят к появлению немилосердных интерпретаций, которые становятся легкой мишенью. Но такие легкие мишени, как правило, не имеют существенной ценности и лишь впустую тратят время и терпение всех участвующих в дискуссии, даже если и приносят удовольствие вашим сторонникам. На мой взгляд, лучшим противоядием против этой тенденции высмеивать оппонента служит список правил, которые много лет назад сформулировал специалист по социальной психологии и теории игр Анатолий Рапопорт (создатель прославленной стратегии “око за око” в легендарном турнире по решению повторяющейся дилеммы заключенного, организованном Робертом Аксельродом)<sup>[9]</sup>.

Как составить успешный критический комментарий:

1. Попробуйте выразить позицию своего оппонента так ясно, живо и объективно, чтобы он признался: “Спасибо, я бы и сам лучше не сказал”.
2. Перечислите всё, с чем вы согласны (особенно если с этим согласны не все).
3. Перечислите, чему вы научились у оппонента.
4. Только после этого вы вправе сказать хоть слово критики.

Если вы будете следовать этим правилам, объекты вашей критики сразу станут к ней восприимчивы: вы показываете, что понимаете их мнение не хуже них самих и мыслите здраво (вы соглашаетесь с ними по ряду важных вопросов и даже признаете убедительность некоторых их доводов)<sup>[10]</sup>.

Следовать правилам Рапопорта всегда довольно сложно, по крайней мере мне. Честно говоря, некоторые объекты критики просто не заслуживают такого внимания и уважения. Должен признать, ругать их и выводить на чистую воду – одно удовольствие. Но когда применение правил напрашивается и все получается, это приносит прекрасные результаты. Я был в высшей степени добросовестен, пытаясь отдать должное представлениям Роберта Кейна (1996) об инкомпатибилизме (представлении о свободе воли, с которым я категорически не согласен) в своей книге “Эволюция свободы” (2003), и ценю его ответ на мое письмо с черновиком соответствующей главы:

...На самом деле мне она очень понравилась, невзирая на наши различия. Вы описываете мою позицию подробно и, в целом, справедливо, гораздо лучше, чем обычно делают критики. Вы подчеркиваете комплексный характер моих взглядов и серьезность моего намерения рассматривать сложные вопросы, вместо того чтобы отмахиваться от них. Я благодарен вам за такое отношение и столь подробный анализ моей позиции.

Другие объекты моей критики, с которыми я вел себя по правилам Рапопорта, были менее радушны. Порой чем справедливее критика, тем сложнее ее принять. Стоит напоминать себе, что героическая, но безуспешная попытка защитить позицию автора иногда вредит сильнее, чем агрессивный разнос в пух и прах. Рекомендую.

## 4. Закон Старджона

На Всемирном конвенте научной фантастики, состоявшемся в сентябре 1953 г. в Филадельфии, писатель-фантаст Тед Старджон сказал:

Говоря о детективах, люди упоминают романы “Мальтийский сокол” и “Глубокий сон”. Говоря о вестернах – “Путь на Запад” и “Шейн”. Но говоря о научной фантастике, они называют ее “белибердой о Баке Роджерсе” и добавляют, что “девяносто процентов научной фантастики – полная чушь”. И они правы. Девяносто процентов научной фантастики – полная чушь. Но полной чушью можно назвать девяносто процентов чего угодно, зато остальные десять процентов хороши. И десять процентов научной фантастики, которые чушью не назовешь, не хуже, если не лучше, всего написанного в других жанрах.

Закон Старджона обычно формулируется менее изящно: “Девяносто процентов чего угодно – полное дерьмо”. Девяносто процентов экспериментов в молекулярной биологии, 90 процентов поэзии, 90 процентов книг по философии, 90 процентов рецензируемых статей по математике – и так далее – полное дерьмо. Правда ли это? Возможно, эта цифра завышена, но давайте признаем, что посредственных работ в любой области гораздо больше, чем стоящих. (Некоторые ворчуны утверждают, что их 99 процентов, но мы не станем им верить.) Мораль сей басни такова: если вам хочется раскритиковать какую-то область, жанр, дисциплину, вид искусства... не тратьте время и силы, освистывая дерьмо! Критикуйте качественные вещи – или не критикуйте вообще. Этот совет очень часто пропускают мимо ушей идеологи, которые намерены разрушить репутацию аналитической философии, эволюционной психологии, социологии, культурной антропологии, макроэкономики, пластической хирургии, театра импровизации, телевизионных ситкомов, философской телеологии, массажной терапии – чего угодно. Давайте сразу договоримся, что в мире полно позорного, плутого, второсортного барахла всех сортов. Чтобы не тратить время и не испытывать ничье терпение, сосредотачивайтесь на лучшем доступном материале – на работах лидеров отрасли, на блестящих образцах, а не на шлаке. Обратите внимание, что этот совет тесно связан с правилами Рапорпорта: если только вы не комик, стремящийся рассмешить людей искрометной буффонадой, избавьте нас от ваших карикатур. Я нахожу, что это особенно верно по отношению к философам. Лучшие теории и аналитические выкладки *любого* философа, от величайших, невероятно восприимчивых мудрецов Древней Греции до интеллектуальных героев недавнего прошлого (Бертрана Рассела, Людвиг Витгенштейна, Джона Дьюи, Жана-Поля Сартра – четырех совершенно не похожих друг на друга мыслителей из множества других), можно превратить в полный бред – или утомительное буквоедство, – сделав всего несколько ловких выпадов. Ужас! Не делайте этого. Таким образом вы дискредитируете только самого себя.



## 5. Бритва Оккама

Изобретение этого инструмента мышления приписывается логике и философу четырнадцатого века Уильяму Оккаму (или Оккамскому), но на самом деле он значительно старше. По-латыни он называется *lex parsimoniae*, или принцип бережливости. Обычно его формулируют следующим образом: “Не следует множить сущее без необходимости”. Идея проста: не стоит создавать запутанную, экстравагантную теорию, если есть теория проще (с меньшим количеством ингредиентов, с меньшим количеством сущностей), которая описывает феномен не хуже. Если воздействие чрезвычайно холодного воздуха объясняет все симптомы обморожения, не стоит заявлять о наличии ненаблюдаемых “снежных бактерий” или “арктических микробов”. Законы Кеплера описывают орбитальное движение планет, поэтому нам нет нужды предполагать, что планетами управляют пилоты, сидящие за скрытыми внутри приборными панелями. Эти примеры не встречают возражений, однако распространение действия этого принципа не всегда находило понимание.

Британский психолог девятнадцатого века Конвей Ллойд Морган использовал эту идею, осуждая тенденцию приписывать разумное поведение животным. Правило экономии Ллойда Моргана советует нам не говорить о наличии сложного сознания у насекомых, рыб и даже дельфинов, кошек и собак, если их поведение можно объяснить проще:

Ни в коем случае ни одно действие животного нельзя считать проявлением высшей психологической функции, если его можно объяснить процессами, характерными для более низкой ступени психологической эволюции. [1894, p. 128]

Злоупотребление этим правилом может сподвигнуть нас считать, что у животных и даже у человека есть мозги, но нет сознания. Как мы увидим, абсолютные запреты не слишком хорошо решают вопросы, которые возникают, когда речь заходит о сознании.

Одной из наименее впечатляющих попыток применить бритву Оккама к сложной задаче можно назвать утверждение (и связанные с ним контрутверждения), что считать Бога создателем вселенной проще и экономнее, чем предлагать альтернативы. Как может такой сверхъестественный и непостижимый аргумент быть экономным? Он кажется мне в высшей степени экстравагантным, но, возможно, существуют и умные способы опровергнуть это предположение. Я не хочу вдаваться в детали: в конце концов, бритва Оккама – это просто общее правило, которое часто оказывается очень полезным. Мысль о том, чтобы превратить его в метафизический принцип фундаментального требования рациональности, которому под силу одним махом доказать или опровергнуть существование Бога, попросту смехотворна. Это сродни попытке опровергнуть теорему квантовой механики, доказав, что она противоречит аксиоме “не ставь все на одну карту”.

Некоторые философы доводили бритву Оккама до крайности, используя ее, чтобы отрицать существование времени, материи, чисел, дыр, долларов, программного обеспечения и многого другого. Одним из первых в высшей степени прижимистых мыслителей был древнегреческий философ Парменид, в представлении которого список существующих вещей был ограничен до минимума. Однажды мой студент написал в своей экзаменационной работе: “Парменид сказал: «Есть только одна вещь – и это не я!»” Как ни печально, кажется, именно этому нас и учил Парменид. Без сомнения, здесь не обошлось без трудностей перевода. Мы, философы, привыкаем воспринимать такие идеи всерьез, пускай и потому лишь, что мы не в силах определить, когда “сумасшедшая” идея окажется несправедливо и неосмотрительно раскритикована и падет жертвой недостатка воображения.

## 6. Метла Оккама

Недавно специалист по молекулярной биологии Сидней Бреннер любопытным образом обыграл бритву Оккама и предложил новый термин – “метла Оккама” – для описания процесса, в ходе которого непорядочные в интеллектуальном отношении сторонники той или иной теории отмечают в сторону все неудобные факты. Это наш первый упорный башмак – первый *антиинструмент* мышления, которого следует остерегаться. Особенно опасна эта практика в руках пропагандистов, которые адресуют свои рассуждения непосвященной публике, поскольку, подобно знаменитой улике Шерлока Холмса о собаке, *не залазавшей* ночью, *отсутствие* факта, сметенного в сторону метлой Оккама, под силу заметить только экспертам. К примеру, креационисты неизменно оставляют за скобками многочисленные неловкие доказательства, с которыми не могут справиться их “теории”, а небиологу их аккуратно сформулированные выкладки кажутся вполне убедительными просто потому, что непосвященный читатель не видит того, чего *там нет*.

Как вообще остерегаться невидимого? Обратитесь за советом к экспертам. В книге “Подпись в клетке” (2009) Стивен Мейер объясняет систематическую невозможность естественного (не сверхъестественного) происхождения жизни и дает, казалось бы, относительно беспристрастное и полное – даже для относительно информированного читателя – описание теорий и моделей, предлагаемых по всему миру, показывая их полную безнадежность. Мейер настолько убедителен, что в ноябре 2009 г. выдающийся философ Томас Нагель назвал его книгу лучшей книгой года в лондонском журнале *Times Literary Supplement*, одном из наиболее уважаемых сборников книжных рецензий! В нашей оживленной переписке после публикации его хвалебной статьи он продемонстрировал, что достаточно много знает об истории работ о происхождении жизни и что эти знания позволяют ему доверяться собственным суждениям. В своем письме в *Times Literary Supplement* (от 1 января 2010 г.) он заметил: “На мой взгляд, книга Мейера написана добросовестно”. Если бы Нагель проконсультировался с работающими в этой сфере учеными, он бы заметил, как часто Мейер пользуется метлой Оккама, отбрасывая неудобные факты, а также, вероятно, насторожился бы, узнав, что экспертов, в отличие от него, не попросили прочитать и отрецензировать книгу Мейера до публикации. Возможно, узнай Нагель, что при публикации понравившейся ему книги издатели пошли на хитрость, его уверенность в собственной оценке пошатнулась бы, а возможно, и нет. Порой научное сообщество несправедливо осуждает критиков-отщепенцев, и, возможно – только возможно! – у Мейера не было другого выхода, кроме как напасть из засады. Но Нагелю все же стоило внимательно изучить этот вопрос, прежде чем высказывать свое мнение. Нужно сказать, что ученые, изучающие происхождение жизни, пока не разработали одной надежной, всех устраивающей теории, но вариантов хватает – от них разбегаются глаза.

Мастерски метлой Оккама орудуют конспирологи. В качестве полезного упражнения можно искать в интернете новые теории заговора и смотреть, сумеете ли вы (не будучи экспертом по теме) найти изъяны в рассуждениях, прежде чем читать опровержения от экспертов. Предлагая термин, Бреннер говорил не о креационизме и не о теориях заговора – он лишь указывал, что порой в пылу спора даже серьезные ученые не могут устоять перед искушением “обойти вниманием” некоторые данные, подрывающие основы их любимой теории. Но этому соблазну необходимо противиться во что бы то ни стало.

## 7. Использование обывателей в качестве ложной аудитории

Уберечь людей от *непредумышленного* использования метлы Оккама может хорошая техника, которую я рекомендую годами и несколько раз проверял на практике, но никогда не использовал с тем размахом, с каким мне хотелось бы. В отличие от инструментов, описанных выше, чтобы использовать ее должным образом, необходимы время и деньги. Надеюсь, другие будут активно использовать эту технику и сообщать о результатах. Я решил описать ее на этих страницах, поскольку она помогает разрешить некоторые из тех же проблем коммуникации, с которыми сталкиваются другие общие инструменты мышления.

Во многих сферах – не только в философии – существуют противоречия, которые кажутся непреходящими и в некотором роде искусственными: люди не слышат друг друга и не прикладывают достаточных усилий, чтобы коммуницировать эффективнее. Накаляются страсти, в ход идут насмешки, уважение пропадает. Сторонние наблюдатели разбиваются на лагеря, хотя и не полностью понимают проблему.

Порой кончается все плохо, и причины этого *бывают* ясны. Когда специалисты говорят со специалистами, будь они представителями одной дисциплины или разных, они *всегда* объясняют свою позицию *недостаточно хорошо*. Причина этого понятна: слишком подробные объяснения для другого специалиста могут стать серьезным оскорблением – “Мне что, объяснить на пальцах?”, – а оскорблять коллегу-эксперта никому не хочется. В результате, чтобы не рисковать, люди склоняются к неполным объяснениям. В основном это происходит непредумышленно, а воздержаться от такого практически невозможно – и это даже к лучшему, поскольку никому не повредит непринужденная вежливость. Но подобное великодушное предположение о том, что уважаемая аудитория на самом деле умнее, чем она есть, имеет печальный побочный эффект: специалисты часто не слышат друг друга.

*Простого* выхода не существует: сколько ни проси всех специалистов, присутствующих на семинаре или конференции, объяснять свои позиции предельно подробно и сколько ни выслушивай их обещаний, ничего из этого не выйдет. Станет только хуже, потому что в таком случае люди будут особенно осторожны, боясь ненароком кого-нибудь оскорбить. И все же есть довольно эффективный способ с этим справиться: нужно, чтобы все специалисты излагали свои взгляды небольшой группе любопытных неспециалистов (здесь, в Университете Тафтса, мне везет общаться со способными студентами), в то время как другие специалисты слушали бы их со стороны. Подслушивать при этом не надо – в этом предложении нет ничего *коварного*. Напротив, все могут и должны понимать, что суть этого упражнения заключается в том, чтобы участникам было комфортно говорить в тех терминах, которые понятны *всем*. Обращаясь к студентам (ложной аудитории), выступающие вообще не будут опасаться оскорбить специалистов, ведь они обращаются не к специалистам. (Полагаю, они, возможно, будут опасаться оскорбить студентов, но это уже другой вопрос.) Если все идет хорошо, специалист А объясняет дискуссионные вопросы студентам, в то время как специалист Б его слушает. В какой-то момент специалист Б может просить. “Так *вот* о чем вы говорили! Теперь я понимаю”. Возможно, положительный эффект наступит лишь тогда, когда настанет очередь специалиста Б объяснять проблему тем же студентам, в процессе чего озарение снизойдет на специалиста А. Порой все складывается не идеально, но обычно такое упражнение проходит хорошо и идет всем на пользу. Специалисты устраняют существующее между ними недопонимание, а студенты получают первоклассный урок.

Я несколько раз проделывал подобное в Университете Тафтса, благодаря великодушной поддержке администрации. Отбирая небольшое количество студентов (меньше дюжины), я объясняю им их роль: от них требуется не принимать на веру ничего непонятного. Они должны поднимать руки, перебивать профессора и указывать специалистам на все аспекты, которые кажутся им неочевидными или туманными. (При этом они получают список литературы для подготовки к семинару, чтобы немного ориентироваться в обсуждаемой теме, но оставаться заинтересованными непрофессионалами.) Им нравится эта роль – и удивляться здесь нечему, ведь они принимают участие в уникальных семинарах больших шишек. Специалисты тем временем часто отмечают, что, получая задачу (заранее) объяснить свою позицию в таких условиях, они находят наилучший способ донести собственные мысли. Иногда этих специалистов годами “прикрывали” коллеги-эксперты, молодые исследователи и способные студенты-магистры, поэтому им действительно нужно попытать свои силы.

## 8. Выпрыгивание

Сложно найти применение метле Оккама, поскольку она помогает отбросить в сторону неудобные факты, но еще сложнее достичь того, что Даг Хофштадтер (1979, 1985) называет *выпрыгиванием* (*jootsing*) – “выпрыгиванием из системы” (*jumping out of the system*). Эта тактика применима не только в науке и философии, но и в искусстве. Креативность – этот востребованный, но редкий дар – часто представляет собой до той поры невообразимое нарушение правил некоторой системы. Это может быть система классической гармонии в музыке, законы размера и ритма в сонетах (или даже в лимериках) или “каноны” вкуса и формы в других художественных жанрах. Это могут быть также аксиомы и принципы какой-либо теории или исследовательской программы. Креативность проявляется не просто в обнаружении чего-то нового – это под силу каждому, ведь новизну можно найти в любой случайной близости вещей, – но в умении заставить эту новизну *выпрыгнуть* из системы, которая долгое время считалась устоявшейся, и не без причины. Когда художественная традиция достигает точки, в которой “все дозволено”, стремящиеся к креативности сталкиваются с проблемой: нет ни фиксированных правил, против которых стоит бунтовать, ни очевидных ожиданий, которые не хочется оправдывать, нечего свергать и нечего использовать в качестве фона для создания чего-либо неожиданного и значимого. Чтобы сломать традицию, ее лучше всего хорошо *изучить*. Именно поэтому так мало дилетантов и любителей предлагают что-то поистине креативное.

Сядьте за фортепиано и попробуйте сочинить хорошую новую мелодию. Вскоре вы поймете, насколько это сложно. Перед вами все клавиши, и нажимать их можно каким угодно образом, но пока вы не найдете, на что опереться, пока не выберете какой-нибудь стиль, или жанр, или нотный рисунок и не используете его либо не сошлетесь на него, пока не поиграете с ним, у вас не выйдет ничего, кроме шума. И не всякое нарушение правил поможет вам достичь желаемого. Насколько мне известно, по меньшей мере два джазовых арфиста сумели добиться успеха – или хотя бы удержались на плаву, – но вряд ли стоит надеяться сделать себе имя, исполняя Бетховена на бонго. В этом искусство похоже на науку: в любой теоретической дискуссии всегда остается множество неисследованных предположений, однако, пытаясь опровергнуть их по очереди в поисках самого уязвимого, вы не добьетесь успеха в науке или философии. (Это все равно что взять мелодию Гершвина и изменять в ней по одной ноте, надеясь получить что-то стоящее. Удачи! Мутации почти всегда пагубны.) На самом деле все сложнее, но порой людям везет.

Советовать человеку идти вперед, выпрыгивая из системы, сродни тому, чтобы советовать инвестору покупать дешево и продавать дорого. Да, конечно, в этом и смысл, но как этого добиться? Обратите внимание, что совет по инвестициям нельзя назвать *совершенно* бесполезным и неприменимым, а призыв к выпрыгиванию даже более содержателен, поскольку он объясняет, как выпадит ваша цель, если вы ее хоть раз видели. (Все знают, как выглядят *больше денег*.) Когда перед вами стоит научная или философская задача, выпрыгивать, как правило, приходится из системы, которая настолько хорошо внедрилась в повседневность, что стала невидимой, как воздух. Как правило, когда длительная полемика ни к чему не приводит, а обе “стороны” упрямо настаивают на своей правоте, проблема часто заключается в том, что они заблуждаются в чем-то, что не вызывает у них разногласий. Обе стороны считают это слишком очевидным, чтобы вообще уделять этому внимание. Искать эти невидимые хитрости нелегко, поскольку то, что кажется очевидным враждующим специалистам, обычно кажется очевидным и всем остальным. Совет не упускать из виду подразумеваемые допущения вряд ли здесь действительно поможет, но вы с большей вероятностью найдете проблему, если надеетесь ее найти и имеете хотя бы примерное представление о том, как она должна выглядеть.

Иногда бывают зацепки. Несколько прекрасных примеров выпрыгивания предполагают отказ от общепризнанной *вещи*, которой, как оказалось, и не было вовсе. *Флогистон* считался элементом огня, невидимый, самонепроницаемый флюид или газ *теплород*, как предполагалось, был основным ингредиентом тепла, пока их существование не опровергли, как произошло и с *эфиром* в качестве среды, где свет в воде и воздухе идет по пути, по которому уже прошел звук. Но другие прекрасные выпрыгивания стали дополнениями, а не изъятиями: микробы, электроны и – возможно даже – многомировая интерпретация квантовой механики! Сначала никогда не понять, стоит ли выпрыгивать из системы. Мы с Рэем Джекендоффом выдвинули предположение, что нужно отказаться от почти всегда признаваемой аксиомы о том, что сознание представляет собой “высший” или “центральный” из всех психических феноменов, а я выдвинул предположение, что пора забыть о типичной и необоснованной

привычке думать о сознании как об особой *среде* (наподобие эфира), в которой осуществляется преобразование или перевод содержимого. Вместе со многими другими мыслителями я также заметил, что если вы считаете совершенно очевидной несовместимость свободы воли и детерминизма, то делаете большую ошибку. Подробнее об этом позже.

Еще одна зацепка: иногда проблема возникает, если кто-то однажды сказал “давайте чисто теоретически предположим...” и все согласились – чисто теоретически, – а затем в ходе последующей пикировки забыли, с чего все началось! Думаю, порой – по крайней мере в моей сфере философии – оппонентам так нравится спорить, что ни одной из сторон не хочется прекращать спор, внимательно изучив его истоки. Вот два старинных примера, которые, само собой, вызывают полемику: (1) “Почему существует нечто, а не ничто?” – глубокий вопрос, требующий ответа. (2) “Выбирает ли Бог добро, потому что оно благое, либо же добро – благое, потому что выбрано Богом?” – другой важный вопрос. Полагаю, было бы чудесно, если бы кто-то сумел предложить *хороший* ответ на любой из этих вопросов, а потому признаю, что называть их псевдопроблемами, недостойными внимания, не слишком *справедливо*, хотя это и не доказывает мою неправоту. Никто не говорит, что истина должна быть любопытной.

## 9. Три вида гулдинга: скорейность, нагромождения и тустеп Гулда

Покойный биолог Стивен Джей Гулд виртуозно создавал и использовал упорные башмаки. Вот три прекрасных вида рода *гулдинг*, который я назвал так в честь человека, использовавшего их лучше всего.

*Скорейность* – это способ быстро и безболезненно отойти от *ложной дихотомии*. Классическая формулировка скорейности такова: “Дело не в том, что *тотойто*, как утверждают консерваторы, а *скорее* в том, что *этоэтойто*, и это все меняет”. В некоторых случаях скорейность работает отлично, потому что вам предлагается выбрать между двумя доступными альтернативами – в этом случае дихотомия оказывается не ложной, а истинной и неизбежной. Но в других случаях скорейность не ограничивается ловкостью рук, поскольку слово “скорее” предполагает – бесспорно, – что между двумя утверждениями есть важное несоответствие, которое разводит их в стороны.

Вот прекрасный пример скорейности, который Гулд использовал при описании своей теории прерывистого равновесия:

Изменения обычно происходят не в результате незаметной постепенной эволюции всего вида, а *скорее* [курсив мой] в результате изоляции небольшой популяции и ее геологически мгновенной трансформации в новый вид. [1992b, p. 12]

Этот фрагмент заставляет нас поверить, что эволюционные изменения не могут быть одновременно “геологически мгновенными” и “незаметно постепенными”. Само собой, это не так. На самом деле именно такими и должны быть эволюционные изменения, если только Гулд не утверждает, что эволюция идет скачками (в проектном пространстве), но в других обстоятельствах он заявлял, что не одобряет гипотезу о скачкообразном развитии. “Геологически мгновенное” видообразование может произойти за “короткий” период времени – скажем, за пятьдесят тысяч лет, период времени, едва заметный в большинстве геологических пластов. За этот краткий миг рост типичного представителя вида может увеличиться, скажем, с полуметра до метра, продемонстрировав 100-процентный прирост, но происходить это будет со скоростью миллиметр в столетие, что кажется мне незаметной постепенной эволюцией.

Давайте *придумаем* несколько других примеров скорейности, чтобы хорошенько разобраться с природой этого трюка.

Дело не в том, что люди – просто “влажные роботы” (как говорит Дилберт, с которым соплашается большинство исследователей в когнитивной науке), а *скорее* в том, что люди наделены свободой воли и несут моральную ответственность за свои хорошие и плохие поступки.

И снова – почему нужно выбрать только одно? Не хватает утверждения, что “влажные роботы” могут также быть людьми, наделенными свободой воли и несущими моральную ответственность за свои поступки. Этот пример основан на распространенном – но спорном – утверждении. Вот другой:

Религия не опиум для народа, как сказал Маркс, а *скорее* пубокоое и утешительное проявление того факта, что человечество признает неизбежность смерти.

И снова – почему религия не может быть опиумом и одновременно утешать? Думаю, вы уже ухватили суть, а потому находить скорейности в документе вам даже легче, чем ложные дихотомии, которые никогда не бывают настолько очевидны: просто напишите “скорее” в строке поиска и изучите результаты. Не забывайте: не все “скорее” представляют собой скорейности, некоторые из них вполне обоснованы. А в некоторых скорейностях не используется слово “скорее”. Вот пример, в котором применяется конструкция “\_\_\_\_\_”, а не “\_\_\_\_\_”. Я составил его из элементов, позаимствованных из работ нескольких идеологов когнитивной науки.

Следует считать, что нервные системы активно генерируют пробы своей среды, а не просто работают, как компьютеры, пассивно обрабатывая поступающие от органов чувств данные.

Кто сказал, что компьютеры, обрабатывающие поступающие данные, не могут активно генерировать пробы? Этот знакомый контраст между ужасно “пассивными” компьютерами и чудесно “активными”

организмами до сих пор не получил должного подтверждения, но при этом стал одним из самых типичных известных мне блокировщиков воображения.

Вариацию скорейности, часто используемую Гулдом, можно назвать *нагромождением*:

Мы говорим о “марше от монады к человеку” (и снова старомодный язык) так, словно эволюция шла по бесконечным путям развития вдоль непрерывных линий наследования. Ничто не может быть дальше от реальности. [1989а, р. 14]

Что не может быть дальше от реальности? На первый взгляд может показаться, будто Гулд говорит, что не существует бесконечной, непрерывной линии наследования между “монадами” (одноклеточными организмами) и нами, но это, само собой, не так. Не существует более надежного следствия великой идеи Дарвина. Так о чем же говорит Гулд? Похоже, акцент надо ставить на фразу “пути *развития*” – именно (и только) вера в развитие “далека от реальности”. Пути наследования действительно бесконечны и непрерывны, но наследуется не (мировое) развитие. Это верно: по (непрерывным) бесконечным линиям наследуется (в основном) местное развитие. Когда мы читаем это высказывание Гулда, нам кажется – если только мы не проявляем особенной осторожности, – что он показал нам, будто стандартное утверждение теории эволюции о едином пути развития (по непрерывной линии наследования) от монад к человеку содержит в себе серьезный изъян. Но, говоря словами самого Гулда, “ничто не может быть дальше от реальности”.

Еще один фокус – *тустеп Гулда* – я несколько лет назад описал в одной из опубликованных работ, после чего специалист по эволюционной теории Роберт Триверс (частная переписка, 1993) присвоил ему имя в честь его изобретателя:

На первом этапе вы подменяете тезис и “опровергаете” его (этот трюк всем знаком). На втором (и это гениально) вы сами привлекаете внимание к тому, что сделали первый шаг – что ваши оппоненты на самом деле не разделяют мнений, которые вы им приписали, – но интерпретируете их высказывания как недовольные ответы на ваш выпад! [Dennett 1993, р. 43]

В своем письме редактору *New York Review of Books* (1993), где Гулд двумя месяцами ранее раскритиковал прекрасную книгу Хелены Кронин “Муравей и павлин” (выпуск от 19 ноября 1992 г.), я привел три примера тустепа Гулда. Вот самый универсальный из них:

Наиболее прозрачный случай – предложенный Гулдом “экстраполяционизм”, который описывается как логическое продолжение “адаптационизма Кронин”. Именно доктрина панпреемственности и панградуализма удачно – и вполне тривиально – опровергается фактом массового вымирания. “Но если массовые вымирания представляют собой настоящие разрывы преемственности, если медленный процесс адаптации в обычное время не приводит к предсказуемому успеху, невзирая на массовые вымирания, то экстраполяционизм терпит поражение, а адаптационизм гибнет”. Я не вижу ни одной причины для любого адаптациониста безрассудно поддерживать идею вроде “экстраполяционизма” в настолько “чистом” виде, дабы отрицать возможность или даже вероятность, что массовое вымирание играет важную роль в подрезке древа жизни, как выражается Гулд. Очевидно, что самый совершенный динозавр погибнет, если комета столкнется с его планетой с силой, в сотни раз превышающей мощность всех когда-либо изготовленных водородных бомб. Ни единое слово из книги Кронин не подтверждает его уверенность, что она допустила эту ошибку. Если Гулд и полагает, что роль массовых вымираний в эволюции имеет отношение к какой-либо из проблем, обсуждаемых Кронин, в частности к половому отбору и альтруизму, то он не объясняет, каким образом и почему. Когда в последней главе Кронин обращается к центральному вопросу теории эволюции, на котором она *не делала* акцента, а именно к происхождению видов, и замечает, что ответа на него по-прежнему нет, Гулд называет это запоздалым откровением, ироничным признанием поражения ее “панадаптационизма”. Возмутительно! [р. 44]

Исследователям риторики могу предложить прекрасное упражнение: пересмотреть огромное количество публикаций Гулда и каталогизировать все типы упорных башмаков, которыми он пользовался, начиная со скорейности, нагромождений и тустепа Гулда.

## 10. Оператор “безусловно”: ментальный блок

Читая или просматривая доказательные эссе, особенно написанные философами, вы можете воспользоваться одним трюком, который сэкономит вам время и силы, тем более в наш век простого компьютерного поиска: поищите в документе слово “безусловно” и проверьте каждый полученный результат. Не всегда (и даже не в большинстве случаев), но часто слово “безусловно” становится маячком, указывающим на слабые места в рассуждениях, предупреждением о вероятном использовании **упорного башмака**. Почему? Потому что этим словом автор отмечает то, в чем он совершенно уверен и в чем, как он надеется, будут уверены его читатели. (Если бы автор был *абсолютно* уверен, что все читатели с ним согласятся, он вообще не стал бы об этом говорить.) Стоя на перепутье, автор должен был решить, стоит ли вообще упоминать об этом спорном моменте или приводить доказательства, и – поскольку жизнь коротка – решил сделать голословное утверждение, вероятно, имея все основания полагать, что оно не вызовет разногласий. Именно в таких местах и стоит искать не слишком хорошо изученные “истины”, которые на самом деле неверны!

Впервые я заметил эту полезную функцию “безусловно”, комментируя эссе Неда Блока (1994), в котором содержалось несколько прекрасных примеров, направленных против моей теории сознания. Вот один из них, который Блок намеренно выделил курсивом<sup>[11]</sup>, чтобы типографскими средствами подчеркнуть его очевидность:

*Безусловно, утверждение, что некоторые умственные представления достаточно устойчивы, чтобы воздействовать на память, контролировать поведение и т. д., есть не что иное, как биологический факт о людях, а не культурный конструкт. [р. 27]*

Предполагалось, что это сразу – без лишних споров – развенчает мою теорию человеческого сознания, в соответствии с которой сознанию, по сути, надо учиться, поскольку оно представляет собой набор когнитивных микропривычек, не гарантированных человеку при рождении. “Всякий раз, когда Блок говорит «безусловно», – заметил я, – ищите так называемый ментальный блок” (Dennett 1994a, p. 549). Блок едва ли не чаще остальных философов использует “безусловно”, но к нему то и дело прибегают и другие мыслители, и каждый раз при этом должен звучать тревожный сигнал. “В этот момент происходит непреднамеренное передергивание, когда ложную посылку всеми правдами и неправдами протаскивают мимо цензоров” (Dennett 2007b, p. 252).

Недавно я решил проверить свои догадки о “безусловно” несколько более систематически. Я просмотрел десятки статей по философии сознания на сайте philpapers.org – всего около шестидесяти – и проверил их на наличие слова “безусловно”. В большинстве статей это слово вообще не встречалось. В тех же, где оно встречалось (от одного до пяти раз в моей выборке), в основном оно использовалось невинно. Несколько случаев я счел спорными, а еще шесть раз отчетливо услышал тревожный сигнал (мне он показался звонким). Само собой, представления других людей об очевидном могут отличаться, поэтому я и не стал сводить в таблицу “данные” этого неформального эксперимента. Я призываю скептиков провести собственные исследования и оценить результаты. Один вопиющий пример использования оператора “безусловно” я подробно разберу далее, в главе 64.



## 11. Риторические вопросы

Остерегаясь “безусловно”, стоит также наметать свой глаз на риторические вопросы, возникающие в ходе спора или полемики. Зачем? Дело в том, что, подобно использованию “безусловно”, они говорят о желании автора срезать углы. Риторический вопрос оканчивается вопросительным знаком, но ответа на него не предполагается. Иными словами, автор и *не думает* ждать, пока вы ответите на вопрос, поскольку ответ настолько очевиден, что вам будет стыдно его озвучить! Получается, что большинство риторических вопросов представляет собой сжатые варианты доведения до абсурда, которые слишком очевидны, чтобы проговаривать их целиком. Если вам встретится хороший риторический вопрос, удивите своего собеседника, ответив на него. Помню, много лет назад эта тактика была прекрасно проиллюстрирована в комиксе *Peanuts*. Чарли Браун задал риторический вопрос: “Кому решать, что верно и неверно?” – и на следующем рисунке Люси ответила: “Мне”.

## 12. Что такое глубокость?

Мой покойный друг, специалист в области информатики Джозеф Вейценбаум, хотел стать философом и на закате своей карьеры попытался отойти от частностей к глубокомыслию. Однажды он рассказал мне, как за ужином, когда он после долгих раздумий ударился в разлагольствования, его дочь Мириам воскликнула: “Ого! Папа только что сказал глубокость!” Какой прекрасный термин!<sup>[12]</sup> Я решил позаимствовать его и найти ему более аналитическое применение.

Глубокостью называется утверждение, которое кажется важным и истинным – и мудрым, – но добывается этого за счет своей двусмысленности. На первый взгляд, она очевидно неверна, но потрясла бы мир, оказалась она верной; на второй – верна, но банальна. Неосторожный слушатель выхватывает зерно истины из второго прочтения и сокрушительную важность из первого и думает: “Ого! Да это глубокость”.

Вот вам пример. (Лучше присядьте, это не для слабонервных.)

Любовь – всего лишь слово.

Ого! Невероятно. Фантастика, правда? Нет. На первый взгляд, это очевидная пупость. Я не знаю, что такое любовь – возможно, чувство или эмоциональная привязанность, возможно, межличностная связь, а возможно, высочайший уровень человеческого сознания, – но всем нам прекрасно известно, что она не просто слово. Любовь в словаре не найдешь!

Мы можем снова прочесть это утверждение, используя правило, которое очень важно для философов: *говоря* о слове, мы ставим его в кавычки, а потому

“Любовь” – всего лишь слово.

Это правда. “Любовь” – слово русского языка, но всего лишь слово, а не предложение, к примеру. Оно начинается на букву “Л”, состоит из шести букв и находится в словаре между “любезным” и “любоznательным”, которые тоже всего лишь слова. “Чизбургер” – всего лишь слово. “Слово” – всего лишь слово.

Но это неправильно, скажете вы. Безусловно, человек, сказавший, что любовь – всего лишь слово, имел в виду совсем другое. Несомненно. Но он этого не сказал. Может, он имел в виду, что “любовь” – это слово, которое вводит людей в заблуждение, заставляя их поверить, что им обозначается нечто чудесное, чего на самом деле вообще не существует, на манер слова “единорог”, а может, он хотел сказать, что это слово настолько туманно, что никто не может знать, означает ли оно конкретную вещь, связь или событие. Но ни одно из этих утверждений нельзя считать достаточно правдоподобным. Возможно, дать определение слову “любовь” непросто, а в состоянии любви сложно быть хоть в чем-то уверенным, но эти утверждения очевидны, не слишком информативны и не мудры.

Не все глубокости так легко анализировать. Недавно Ричард Докинз рассказал мне о прекрасной глубокости архиепископа Кентерберийского Роуэна Уильямса, который описал свою веру как

молчаливое ожидание истины, смиренное пребывание в тени вопросительного знака.

Предлагаю вам самим проанализировать этот пример.

## Резюме

Умело используемый инструмент становится частью вас, подобно рукам и ногам, и это особенно верно в отношении инструментов мышления. Вооружившись этими простыми универсальными инструментами мышления, вы будете лучше подготовлены к последующим изысканиям: вы будете видеть возможности, слышать предупредительные сигналы, чувствовать подвохи и остерегаться ошибок, которые могли бы сделать в ином случае. Вам стоит также держать в уме некоторые максимы – к примеру, правила Рапопорта и закон Старджона, – которые будут нашептывать вам советы, подобно говорящему сверчку, напоминая о необходимости контролировать свою агрессию, когда вы вторгнетесь в дебри, размахивая оружием. Да, инструменты мышления можно считать оружием, поэтому сравнение с битвой вполне уместно. Дух соперничества, очевидно, представляет собой естественный побочный продукт интеллектуальных амбиций и дерзости, необходимой для работы над самыми сложными задачами. Мы увидели, что в пылу битвы даже великие мыслители порой обращаются к грязным трюкам в своем стремлении убедить вас принять их точку зрения, а конструктивная критика превращается в высмеивание, как только появляется возможность отпустить пару острот.

Мы будем рассматривать насущные вопросы смысла, эволюции, сознания и особенно свободы воли. При столкновении с некоторыми их аспектами вы почувствуете ужас или отвращение – и будьте уверены, вы в этом не одиноки, ибо даже именитые эксперты подвержены самообману и могут не видеть истины из-за убеждений, которые подкрепляются скорее эмоциональной привязанностью, чем логикой. Людям не все равно, наделены ли они свободой воли, каким образом сознание заключается в их теле и откуда смысл в мире, состоящем из одних атомов, молекул, фотонов и бозонов Хиггса, да и есть ли этот смысл вообще. Людям и должно быть не все равно. В конце концов, что важнее вопросов, кто мы в этом мире и что нам с этим делать? Так что будьте осторожны. Впереди зыбкая почва, а карты ненадежны.

### III.

## Инструменты мышления о значении или содержании

Зачем начинать со значения (*meaning*)? Я начинаю со значения, поскольку этот вопрос лежит в основе всех сложных проблем, и причина проста: этих проблем не возникает, пока мы не начинаем говорить о них друг с другом и с самими собой. Барсуков не заботит свобода воли – проблема сознания не волнует даже дельфинов, поскольку *задавать вопросы* вообще не в их репертуаре. Все мы знаем, что любопытной Варваре на базаре нос оторвали, но именно любопытство толкает нас, мыслящих людей, забираться в дебри трудностей. Возможно, это негативный аспект языка и мы были бы счастливее – как спокойные, здоровые млекопитающие, – если бы не обращали внимания на эти вопросы, как наши сородичи-обезьяны. Но у нас есть язык, поэтому нам никуда не деться от Великих Вопросов, и простыми они не кажутся – хорошо это или плохо.

Первым делом в любом успешном исследовании нужно как можно лучше понять, от чего мы отталкиваемся и какое оборудование используем. Слова имеют значения. Как это возможно? Мы пользуемся словами и обозначаем вещи, говоря о них. Как это возможно? Как мы вообще понимаем друг друга? Кажется, собаки тоже “понимают” (вроде бы) несколько слов, даже несколько сотен слов, но без учета этих одомашненных трюкачей и других видов, использующих рудиментарные сигнальные системы (приматов, птиц... даже каракатиц!), словами пользуется только человек – и именно слова отличают наше сознание (*mind*) от сознания других животных. Это разительное отличие, но все равно кажется, что другие животные – “высшие” животные – наделены сознанием, а следовательно, в какой-то степени, пускай и небольшой, взаимодействуют со значениями: со значениями их перцептивных состояний, их побуждений и желаний и даже их снов.

Порой нам кажется, что животные совсем такие, как мы, – будто они люди, одетые в костюмы кошек, медведей и дельфинов. Это характерно для любой человеческой культуры: животные представляются нам знающими, понимающими, желающими, дерзающими, опасаящимися, решающими, жаждущими, помнящими и так далее. Иными словами, они кажутся такими же, как мы: наделенными *сознанием*, полным *многозначных...* вещей (идей? убеждений? ментальных представлений?). Как могут значения содержаться в мозге? Нам нравится думать, что, раз слова имеют значение, вероятно, и осмысленные вещи у нас в голове – и в голове у животных – подобны словам, составляются в мысленные предложения, выражают наши убеждения и так далее. Но если слова берут свои значения от сознаний, которые их высказывают, то откуда берут свои значения эти слова сознания (*mindwords*)? Хранятся ли в сознании животных слова сознания и их определения, собранные в своеобразный церебральный словарь? А если животные – по крайней мере “высшие” животные – всегда были наделены сознанием, полным слов сознания (*mindwords*), то почему они не умеют говорить?<sup>[13]</sup> Идея языка мышления глубоко проблематична, но наши мысли и убеждения должны из чего-то состоять. Из чего же еще?<sup>[14]</sup>

### 13. Убийство на Трафальгарской площади

Вот наш первый насос интуиции. Жак выстрелом убивает своего дядю на Трафальгарской площади; Шерлок ловит его на месте преступления; Том читает об этом в газете “Гардиан”, а Борис – в газете “Правда”. Опыт Жака, Шерлока, Тома и Бориса существенно различается – не говоря уже об их биографиях и перспективах, – но у них есть кое-что общее: все они считают, что француз совершил убийство на Трафальгарской площади. Они не говорили об этом, даже “самим себе”, эта фраза, как мы можем предположить, не “приходила им в голову”, а даже если бы и приходила, Жак, Шерлок, Том и Борис восприняли бы ее совершенно по-разному. И все же все они считают, что француз совершил убийство на Трафальгарской площади. Фактически это общее свойство можно рассмотреть только с одной, очень ограниченной точки зрения – с точки зрения **народной психологии**. Обычные народные психологи – все мы – без труда находят такие полезные сходства между людьми. Мы делаем это, практически ничего не зная о том, что находится в пространстве между ушами тех людей, которым мы приписываем такие убеждения. Мы можем полагать, что между четырьмя этими людьми есть *еще* что-то общее – сходным образом оформленное то-или-иное, что каким-то образом регистрирует это общее убеждение, – но в таком случае мы ступаем на нетвердую почву сомнительного теоретизирования. Какие-то нервные структуры действительно *могут* быть сходными – если все четыре мозга “сформулировали” утверждение, что француз совершил убийство на Трафальгарской площади, одинаковым образом, – но это необязательно и даже крайне маловероятно по причинам, которые мы кратко опишем ниже.

Предложения “Я голоден” и *J’ai faim* имеют кое-что общее, несмотря на то что они составлены из разных букв (или разных фонем при произнесении вслух), на разных языках с разными грамматическими структурами: они *означают* или *описывают* одно и то же – голод говорящего. Это общее свойство, *значение* (двух предложений на каждом из языков) или *содержание* (представлений, которые они выражают), играет основополагающую роль в философии и когнитивной науке. *Очёмность* (*aboutness*), свойственная, к примеру, фразам, картинам, утверждениям и (без сомнения) некоторым состояниям сознания, на философском жаргоне называется *интенциональностью* (от лат. *intentio* – намерение), и этот термин не слишком удачен, поскольку обыватели часто путают его с будничной идеей делать что-либо намеренно (как в вопросе “Благородны ли твои намерения?”). Вот как можно запомнить различия: сигарета не говорит о курении и вообще ни о чем не говорит, хотя и подразумевает наличие намерения ее выкурить; знак “не курить” говорит о курении, а следовательно, демонстрирует интенциональность; убеждение, что за деревом скрывается грабитель, демонстрирует интенциональность (оно говорит о – возможно, несуществующем – грабителе), но его возникновение явно не может считаться намеренным в традиционном смысле (вы не “думаете о нем намеренно”, оно просто приходит к вам); решив убежать от дерева, вы продемонстрируете намерение в традиционном смысле, но ни о чем таким образом не расскажете. Если вы просто возьмете в привычку заменять неуклюжим термином “очёмность” философское понятие “интенциональности”, большинства ошибок удастся избежать. Не считая взаимного согласия, что значение и содержание представляют собой тесно связанные и взаимозависимые феномены – или даже один феномен (интенциональность), – единого мнения о том, что есть содержание (или значение) и как лучше всего его ухватить, пока не существует. Именно поэтому об этой теме следует говорить с осторожностью. В ней множество проблем, но мы можем подходить к вопросам постепенно, маленькими шагами.

Пример с четырьмя людьми показывает, что мозги этих людей могут иметь мало общего, при этом характеризуясь одной и той же “интенциональностью”, поскольку они верят “в одно и то же”. Жак был непосредственным свидетелем убийства – он сам убийца, – эмпирическая близость Шерлока к событию лишь немногим менее непосредственна, но Том и Борис узнали о нем совершенно иначе. Существует неограниченное количество способов получить информацию о том, что француз совершил убийство на Трафальгарской площади, и неограниченное количество способов использовать эту информацию себе во благо (отвечать на вопросы на викторинах, играть на ставках, дразнить французских туристов в Лондоне и т. д.). Далее мы узнаем, есть ли достаточное основание полагать, что все эти источники и следствия обрабатываются в одинаковых структурах мозга, но пока нам не стоит торопиться с выводами.

Прежде чем закончить с этим насосом интуиции, нам стоит последовать совету Дага Хофштадтера и *покрутить регуляторы*, чтобы понять, как работают его элементы. Почему я выбрал именно такую посылку? Потому что мне нужно было достаточно запоминающееся и из ряда вон выходящее событие, чтобы о нем сообщили на разных языках вдали от места преступления. Важно упомянуть, что большую

часть знаний мы получаем совершенно иным путем. К примеру, Жак, Шерлок, Том и Борис разделяют неограниченное количество других представлений, которые возникли не столь драматично: скажем, все они верят, что стулья больше ботинок, что суп жидкий и что слоны не летают. Если бы я попытался сообщить вам, что лосось в дикой природе не пользуется слуховыми аппаратами, вы бы ответили, что и так это знаете, но откуда? Вы не родились с этим знанием, не получили его в школе – и крайне маловероятно, что вы вообще когда-либо формулировали в голове такое утверждение. В связи с этим, хотя и может показаться очевидным, что Борис узнал о французе, “просто загрузив” соответствующее русское предложение из “Правды” себе в голову и затем переведя его на, хм, рассудочный, нет никаких оснований предполагать, что мозг Бориса провел такую же операцию (перевод с *какого* на рассудочный?) над фактом о лососе.

Вот еще один регулятор: допустим, что собака Фидо и голубь Клайд тоже были непосредственными свидетелями убийства. Они могут *что-то* запомнить об этом событии, в результате чего в их мозгу произойдет изменение, которое окажет влияние на их дальнейшее поведение, но запомнят они не факт, что француз совершил убийство на Трафальгарской площади, хотя информация об этом факте и может передаваться светом и звуком, улавливаемыми их органами чувств. (Видеозапись события, к примеру, может стать надлежащим доказательством свершившегося факта, но Фидо и Клайд ее не поймут.) Таким образом, этот насос интуиции может привнести антропоцентрический уклон в наше исследование значения. В качестве типичных проводников значения выступают слова и предложения, однако животные ими пользоваться не умеют, а идея о том, что их *мозг* все равно их использует, кажется притянутой за уши – что, впрочем, не доказывает ее ложности. Если она окажется истинной, это станет поразительным открытием, но делать такие открытия нам не впервой.

Феномен интенциональности одновременно совершенно привычен – и так же свойственен нашей повседневной жизни, как пища, мебель и одежда, – и крайне трудноуловим в научной перспективе. У нас с вами редко возникают сложности с тем, чтобы отличить поздравление с днем рождения от смертельной угрозы, но представьте себе стоящую перед инженерами задачу разработать надежный детектор смертельных угроз. Что общего у всех смертельных угроз? Казалось бы, только их значение. А значение *не похоже* на радиоактивность или кислотность – одно из свойств, легко определяемых хорошо настроенным детектором. Ближе всего мы подошли к созданию универсального детектора значений, когда разработали компьютер *IBM Watson*, который сортирует значения гораздо лучше более ранних систем искусственного интеллекта. Однако обратите внимание, что все не так просто: даже этот компьютер (вероятно) не заметит ряд смертельных угроз, которые под силу уловить ребенку. Даже маленькие дети понимают, что, когда один ребенок, хохоча, кричит другому: “Ну держись! Еще раз так сделаешь – убью!” – он на самом деле не грозит никому смертью. Внушительные размеры и сложность компьютера *Watson* хотя бы косвенным образом показывают, насколько трудноуловимо знакомое нам свойство значения.

## 14. Старший брат из Кливленда

И все же значение нельзя считать совершенно *непостижимым* свойством. Так или иначе структуры в нашей голове каким-то образом “хранят” наши убеждения. Когда вы узнаете, что пуду – млекопитающие, *нечто* в вашей голове должно измениться; нечто должно закрепиться в том положении, в котором оно не было закреплено, прежде чем вы об этом узнали, и это нечто так или иначе должно обладать достаточной *очётностью*, чтобы объяснять вашу новообетенную способность определять, что пуду состоят в более близком родстве с бизонами, чем с барракудами. Несомненно, весьма заманчива идея о том, что убеждения “хранятся в голове”, подобно тому, как данные хранятся на жестком диске, зашифрованные систематическим кодом, который может быть уникальным для каждого человека, как уникальны отпечатки пальцев. Убеждения Жака будут записаны в его голове по-жаковски, а убеждения Шерлока – по-шерлоковски. Но эта привлекательная идея проблематична.

Допустим, мы вступили в золотой век нейрокриптографии, когда “когнитивные микронейрохирурги” получили возможность путем несложных операций *помещать* убеждения в мозг человека, записывая необходимую посылку в нейронах, само собой с использованием местного ментального языка. (Если мы сумеем научиться *читать* ментальные записи, вероятно, мы сумеем и *писать* на ментальном языке, используя достаточно тонкие инструменты.) Допустим, мы хотим поместить в мозг Тома следующее ложное убеждение: “У меня есть старший брат, который живет в Кливленде”. Допустим, когнитивный микронейрохирург сумеет сделать необходимую перенастройку – настолько пубоко и аккуратно, насколько вам хочется. Эта перенастройка либо повредит базовую рациональность Тома, либо нет. Рассмотрим оба варианта. Том сидит в баре, и друг спрашивает его: “У тебя есть братья или сестры?” Том отвечает: “Да, у меня есть старший брат, который живет в Кливленде”. Друг продолжает: “Как его зовут?” Что теперь? Том может ответить: “Зовут? Кого зовут? Боже, что я вообще несу? Нет у меня старшего брата! На мгновение мне показалось, что у меня есть старший брат и он живет в Кливленде!” Или же он может сказать: “Я не знаю, как его зовут” – и далее станет отрицать, что знаком с этим братом и говорить что-то вроде: “Я единственный ребенок, а мой старший брат живет в Кливленде”. Ни в одном из случаев наш когнитивный микронейрохирург не преуспел с внедрением в мозг нового убеждения. В первом случае нетронутая рациональность Тома вытесняет (одинокое, не имеющее поддержки) чужеродное убеждение, как только оно заявляет о себе. Мимолетное желание сказать “У меня есть старший брат, который живет в Кливленде” на самом деле даже нельзя назвать убеждением, поскольку оно скорее напоминает тик, подобно проявлению синдрома Туретта. Если бедный Том упрямится в своей убежденности, как во втором случае, его очевидная иррациональность по вопросу о старших братьях лишает его права считаться убежденным. Человек, который не понимает, что невозможно быть единственным ребенком и иметь при этом старшего брата, живущего в Кливленде, на самом деле не понимает произносимого им предложения, а человек не может быть убежденным в том, чего он по-настоящему не понимает, хотя и может повторять эту посылку, как попугай.

Этот пример из сферы научной фантастики подчеркивает безмолвную презумпцию ментальной компетентности, которая лежит в основе проверки подлинности всех убеждений: если вы не располагаете неограниченным репертуаром способов использования проверяемого убеждения (если это действительно убеждение) в различных контекстах, его нельзя считать убеждением ни в каком из знакомых нам смыслов. Если хирург сделал свою работу аккуратно, сохранив функции мозга, этот мозг сведет на нет его работу, как только представится случай, или же патологически окружит работу хирурга многочисленными цветистыми фантазиями (“Его зовут Себастьян, он цирковой акробат и живет в воздушном шаре”). Подобные фантазии известны науке: люди, страдающие от синдрома Корсакова (амнезии, которой часто подвержены алкоголики), могут невероятно убедительно рассказывать истории о собственном прошлом, в которых нет ни капли правды. Но само возникновение таких подробностей явно свидетельствует о том, что “посылки” в человеческом мозге не хранятся изолированно – даже ложное убеждение *требует* поддержки множества истинных убеждений и способности осознавать их следствия. Если человек не верит, что его старший брат – мужчина, дышит воздухом, живет к западу от Бостона и к северу от Панамы и так далее, совершенно точно нельзя сказать, что хирург сумел поместить в его голову *убеждение*.

Этот напор интуиции показывает нам, что невозможно иметь всего одно убеждение. (Считать, что у собаки четыре ноги, невозможно, не имея убеждений о том, что ноги – это конечности, четыре больше трех и т. д.)<sup>[15]</sup> Он демонстрирует и многое другое, но я не буду всего этого перечислять. Я не стану

и пытаться объяснить, как использовать вариацию этого инструмента мышления с другими целями, хотя вы можете сами покрутить регуляторы и проверить, что вам это даст. Я хочу описать целый ряд подобных инструментов мышления, прежде чем мы лучше проанализируем их свойства.



## 15. “Мой папа – врач”

Маленькую девочку спрашивают, чем занимается ее отец, и она отвечает: “Мой папа – врач”. Верит ли она своим словам? В некотором смысле верит, но что ей нужно знать, чтобы поверить в них *по-настоящему*? (Что если бы она сказала: “Мой папа – арбитражер” или “Мой папа – актуарий”?) Допустим, мы заподозрили, что она не понимает своих слов, и решили ее проверить. Должна ли она быть в состоянии переформулировать свое утверждение или снабдить его подробностями, сказав, что ее отец лечит больных? Достаточно ли ей знания, что ее папа – врач, а следовательно, не может быть ни мясником, ни пекарем, ни свечником? Знает ли она, что такое врач, если у нее нет представления о лжедокторе, знахаре, неквалифицированном специалисте? Если уж на то пошло, как много она должна понимать, чтобы знать, что папа – это ее отец? (Приемный отец? “Биологический” отец?) Само собой, ее понимание того, что значит быть врачом, а также что значит быть отцом, с годами углубится, а следовательно, углубится и ее понимание собственной фразы “Мой папа – врач”. Можем ли мы сказать – достаточно объективно, – сколько ей надо знать, чтобы понимать эту посылку “полностью”? Если понимание, как показывает этот пример, приходит постепенно, то убеждение, которое зависит от понимания, тоже должно формироваться постепенно, даже для таких простых посылок. Девочка **“вроде как”** верит, что ее отец работает врачом, и нельзя сказать, что она сомневается в этом, но все же она не до конца понимает значения собственных слов, а без понимания невозможно сформировать адекватное представление или убеждение.

## 16. Манифестная и научная картина мира

Прежде чем дальше разбираться с природой значения, нам необходимо возвести строительные леса. Следующий инструмент мышления позволяет взглянуть на многие вопросы с такого удачного ракурса, что его следует взять на вооружение всем и каждому, но пока он не получил большого распространения за пределами породившей его философии. В 1962 г. его разработал философ Уилфрид Селларс, который хотел показать, что именно наука говорит о мире вокруг нас. *Манифестная* картина мира – это мир, каким он предстает нам в повседневности, полный твердых предметов, цветов, запахов и вкусов, голосов и теней, растений и животных, людей и их вещей, причем не только столов и стульев, мостов и церквей, долларов и контрактов, но и таких неосознаваемых вещей, как песни, стихи, возможности и свобода воли. Представьте, сколько сложных вопросов возникает, когда мы пытаемся сопоставить все это с элементами *научной* картины мира: молекулами, атомами, электронами, кварками и т. п. Существуют ли *по-настоящему* твердые предметы? Физик Артур Эддингтон в начале двадцатого века писал о “двух столах” – один, твердый, был знаком нам по опыту, в то время как другой, состоящий из атомов, разделенных пустым пространством, напоминал скорее галактику, чем кусок дерева. Высказывались мнения, будто наука доказала, что *по-настоящему твердых предметов не существует*, а твердость – всего лишь иллюзия, но Эддингтон не заходил так далеко. Некоторые утверждали, что цвет – тоже иллюзия. Так ли это? Частицы электромагнитного излучения узкого спектра, различимого человеческим глазом (от инфракрасного до ультрафиолетового), не имеют цвета. Не окрашены и атомы, даже атомы золота. И все же цвет нельзя считать иллюзией: мы не уличаем Sony во лжи, когда в рекламе говорят, что телевизоры компании открывают нам мир цвета, и не подаем в суд на Sherwin-Williams, торгующую разными цветами в форме краски. А что насчет долларов? Сегодня подавляющее большинство долларов изготовлено не из серебра и даже не из бумаги. Они *виртуальны* – сделаны из информации, а не из вещества, прямо как стихи и обещания. Значит ли это, что они иллюзорны? Нет. Но искать их среди молекул не стоит.

Как известно, Селларс (1962, р. I) сказал: “Задача философии, если сформулировать ее абстрактно, заключается в том, чтобы понять, как вещи (в наиболее широком смысле) связаны между собой (в наиболее широком смысле)”. Я не встречал более удачного определения философии. Задача выяснить, как привести все знакомые *вещи* нашей манифестной картины мира в соответствие со всеми относительно неизвестными *вещами* научной картины мира, часто оказывается ученым не по зубам. Скажите-ка, доктор Физик, что такое *цвет*? Существуют ли цвета в соответствии с вашей теорией? Доктор Химик, дайте мне химическую формулу *халявы*! Безусловно (динь!), халява существует. Из чего она состоит? Хм-м-м. Может, никакой халявы и нет вовсе? Но в чем тогда различие – *химическое* различие! – между настоящей халявой и мнимой? Можно и дальше продолжать в том же духе, перечисляя многочисленные головоломки, которые только философы и пытались разрешить, но давайте лучше сделаем шаг назад, как советует нам Селларс, и осознаем, что существует два не похожих друг на друга представления о мире. Почему именно два? Может, их больше? Попробуем ответить на этот вопрос, начав описывать научную картину мира, и посмотрим, сумеем ли мы заметить появление манифестной картины мира с нашей наблюдательной точки.

Любой организм, будь то бактерия или представитель вида *Homo sapiens*, имеет в этом мире определенный набор вещей, которые многое для него значат и которые, как следствие, необходимо как можно лучше отличать и предвосхищать. Философы называют список вещей, обреченных на существование, *онтологией* (от греческого слова “сущее”, или “вещь”, вот так сюрприз!). Таким образом, онтологией обладает любой организм. Онтологию организма также называют его *Umwelt* (von Uexküll 1957; *Umwelt* в переводе с немецкого означает “среда” – и снова сюрприз!). *Umwelt* животного состоит в первую очередь из аффордансов (Gibson 1979) – того, что можно съесть, с чем можно спариться и чего следует опасаться, удобных проходов, наблюдательных позиций, укромных нор, высоких точек и так далее. *Umwelt* организма в некотором смысле представляет собой *внутреннюю* среду, “субъективную” и даже “нарциссическую” онтологию, состоящую исключительно из вещей, которые имеют первостепенное значение *для этого организма*, но при этом его *Umwelt* не обязательно носит внутренний или субъективный характер в значении сознательности. Фактически *Umwelt* представляет собой инженерную концепцию: представьте себе онтологию контролируемого компьютером лифта – набор всех вещей, которые необходимо отслеживать для корректной работы системы<sup>[16]</sup>. В одном из своих исследований Икскуль изучал онтологию клеща. Мы можем предположить, что *Umwelt* морской звезды,

червя или маргаритки имеет больше общего с онтологией лифта, чем *naun Umwelt*, который фактически представляет собой нашу манифестную картину мира.

Наша манифестная картина мира, в отличие от онтологии или *Umwelt* маргаритки, действительно имеет декларационный и в высшей степени субъективный характер. Это мир вокруг нас, который интерпретируем мы сами<sup>[17]</sup>. Однако, подобно онтологии маргаритки, существенная часть нашей манифестной картины мира была сформирована под влиянием естественного отбора, продолжавшегося целую вечность, и стала частью нашего генетического наследия. В одном из моих любимых примеров многообразие *Umwelt* показывается на примере муравья и насекомоядной птицы (Wimsatt 1980). Птица выслеживает насекомых поодиночке и вынуждена компенсировать их беспорядочное движение высокой частотой слияния мельканий (фактически она видит больше “кадров в секунду”, чем мы, а следовательно, фильм для нее был бы похож на слайд-шоу). Муравей просто подходит к месту скопления муравьев и пускает в ход свой длинный язык. Философ сказал бы, что понятие “муравей” для муравья *собирательное*, подобно понятиям “вода”, “лед” и “мебель”, в отличие от *несобирательных* понятий “оливки”, “капли” и “стулья” (их можно сосчитать). Когда муравей видит лакомый кусок *муравья*, он заглатывает его целиком, не замечая отдельных муравьев, как мы не замечаем отдельные молекулы глюкозы, когда едим конфету.

Большая часть нашей манифестной картины мира не записана в нашем генетическом коде, а каким-то образом сформирована в раннем детстве. Слова для нас – очень важная категория сущего. Они выступают в качестве проводников манифестной картины мира, но вполне возможно, что способность классифицировать некоторые события мира как слова и желание говорить хотя бы отчасти генетически унаследованы – подобно способности птицы на лету ловить насекомых или желанию осы строить гнездо. Даже в отсутствие грамматики, которая связывает их в предложения, слова могут использоваться в качестве ярлыков, помогающих создать категории внутри манифестной картины мира: мама, песик, печенье. Могли бы вы сформулировать ясный образ халявы, ошибки или обещания – не говоря уже об оглушительном успехе или холостяке, – не прибегая к помощи слов, которыми я только что их назвал? Изучив список любимых инструментов мышления Дага Хофштадтера, мы уже увидели, как термины структурируют и организуют наше сознание, обогащая наши личные манифестные картины мира *вещами* – пороховыми бочками, пустыми словами и обратной связью, – заметить которые в ином случае практически невозможно.

## 17. Народная психология

Вероятно, самым важным паттерном в рамках нашей манифестной картины мира следует считать паттерн *народной психологии*, который служит опорой для многих других значимых для нас категорий. Я предложил этот термин в его текущем значении в 1981 г., однако он, очевидно, и ранее появлялся в работах Вильгельма Вундта, Зигмунда Фрейда и других мыслителей (*Volkpsychologie*), где описывал определенную черту немецкого национального характера (дух немецкого народа – не будем в это уплываться). Как и многие другие, я этого не учел и предложил использовать термин “народная психология” для обозначения присущего каждому из нас *таланта* видеть в людях вокруг нас – а также в животных, и в роботах, и даже в ничем не примечательных термостатах – *агентов*, имеющих информацию о мире, в котором они функционируют (*убеждения*), и цели (*желания*), которых они стремятся достичь, и выбирающих наиболее *рациональный* порядок действий с учетом своих убеждений и желаний.

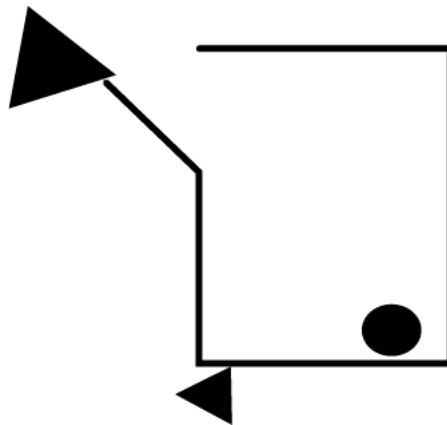
Некоторые исследователи предпочитают называть народную психологию “теорией сознания” (*theory of mind*), но этот термин кажется мне некорректным, поскольку он отмечает вопрос о том, откуда у нас этот талант, наталкивая на мысль, что мы разработали и успешно применяем некоторую теорию. Подобным образом мы сказали бы, что у вас есть *велосипедная теория*, если вы умеете ездить на велосипеде, и *теория питания*, которая объясняет вашу способность не умирать с голоду и не есть песок. На мой взгляд, это не слишком удачный способ мыслить об этих навыках. Поскольку никто не возражает, что человек обладает талантом к интерпретации, но никто не знает точно, как ему удастся быть настолько компетентным, думаю, лучше пока отказаться от слова “теория” и использовать несколько более нейтральный термин. Академическая, или научная, психология тоже занимается объяснением и прогнозированием сознания людей, и в рамках нее действительно существуют различные теории: бихевиоризм, когнитивизм, нейровычислительные модели, гештальтпсихология и др. Чтобы преуспеть в сфере народной психологии, нам не нужно традиционное образование. По аналогии, *народная физика* – это *талант* ожидать, что жидкости потекут, не имеющие опоры предметы упадут, горячие вещества обожгут нас, вода утолит нашу жажду, а под лежащий камень вода не затечет. Весьма интересен вопрос, каким образом наш мозг так легко генерирует *все эти* ожидания, которые почти всегда оказываются верными, даже если мы никогда не изучали физику.

Народная психология – это то, что “всем известно” о собственном сознании и сознании окружающих: люди чувствуют боль, голод и жажду и умеют отличать одно от другого, они помнят события прошлого, многое предугадывают, видят все на расстоянии видимости, слышат все на расстоянии слышимости, обманывают и обманываются, понимают, где они находятся, узнают друг друга и так далее. Уверенность, с которой мы делаем эти предположения, поражает, ведь на самом деле мы крайне мало знаем о том, что творится в головах у этих людей (не говоря уже о животных). Мы настолько уверены во всем этом, что нам приходится специально *дистанцироваться* от происходящего, чтобы вообще заметить, что мы это делаем.

Художники и философы соглашались в одном: перед ними стоит задача “сделать привычное необычным”<sup>[18]</sup>. Великие проявления творческого гения заставляют нас вырваться из плена привычности и **выпрыгнуть** к новым горизонтам, где мы сможем взглянуть на знакомые, будничные вещи новыми глазами. Ученые совершенно с этим согласны. Моментом откровения для Ньютона стал тот миг, когда он задался странным вопросом, почему яблоко упало с дерева *на землю*. (“Почему бы ему *не упасть*? – спрашивает любой, кто далек от гениальности. – Оно ведь *тяжелое!*” Как будто такого объяснения достаточно!) Если вы зрячий, но общаетесь с несколькими слепыми друзьями, вероятно, вы подтвердите, что *все равно* время от времени указываете руками на вещи или очерчиваете контуры предметов перед их невидящими глазами, чтобы ваши друзья поняли то, что понимаете вы. “По умолчанию” в присутствии другого человека вы ожидаете, что вы оба видите одно и то же, слышите одно и то же, чувствуете одни и те же запахи. Вы едете по шоссе со скоростью девяносто километров в час и нисколько не переживаете по поводу машины, которая на той же скорости едет вам навстречу по соседней стороне дороги. Откуда вы знаете, что не случится аварии? Сами того не сознавая, вы предполагаете, что водитель (которого вы даже не видите и почти наверняка не знаете) *хочет* жить и *знает*, что для этого ему следует держаться правой стороны дороги. Обратите внимание, что вам стало бы гораздо тревожнее, если бы по радио сообщили, что в этот день на вашем шоссе тестируют новый беспилотный автомобиль. Да, он *создан* быть безопасным, и разработали его в *Google*, но вы гораздо более уверены в рациональности и компетентности

среднего водителя-человека, чем в способностях хваленого робота (если только все это происходит не субботним вечером).

Откуда “все знают” народную психологию? Неужели это все-таки теория сознания, которую человек постигает в детстве? Можно ли сказать, что некоторые ее элементы мы знаем с рождения, или же мы учим ее полностью – но как и когда? В последние тридцать лет этим темам посвящалось огромное количество исследований. Я полагаю (даже после тридцати лет исследований), что это не столько теория, сколько практика – некий способ познания мира, настолько естественный для человека, что в нашей голове, должно быть, есть какая-то генетическая основа для него. Отчасти он постигается “с молоком матери”, поэтому, если бы вы выросли в полной изоляции от людей, в народной психологии вы бы не преуспели (и нажили бы множество других серьезных проблем), но человек испытывает сильное желание считать агентами объекты, которые движутся беспорядочно (*не как маятник или мяч, катящийся с горы*). Мне нравится пример, который используется для демонстрации этого желания на вводном курсе психологии: в 1944 г. Фриц Хайдер и Марианна Зиммель создали короткий анимационный фильм (используя покадровую анимацию), где два треугольника и круг перемещаются по экрану, то заходя на огороженную территорию, то выходя обратно. Геометрические фигуры совсем не похожи на людей (или животных), но считать их взаимодействие лишенным смысла (*purposeful*) практически невозможно: нам кажется, что ими руководят вожеление, страх, смелость и гнев. Этот старинный фильм можно найти на многих сайтах, например на <http://www.psychexchange.co.uk/videos/view/20452/>.



Мы рождаемся с “детектором агентов” (Barrett 2000; см. также Dennett 1983), который заводится с пол-оборота. Когда он допускает ошибку, что часто случается в напряженных ситуациях, мы видим призраков, гоблинов, чертей, лепреконов, фей, гномов, демонов и подобных существ, хотя на самом деле это просто ветки, обрушившиеся стены или скрипящие двери (Dennett 2006a). С детских лет мы без труда и без отчета считаем других агентами, причем эти агенты не просто счастливы, рассержены, озадачены или испуганы, но *посвящены в тайну*, или *гадают, куда свернуть*, или даже *не хотят принимать предлагаемые условия*. Это не операция на мозге и не высшая математика – это просто. На мой взгляд, сила народной психологии и простота ее использования объясняются лежащими в ее основе упрощающими допущениями. Народная психология *напоминает* идеализированную научную модель – максимально абстрактную и свободную от всего лишнего. Я называю эту модель *интенциональной установкой*.

## 18. Интенциональная установка

Пока все просто. Почти все мы хороши в народной психологии<sup>[19]</sup>. У нас талант считать, что другие (как и мы сами) обладают сознанием, и обычно мы не прикладываем к этому усилий, как не прикладываем усилий к тому, чтобы дышать. Мы без всяких задних мыслей зависим от этого таланта, и он почти никогда нас не подводит. Почему народная психология дается нам так просто? Как она работает? Здесь нам стоит сделать паузу и возвести очередной уровень лесов, которые существенно облегчат нам жизнь, когда мы двинемся дальше.

Как работает народная психология? Чтобы разобраться в этом, посмотрим, как она работает, когда мы применяем ее *не к людям*. Допустим, вы играете в шахматы с компьютером. Вы хотите победить, и единственный способ приблизить победу – попытаться предугадать ответные ходы компьютера: “Если я пойду слонем, компьютер его съест; если же я пойду пешкой, компьютеру придется передвинуть ферзя...” Откуда вам знать, что сделает компьютер? Вы заглядывали в него? Изучали его шахматную программу? Конечно нет. Вам это и не нужно. Вы делаете уверенные предсказания на основе вполне очевидных допущений, что компьютер

1. “знает” правила и “умеет” играть в шахматы;
2. “хочет” победить, а также
3. “увидит” эти возможности, как видите их вы, и поступит соответствующим образом (то есть рационально).

Иначе говоря, вы считаете компьютер *хорошим* шахматистом – или, по крайней мере, не шахматистом-идиотом, играющим себе во вред. *Иначе* иначе говоря, вы обращаетесь с ним так, словно это человек, наделенный сознанием. И совсем иными словами, используя народную психологию, чтобы предвидеть и понимать его ходы, вы *принимаете интенциональную установку*.

Интенциональная установка – это стратегия интерпретации поведения объекта (человека, животного, предмета, чего угодно) *в качестве* рационального агента, который “выбирает”, как “действовать”, “руководствуясь” своими “убеждениями” и “желаниями”<sup>[20]</sup>. Пугающие кавычки, в которые заключены эти термины, привлекают внимание к тому факту, что ряд их стандартных коннотаций можно отбросить, чтобы использовать их главные характеристики – их роль в практическом мышлении, а следовательно, в предсказании поведения практических мыслителей. Все, что поддается эффективному и подробному предсказанию из интенциональной установки, по определению представляет собой *интенциональную систему*. Как мы увидим, множество удивительных и сложных объектов, не имеющих мозгов, ушей, глаз или рук, а следовательно, не имеющих сознания, все равно можно считать интенциональными системами. Главный фокус народной психологии, так сказать, находит применение и за пределами мира человеческого взаимодействия. (Мы увидим, как его применить не только в компьютерных технологиях и когнитивной нейронауке, но и в эволюционной биологии и биологии развития – и это только самые важные сферы.)

Я предлагаю просто *отложить* неудобный вопрос о том, что *на самом деле* обладает сознанием и представляет собой *надлежащее* вместилище для интенциональной установки. Каким бы ни был верный ответ на этот вопрос – если верный ответ на него вообще существует, – он не отменит того факта, что интенциональная установка прекрасно *работает* для предсказаний в других сферах и лишь немногим лучше функционирует в повседневной жизни, где мы применяем народную психологию при взаимодействии с другими людьми. Этот мой *шаг* раздражает некоторых философов, которые готовы бить тревогу, потому что им важно первым делом дать четкие определения сознанию, убеждению и желанию и лишь потом идти дальше. *Определите термины, сэр!* И не подумаю. Это преждевременно. Сначала я хочу изучить силу и спектр применения этого ловкого фокуса, интенциональной установки. Увидев, на что она годится и почему, мы сможем вернуться и спросить себя, нужны ли нам еще эти строгие, формальные определения. Я выбираю этот путь, чтобы *раздробить* сложную проблему, вместо того чтобы попытаться проплотить (и переварить) все разом. Многие инструменты мышления, о которых я расскажу в этой книге, хорошо помогают с дроблением задач, позволяя приблизительно нащупать несколько “фиксированных” точек и составить общее представление об очертаниях проблемы. В книге “Пространство для маневра” (1984а) я сравнил свой метод с работой скульптора, который сначала высекает в мраморе общую форму статуи, а затем работает осторожно, сдержанно, руководствуясь

принципом последовательного приближения. Очевидно, многим философам этот способ не подходит – они предпочитают четко определить (по крайней мере, так им кажется) границы своих проблем и возможных решений, прежде чем выдвигать какие-либо гипотезы.

### Три установки

Давайте посмотрим, в чем заключается сила интенциональной установки, сравнив ее с тактикой прогнозирования. Начнем с определения трех основных установок (в рамках которых можно выделить подкатегории, но здесь нужды в этом не возникает): физической установки, конструктивной установки и интенциональной установки.

*Физическая установка* – это используемый в физических науках стандартный трудоемкий метод предсказаний, который предполагает применение всех известных нам законов физики и имеющихся данных о строении исследуемых объектов. Предсказывая, что выпущенный у меня из руки камень упадет на землю, я использую физическую установку. Как правило, физическая установка – единственная работающая стратегия для интерпретации неодушевленных нерукотворных объектов, однако, как мы увидим, существует ряд исключений. Любой физический объект, будь он создан природой или человеком, одушевлен или неодушевлен, подчиняется законам физики, а следовательно, его поведение, *в принципе*, может объясняться и прогнозироваться из физической установки. Неважно, что именно я выпускаю из руки – будильник или золотую рыбку, – я делаю одинаковое предсказание о падении этого предмета, опираясь на одинаковые основания. Обычно физическая установка не дает оснований предполагать, что будильник или золотая рыбка поведет себя более интересным образом.

Поведение будильников, которые представляют собой сконструированные объекты (в отличие от камней), также поддается прогнозированию более затейливым образом – прогнозированию из *конструктивной установки*. Допустим, я посчитал новый объект будильником. Я могу сделать вывод, что *если* я нажму несколько кнопок определенным образом, *то* через некоторое время будильник издаст громкий звук. Мне не нужно выводить конкретные физические законы, которые объясняют эту удивительную закономерность; я просто *полагаю*, что объект сконструирован особым образом – такие конструкции мы называем будильниками – и будет выполнять заявленную функцию. Прогнозы из конструктивной установки сопряжены с большим риском, чем прогнозы из физической установки, поскольку нам необходимо учитывать ряд дополнительных допущений:

1. что объект *действительно* сконструирован именно так, как я думаю, а также
2. что он будет функционировать в соответствии со своей конструкцией – иными словами, что он не даст сбой.

Иногда при конструировании объектов возникают ошибки, а иногда сконструированные объекты ломаются. Камень не может функционировать неправильно, поскольку его функция вообще не определена, а если он сломается надвое, в результате появятся два камня, а не один, расколотый пополам. Когда сконструированный объект относительно сложен (к примеру, цепная пила в сравнении с топором), умеренная цена риска более чем компенсируется невероятной простотой прогнозирования ее поведения. Никто не станет опираться на фундаментальные законы физики при прогнозировании поведения цепной пилы, имея в своем распоряжении удобную схему ее движущихся частей.

Еще более рискованной и изменчивой представляется интересующая нас *интенциональная установка* – подвид конструктивной установки, в котором сконструированный объект считается своего рода агентом, имеющим убеждения и желания и наделенным достаточной рациональностью, чтобы совершать необходимые действия на основании этих убеждений и желаний. Будильник так прост, что нам, строго говоря, нет нужды обращаться к этому затейливому антропоморфизму, чтобы понять, почему он ведет себя именно так, а не иначе, но если артефакт устроен гораздо сложнее будильника, использование интенциональной установки приносит больше пользы – и становится практически обязательным. Чтобы не упустить ничего важного, давайте пошагово разберем механизм принятия интенциональной установки на уже знакомом нам примере играющего в шахматы компьютера:

Сначала перечислим все возможные ходы, которые доступны компьютеру, когда наступает его очередь ходить (обычно их несколько десятков).

Затем проранжируем все возможные ходы от лучшего (самого разумного, самого рационального) к худшему (самого глупого, самого бесперспективного).

Наконец, сделаем предсказание: компьютер сделает лучший ход.

Возможно, вам будет сложно однозначно определить лучший ход (компьютер может “проанализировать” ситуацию лучше вас!), но почти всегда вы сумеете отбросить все ходы, кроме четырех-пяти самых разумных, что и так дает вам огромные возможности для прогнозирования. Вы *могли бы* повысить точность прогнозирования и заранее определять, какой именно ход сделает компьютер – ценой огромных усилий и временных затрат, – вернувшись в конструктивную установку. Для этого вам нужно получить “исходный код” программы (см. лаву 27) и “смоделировать его вручную”, сделав миллионы миллиардов крошечных шагов, которые компьютер предпримет, решая, как ответить на задуманный вами ход. Так вы точно поймете, как поступил бы компьютер, если бы вы сделали свой ход, но отведенное на раздумье время истекло бы задолго до того, как вы достигли бы результата: вам не хватило бы и целой жизни! Слишком много информации! И все же использовать конструктивную установку *гораздо* проще, чем вернуться к физической установке и попробовать описать поток электронов, возникающий после нажатия на клавиши компьютера. Таким образом, физическая установка совершенно непрактична при прогнозировании и объяснении ходов компьютера, а конструктивная установка требует слишком большого труда, если только не использовать другой компьютер (а это жульничество). Используя интенциональную установку, человек увильивает от трудоемкого сбора информации и расчетов и делает рискованное, но правдоподобное предположение: компьютер поступит “рационально”, то есть найдет и сделает лучший ход (учитывая, что он хочет победить и знает положение и вес всех фигур на доске). Во многих ситуациях, особенно если лучший ход для компьютера так очевиден, что он считается “вынужденным” или “элементарным”, интенциональная установка позволяет предсказать ход почти со стопроцентной точностью без особенного труда.

Очевидно, что интенциональная установка эффективно работает при прогнозировании поведения играющего в шахматы компьютера, поскольку он был сконструирован с целью “обдумывать” и выбирать лучшие ходы в рамках в высшей степени рационалистической игры в шахматы. Если компьютерная программа управляет нефтеперерабатывающим заводом, почти столь же очевидно, что она будет предпринимать различные шаги, анализируя условия, которые в большей или меньшей степени диктуют ее поведение с учетом ее общего целевого назначения. Здесь важно, что программа по умолчанию считается *совершенной или рациональной*, хотя, если программист был некомпетентен, она редко будет работать именно так, как должна работать в конкретных обстоятельствах, по мнению экспертов. Если информационные системы (или системы контроля) сконструированы хорошо, логика их действий распознаваема и в высокой степени предсказуема, даже если авторы программ, вопреки устоявшимся нормам, не снабдили исходный код “комментариями”, объясняющими логику работы сторонним наблюдателям. (Подробнее об этих компьютерных тонкостях позже.) Чтобы прогнозировать поведение системы, нам *ничего* не нужно знать о компьютерном программировании; нам нужно лишь понимать, каковы рациональные требования, предъявляемые к управлению нефтеперерабатывающим заводом.

Теперь мы видим, как и зачем интенциональная установка работает для прогнозирования *нашего* поведения. Воспринимая *друг друга* в качестве интенциональных систем, мы даем себе возможность не углубляться в детали процессов, происходящих у нас в голове, и бессознательно полагаемся на факт, что в удивительно точном первом приближении люди рациональны. Неожиданно оказавшись в новом человеческом сценарии, обычно мы понимаем его без труда, даже не отдавая себе отчета, благодаря нашей врожденной способности понимать, во что должны верить люди (истину о том, что им известно) и чего они должны желать (что для них хорошо).

Никто не спорит, что мы выдвигаем очень точные гипотезы, используя народную психологию, однако споры о том, чем объясняется этот дар, не утихают. Может, мы заучиваем десятки или сотни тысяч “законов природы” наподобие

“Если человек бодрствует, его глаза открыты и он стоит лицом к автобусу, он склонен полагать, что перед ним стоит автобус” и

“Всекий раз, когда люди полагают, что могут получить выгоду по низкой цене, они склонны сотрудничать с другими, даже с незнакомцами”?



Или же эти грубо отесанные законы генерируются по запросу присущим нам чувством, что именно таковы рациональные ответы в соответствующих обстоятельствах? Я считаю, что верно второе. Хотя огромное количество стереотипных моделей поведения действительно описывается подобными обобщениями (которые, в принципе, могут усваиваться постепенно в течение всей нашей жизни), достаточно сложно создать научно-фантастический сценарий такой новизны, чтобы он шел вразрез с любыми прогнозами и никто просто не мог бы представить, каким могло бы быть поведение людей в описываемых обстоятельствах. Возникает закономерный вопрос: “Что бы *ты* сделал, если бы такое случилось с тобой?” На него есть куча бесполезных ответов вроде: “Я бы сразу свалился замертво”, – но среди них находится и характерно *рациональный*: “Надеюсь, я бы догадался, что стоит сделать X”. Наблюдая за героями, которые в этих нестереотипных обстоятельствах поступают на редкость разумно, мы без труда понимаем, что они делают и почему. Подобно нашей способности понимать новые предложения родных языков – предложения, которые мы никогда прежде не слышали, – наша способность понимать огромное множество человеческих взаимодействий свидетельствует о наличии генеративной способности, в некоторой степени от рождения присущей нормальным людям.

Мы столь же естественно и безотчетно распространяем интенциональную установку на животных. Без нее не обойтись, если мы хотим поймать коварного зверя, но она также полезна, если мы пытаемся организовать свое понимание поведения более простых животных и даже растений. Моллюск ведет себя рационально, опираясь на свое ограниченное представление о мире. Нас не удивляет, что деревья, которые чувствуют постепенное приближение конкурентов (поскольку все больше попадающего на них солнечного света отражается от высоких зеленых объектов поблизости), перераспределяют ресурсы, чтобы расти быстрее. В конце концов, в таких обстоятельствах для растения это умный ход. В мире артефактов примитивную интерпретацию из интенциональной установки выдерживает даже презренный термостат: он хочет поддерживать температуру на заданном уровне, часто измеряет температуру, чтобы иметь актуальное представление о ней, сравнивает полученное значение с заданной температурой и действует соответствующим образом. Именно так можно объяснить работу термостата ребенку, не вдаваясь в технические подробности.

Простая теория интенциональных систем объясняет, как и почему мы понимаем поведение огромного количества сложных объектов, считая их *агентами*. Нельзя сказать, что эта теория *непосредственно* объясняет работу внутренних механизмов, которые каким-то образом достигают предсказанного рационального ориентира. Интенциональная установка дает “спецификацию” интенциональной системы, технические требования к ее работе – к примеру, что она должна распознавать, запоминать и делать, – а осуществляют это техническое задание инженеры (или эволюция и развитие, если в качестве интенциональной системы выступает живой организм). Дайте мне агента, который понимает разницу между долларовыми и 10-долларовыми купюрами, умеет давать сдачу и определять фальшивки, а также готов и способен круглосуточно доставлять покупателям необходимые продукты. Эта характеристика из интенциональной установки может быть как описанием торгового автомата, так и примитивной должностной инструкцией продавца круглосуточного магазина. При этом совершенно неважно, какое строение имеет искомая сущность и какими еще талантами она наделена.

Такая равнозначность, или нейтральность, не баг, а фича, как выражаются программисты. Как мы увидим в последующих трех главах, она позволяет теории интенциональных систем играть основополагающую роль при наведении мостов через пропасть смятения между нашими сознаниями и мозгами. Не углубляясь в тонкости, можно сказать, что она позволяет нам увидеть сходства между “реальными” убеждениями (убеждениями *личностей*) и “всего лишь” их имитацией (характерной для торговых автоматов, животных, маленьких детей и, что особенно важно, субличностных *фрагментов* личностей). Чтобы делать прогнозы, мы можем использовать интенциональную установку и давать спецификацию *компетенций* подсистем мозга, еще не зная, как именно реализованы эти компетенции. Мы анализируем большую, сложную “реальную” личность, разбивая ее на множество субличностных агентов, имеющих собственные задачи и методы, а затем анализируем этих субличностных агентов, разбивая их на множество еще более простых и плутовых агентов. В итоге мы получаем интенциональные системы, которые достаточно просты, чтобы без труда описать их из интенциональной установки. Перекидывая мост через пропасть между **народной психологией на личностном уровне** и **субличностной** активностью нейронных сетей, мы задействуем воображение, которому идет на пользу это принципиальное послабление условий, выдвигаемых философами для убеждений и желаний (настоящего, взрослого) человека. Где на пути от человека к неодушевленным объектам “настоящие” убеждения и желания уступают место “мнимым”? Как мы увидели в главе 15 и как увидим снова, применив несколько других инструментов мышления, это стремление провести черту безосновательно.

Интенциональная установка широко применяется в информатике и психологии животных, и теория интенциональных систем объясняет почему. Некоторые теоретики эволюционной биологии утверждают, что обходятся без нее, но на самом деле они обманывают себя, как мы увидим в разделе об эволюции<sup>[21]</sup>.

## 19. Личностное и субличностное

Ваши глаза не видят – видите вы. Ваш рот не наслаждается вкусом шоколадного торта – наслаждаетесь вы. Ваш мозг не злится из-за острой боли в плече – злитесь вы. Ваша рука не подписывает контракт – подписываете вы. Ваше тело может чувствовать возбуждение, но влюбляетесь все равно вы. Это не *просто* “грамматические” тонкости вроде того, как мы говорим “идет дождь”, когда на улице гроза, вместо того чтобы сказать “гроза идет дождем”. И это не *просто* вопрос единства определений. Люди часто отмахиваются: “Разве это не просто игра в слова?” – имея в виду, что обычно не так уж важно, как именно мы “определяем термины”. Но порой определения имеют решающее значение, и это как раз один из таких случаев. Мы говорим о *личностях* и о том, что они могут делать и выносить, опираясь на ряд важных фактов.

На первый взгляд, существуют вещи, которые под силу делать целой личности, но не под силу ни одному из ее фрагментов. Это почти так, но воображение подсказывает нам жутковатую мысль, что в таком случае личность (грубо говоря) представляет собой фрагмент человеческого тела, а именно полноценно функционирующий мозг. (*Обладаете* вы мозгом – или вы *есть* мозг? Ответить на этот вопрос не так легко.) Если отрубить мне руки, я все равно смогу подписать контракт (держа ручку ногой или отдав указание голосом), но если прекратить работу моего мозга, никакие действия моих рук и ног нельзя будет приравнять к подписанию контракта. Если выколоть мне глаза, я потеряю способность видеть, но смогу вернуть ее, установив плазные протезы, которые уже не кажутся элементом научной фантастики. Что если “ампутировать” некоторые части моего мозга? Если удалить затылочную кору, оставив невредимыми глаза и зрительный нерв, я стану “кортикально слеп”, но, возможно, сохраню остаточное зрение (к примеру, знаменитое “слепозрение”). Несомненно, ампутировав немного больший фрагмент мозга, мы сможем избавиться от слепозрения, сохранив *вам* жизнь. Нам нравится мысль, что в процессе постепенного исключения остальных чувств – слуха, осязания, обоняния и вкуса – можно урезать мозг до вместилща *вас*, обнаружив таким образом, где находится личность, и поняв, что она собой представляет. Эта мысль интересна, но ошибочна. Многочисленные функции мозга так тесно переплетаются и так активно взаимодействуют, что места, “где все сходится” в единое сознание, просто не существует<sup>[22]</sup>. Если уж на то пошло, многие способности, склонности, предпочтения и причуды, которые делают вас *вами*, зависят от других особенностей вашего тела, а не только от мозга; и неизменно популярный философский мысленный эксперимент о пересадке мозга (кем бы предпочли стать вы – “донором” или “реципиентом”?) основан на искажающей реальность идеализации. Как я однажды сказал, “нельзя отсечь меня от моего тела, оставив чистый срез” (1996а, р. 77).

Вероятно, главное свойство интернета – его децентрализация. У него нет ни штаб-квартиры, ни головного офиса, на который можно сбросить бомбу, чтобы его отключить. Его составные части характеризуются высоким запасом мощности и оперативной гибкостью, поэтому при отключении одной из частей он “выходит из строя с достоинством”, если выходит из строя вообще. У умного компьютера HAL из “Космической одиссеи 2001 года” был “логический центр памяти” – комната, полная банков данных. Отсоединив их один за одним, Дэйв выключил HAL навсегда. У интернета подобного центра нет. Хотя природа не наделила нас таким высоким уровнем распределенной неуязвимости, в своем теле *вы* в значительной степени децентрализованы, а отдельные части вашего тела имеют запас оперативной гибкости. Ваш мозг на удивление пластичен (способен к трансформации), поэтому вы можете и дальше быть собой, стремиться к мечтам, проклинать врагов, плести интриги, переживать трагедии и триумфы после отсечения важных, но не “жизненно важных” частей тела. Именно поэтому *вы* можете обладать способностями, которыми не обладает ни одна из ваших частей. Эту мысль можно также вывернуть наизнанку: единственный способ понять, какими способностями обладают *части* живого тела, – это проанализировать, какой вклад они вносят в работу всей грандиозной системы.

Вот еще несколько примеров. Ваш мозг не понимает русский – его понимаете вы. Ваш мозг не считает шутку смешной – так считаете вы. Даже если работа соответствующих структур мозга обуславливает ваше понимание и восприятие, эти структуры не могут функционировать должным образом без тренировки и поддержки, которую им годами оказывают органы чувств, конечности и другие эффе́кторы<sup>[23]</sup>.

Получается, что мы не просто так считаем *личность* – этого выносливого, сознательного, рационального агента, которому принадлежит тело, – *субъектом* большинства наших ежедневных

атрибуций: это *вы* сделали ошибку, победили в гонке, влюбились в Лесли, научились сносно говорить по-французски, хотите поехать в Бразилию, предпочитаете блондинок, оклеветали кого-то. (См. также главу 62.) Вы проголодались, устали и рассердились – и все благодаря субличным элементам и ничему более.

Но что это за элементы? Неужели человеческое тело состоит из них, как из кирпичиков? Ответом будет да, если мыслить на уровне мельчайших частиц – атомов. И нет – на любом другом уровне, от молекул, клеток и выше. Белки – рабочие лошади ваших клеток – это невероятно компетентные и проницательные маленькие роботы. Мы можем назвать их наноботами. Нейроны, которые отвечают за распространение и переключение сигналов в мозге, а также за его адаптивность, можно считать более многофункциональными и умелыми роботами – назовем их микроботами. Они формируют коалиции, соревнуются и сотрудничают друг с другом в рамках более крупных структур, коммуницируют, подавляют друг друга, анализируют поток информации от органов чувств, пробуждают дремлющие информационные структуры “в памяти” (которая не имеет отдельного места в мозге) и руководят сложными каскадами сигналов, приводящих в движение мускулы.

Все уровни выше простейших атомных кирпичиков в некоторой степени характеризуются *субъектностью*. Иными словами, они поддаются интерпретации в качестве *интенциональных систем*. На молекулярном уровне (двигательные белки, ферменты, осуществляющие коррекцию ДНК, контроллеры триллионов шлюзов в клеточных мембранах и т. п.) их компетенции сильно “роботизированы”, но все равно поразительны, будто армии марширующих метел в “Ученике чародея” или демон Максвелла, если взять два вымышленных примера. На клеточном уровне отдельные нейроны демонстрируют большую склонность к экспериментам: они ищут более надежные связи и меняют принципы своей активации на основании недавнего опыта. Они напоминают скорее узников или рабов, чем обычные машины (вроде белковых наноботов); их можно представить в качестве запертых в камерах близоруких нервных клеток, вовлеченных в массовые проекты, которых им не понять, но готовых совершенствоваться, меняя принципы своего функционирования. На высших уровнях близорукость отступает, а группы клеток – пучки, колонки, ганглии, “ядра” – получают специализированные роли, зависящие от все большего количества условий, включая условия внешнего мира. Здесь степень субъектности становится выше, поскольку “выполнение работы” требует немалой *рассудительности* и даже *умения принимать решения*.

Эти агенты подобны белым воротничкам – аналитикам и управленцам, имеющим конкретные обязанности. Как и все белые воротнички, они не лишены здоровой соревновательности и готовы принять все, с чем сталкиваются в процессе своей деятельности, и даже подмять под себя соседей или тех, кто состоит с ними в контакте. На этом уровне компетентности выступающие в качестве агентов субличные элементы представляют собой умные кирпичики, и мы начинаем понимать – по крайней мере, в общих чертах, – как сложить из них цельную, способную к пониманию личность. (“Требуется сборка”, как написано на коробке с деталями велосипеда, но нам хотя бы не нужно резать и гнуть металл, делать гайки и болты.)

Представить немыслимое – как личность может состоять из (одних лишь) бездумных молекул – можно двумя способами: пойти от частного к общему, как мы только что поступили, или от общего к частному, начав с цельной личности и изучая, каково наименьшее число очень умных гомункулов, способных договориться между собой и выполнять все задачи, которые необходимо выполнять, чтобы личность продолжала функционировать. Первым по пути от общего к частному прошел Платон. Он предположил, что душа состоит из трех частей-агентов, аналогичных правителям, стражам и работникам, или разумному, яростному и страстному началу, но за прошедшие с тех пор два тысячелетия было выявлено немало причин, по которым его анализ нельзя считать удачным. Предложенное Фрейдом около ста лет назад деление на Ид, Эго и Супер-Эго было несколько лучше, но активные попытки поделить сознание на подсознания начались только после изобретения компьютера и появления науки об искусственном интеллекте (ИИ), задача которой изначально заключалась в анализе когнитивных способностей цельной (взрослой, сознательной, использующей язык) личности путем представления ее в качестве огромной сети субличных специализированных элементов, таких как генератор целей, поисковик воспоминаний, оценщик планов, анализатор восприятия, анализатор предложений и так далее.

## 20. Каскад гомункулов

В старом как мир стремлении понять сознание мыслители часто поддавались искушению представить внутреннего агента, маленького человечка – *homunculus* по-латыни, – который сидит в диспетчерской мозга и выполняет всю умную работу. Если представить человеческую нервную систему, скажем, в виде огромного телефонного коммутатора (философы обожали этот прием еще в 1950-х и 1960-х), возникнет проблема работающего с ним оператора: можно ли считать, что его сознание представляет собой телефонный коммутатор меньшего размера, имеющий собственного оператора, сознание которого, в свою очередь, сравнимо с... Эта бесконечная регрессивная последовательность обрекает на провал любую теорию, предполагающую наличие центрального гомункула.

Но, может, проблема не во введении гомункула как такового, а во введении *центрального* гомункула? В своей первой книге “Содержание и сознание” (1969) я допустил серьезную ошибку, не удержавшись от насмешки. Я написал:

“Человечек внутри мозга”, “дух в машине” Райла не решает проблему сознания. Хотя нельзя исключить, что аналогии с “письменностью мозга” найдется какое-нибудь полезное применение, кажется, что она лишь заменяет маленького человечка целым комитетом. [p. 87]

И чем же плох этот комитет? (*Aga!* Вот и критика моей попытки доведения до абсурда!) В итоге я пришел к выводу (в книге “Мозговые штурмы”, 1978а), что идея заменить человечка внутри мозга целым комитетом не так уж плоха – пожалуй, это одна из лучших идей когнитивной науки. Такой была классическая стратегия символического искусственного интеллекта GOF AI (“старый добрый искусственный интеллект”; Haugeland 1985), которая получила название *гомункулярного функционализма*:

Отталкиваясь от определенным образом описанной проблемы, программист ИИ фактически считает компьютер антропоморфным: если компьютер *решил* проблему, программист скажет, что создал компьютер, который может [например] понимать вопросы на русском языке. На первом и высшем уровне конструкции компьютер разбивается на две подсистемы, каждая из которых получает определенным образом описанную задачу. Программист составляет схему блоков оценки, устройств памяти, селекторов, контроллеров и тому подобных элементов. Это многочисленные *гомункулы*... Каждый гомункул, в свою очередь, делится на *меньших* гомункулов, но – что важнее – на *менее умных* гомункулов. Достигнув уровня, на котором гомункулы представляют собой не более чем суммирующие и вычитающие устройства, а их интеллект ограничивается способностью по команде выбирать большее из двух чисел, гомункулы превращаются в функционеров, *которые можно заменить машиной*. [p. 80]

Большой плюс этой стратегии заключается в том, что она опровергает возражение о *бесконечном регрессе*. Гомункулярный функционализм позволяет обойти губительный бесконечный регресс и заменить его *конечным*, который в итоге приводит нас, как мы только что заметили, к операторам, задача коих настолько проста, что их можно заменить машинами. Главное было отказаться от мысли, что всю работу делает центральный оператор, и распределить задачи между небольшими, менее умными агентами, работа которых также распределяется между другими агентами, и так далее.

Таким образом, идя от общего к частному, классический GOF AI позволил нам сделать большой шаг вперед, однако предложенная им бюрократическая система организации вышла хотя и эффективной, но излишне механистичной! Может, нам и показалось, что мы избавились от короля или директора, но у нас все равно осталась армия руководителей среднего звена, подчиняющихся множеству вице-президентов (взаимодействия которых формируют высший уровень системы) и отдающих приказы своим подчиненным, которые, в свою очередь, делегируют задачи нижестоящим конторским работникам, и так далее. Существование этой гиперэффективной организации, где ни один элемент не работает вхолостую, где не происходит искусственного раздувания штатов, а все приказы беспрекословно выполняются, в основном обуславливалось тем фактом, что огромные компьютеры, на которых разрабатывались ранние модели ИИ, по сегодняшним стандартам были медленными и маломощными, а люди хотели как можно быстрее получить результат. Чтобы произвести впечатление на инвестора, нужно было создать такой ИИ, которому для ответа на простой вопрос не нужно думать часами. Он должен был работать четко. Кроме того, на создание тысяч строк кода требуется немало времени, а если вы сумели разбить итоговую цель –

к примеру, научить компьютер отвечать на вопросы о лунных камнях, диагностировать болезни почек или играть в шахматы – на серию решаемых задач и знаете, как создать для них программы, которые затем можно будет интегрировать в рабочую систему, вы получите свой “экспериментальный образец”<sup>[24]</sup> при целесообразно низких временных и финансовых затратах.

Обратите внимание, что компьютеры всегда конструируются таким образом, чтобы потребности и производительность практически не зависели друг от друга. Аппаратное обеспечение распределяет электрический ток щедро и беспристрастно, поэтому ни одна из микросхем не рискует оказаться обделенной. На программном уровне великодушная диспетчерская система распределяет машинные циклы в зависимости от приоритизации задач, и хотя приоритет процессам может присваивать специальный распределительный механизм, в итоге получается организованная очередь, а не беспорядочная борьба за выживание. Как сказал бы Маркс, “от каждого по способностям, каждому по потребностям”. Специалист по теории вычислительных систем Эрик Баум метко назвал эту иерархию системой “*политбюро*”. Возможно, смутное понимание этого факта лежит в основе типичного человеческого представления о том, что компьютер ни к чему не может *проявлять участие*. И проблема не в том, что он сделан из неправильных материалов – разве можно сказать, что кремний не подходит для проявления участия в той степени, в которой подходит углерод? – а в том, что его внутренняя организация не предполагает наличия рисков и возможностей и потому компьютер и *не должен* проявлять участие.

Нейроны не такие. Обычные клетки, из которых состоят наши тела, вероятно, можно сравнить с послушными рабами – наподобие самоотверженных, бесплодных рабочих муравьев, которые живут в колонии, выполняют стереотипные задачи и существуют в относительно неконкурентной (марксистской) среде. Но клетки мозга – как я теперь думаю – ожесточенно конкурируют на рынке. За что? Чего может *желать* нейрон? Энергии и ресурсов, чтобы жить дальше – прямо как его одноклеточные предки-эукариоты и более дальние кузины, бактерии и археи. Нейроны – это своего рода биологические роботы; они явно не наделены сознанием в полной мере. Не забывайте, это эукариоты, подобные грибкам и дрожжевым клеткам. Если отдельные нейроны сознательны, то сознательна и нога спортсмена! Но нейроны, как и их безмозглые одноклеточные кузины, представляют собой весьма компетентных агентов в борьбе за существование – только не между пальцами ваших ног, а в конкурентной среде мозга, где победу одерживают клетки, способные более эффективно устанавливать связи и участвовать в более важных процессах на уровне **виртуальных машин**, где различаются масштабные человеческие стремления и цели. Многие подсистемы нервной системы организованы в качестве *оппонентных процессов*, втянутых в войны между подподсистемами, каждая из которых стремится перетянуть одеяло на себя. (К примеру, наши эмоции можно представить в качестве соперничающих бурь, стремительно вытесняющих друг друга, мешающих друг другу или вступающих в сговор против другой бури.) На мой взгляд, оппонентная динамика эмоций и та роль, которую они играют в контроле над нашим сознанием, определяются *экономным характером* нейробиологии, сдерживающим склонность отдельных нейронов к конкуренции. (Обратите внимание, что нейроны все равно умеют работать в команде в рамках более крупной организации, в отличие от более радикально эгоистичных агентов, раковых клеток. Нобелевский лауреат биолог Франсуа Жакоб, как мне помнится, сказал, что любая клетка мечтает разделиться на две. Нейроны же стремятся сохранить активность и влияние, но не мечтают о размножении.) С такой точки зрения способность животных разумно контролировать свое поведение все равно остается моделируемым процессом – как транзакция на бирже, – но, как сказал нейробиолог Себастьян Сеунг (2007), нейроны эгоистичны и стремятся максимизировать свою прибыль во всех валютах, которые только доступны в мозге. Что же нейроны *покупают*, отдавая дофамин, серотонин и окситоцин? Они приобретают большее влияние в сетях, в которые вовлечены, а вместе с ним и большую безопасность. (Тот факт, что мулы бесплодны, не мешает им добывать себе пропитание, а следовательно, нейроны также могут руководствоваться инстинктами самосохранения, унаследованными от способных к размножению предков.)

Итак, если идти от частного к общему, вдохновляясь нейробиологией, гомункулярный функционализм начинает казаться все более точной моделью работы мозга, поскольку порождаемые им хаотичные и конкурентные “вычислительные архитектуры” выглядят убедительнее с биологической точки зрения: мы начинаем понимать, какие эволюционные процессы ответственны за построение этих архитектур как на стадии зародыша, так и во взрослой жизни. Кроме того, мы можем проследить их происхождение от более простых нервных систем, состоящих из менее квалифицированных гомункулов, которые лишь **вроде как** обладают восприятием, дают друг другу сигналы и имеют память.<sup>[25]</sup>

## 21. Оператор “вроде как”

Зачем нам вообще оператор “вроде как”? Дело в том, что, разбирая – или собирая – комплект все более компетентных уровней, мы должны знать о каждом уровне два факта: что он *собой представляет* и что он *делает*. Чтобы понять, что он *собой представляет*, достаточно описать структурную организацию элементов, из которых он состоит, при условии что элементы, насколько нам известно, функционируют именно так, как положено. Чтобы понять, что он *делает*, нужно проанализировать (когнитивную) функцию, которую он (вроде как) выполняет, причем выполняет достаточно хорошо, чтобы на следующем уровне мы могли сказать, что имеем в своем распоряжении компетентный кирпичик, выполняющий только эту функцию – выполняющий ее вроде как, чего нам вполне достаточно. Это ключ к решению невероятно сложного вопроса о том, как сознание вообще может состоять из материальных механизмов. На заре компьютерной эры этот потенциал разглядел Алан Тьюринг, который больше всех заслуживает признания за изобретение компьютера. Он брал безмозглые фрагменты механизмов, не имеющие никакой искры сознания, и составлял из них более компетентные механизмы, которые, в свою очередь, могли стать фрагментами еще более компетентных механизмов – и так далее, без очевидных ограничений. Оператор “вроде как” когнитивной науки примерно соответствует дарвиновскому градуализму эволюционных процессов (подробнее об этом в части VI). До бактерий существовали вроде как бактерии, до млекопитающих – вроде как млекопитающие, до собак – вроде как собаки и так далее.

Нам не обойтись без дарвиновского градуализма, чтобы объяснить огромную разницу между обезьяной и яблоком, а без градуализма Тьюринга не обойтись, чтобы объяснить огромную разницу между человекоподобным роботом и карманным калькулятором. Обезьяна и яблоко состоят из одних и тех же базовых ингредиентов, по-разному структурированных и используемых в многоуровневом каскаде различных функциональных компетенций. Не существует четкой границы, отделяющей вроде как обезьяну от обезьяны. Как человекоподобный робот, так и карманный калькулятор сделаны из одинаковых базовых, бездумных, бесчувственных кирпичиков Тьюринга, но мы составляем из них более крупные, более компетентные структуры, которые затем становятся элементами еще более компетентных структур на более высоких уровнях, и в конце концов добираемся до таких (вроде как) умных фрагментов, что их можно назвать понимающими. Мы используем **интенциональную установку**, чтобы отслеживать убеждения и желания (или “убеждения” и “желания”, или вроде как убеждения и вроде как желания) (вроде как) рациональных агентов на каждом уровне, начиная с простейших бактерий, через все схемы распознавания, сигнализации, сравнения и запоминания, которые включает в себя мозг животных, от морских звезд до людей-астрономов. Не существует четкой границы, отделяющей истинное понимание, и это верно даже в случае с людьми. Маленький ребенок вроде как понимает собственную фразу “мой папа – врач”, а я вроде как понимаю формулу  $E=mc^2$ . Некоторые философы противятся этому *антиэссенциализму* (см. главу 43): либо вы верите, что снег белый, либо нет; либо вы наделены сознанием, либо нет; ничто нельзя считать аппроксимацией психического феномена; здесь либо все, либо ничего. Для таких мыслителей свойства сознания остаются неразрешимой загадкой, поскольку сознание “совершенно” и совершенно не похоже ни на что из того, что можно найти в сугубо материальных механизмах.

## 22. Чудо-ткань

В превосходной книге об индийской уличной магии “Сеть магии. Чудеса и уловки в Индии” Ли Сигел (1991) пишет:

“Я пишу книгу о магии”, – объясняю я, и меня спрашивают: “О настоящей магии?” Под настоящей магией люди подразумевают чудеса, чудотворные деяния и сверхъестественные силы. “Нет, – отвечаю я, – о фокусах. Не о настоящей магии”. Иными словами, настоящей магией называют магию, которая на самом деле не настоящая, а настоящую магию, которую действительно можно творить, настоящей никто не считает. [р. 425]

“Настоящая магия” – по определению, скажете вы – чудесна. Она нарушает законы природы. Многие люди хотят верить в настоящую магию. Повторяя трюки самопровозглашенных ясновидцев вроде Ури Геллера, фокусник, скептик и охотник за привидениями Джеймс Рэнди показывает, что поразительные эффекты нельзя считать настоящей магией, поскольку на самом деле это лишь фокусы. Но некоторых людей этим не убедить. Много лет назад на обсуждении после представления в Виннипеге один из зрителей обвинил Рэнди в двойном обмане: якобы Рэнди был таким же ясновидцем, как и Геллер, но специально притворялся простым фокусником, чтобы разоблачать двуличность более известного Геллера и забирать себе частичку его славы! Сложно опровергнуть такое обвинение, не объяснив всем присутствующим, в чем именно состоит секрет фокусов, чего Рэнди, уважая традиции фокусников всего мира, делать не спешит. (Первыми секреты фокусов стали раскрывать Пенн и Теллер. Их осудили многие коллеги по цеху, но в итоге они сумели доказать, что традиционное табу можно нарушить, не испортив при этом шоу иллюзионистов.)

Подобное стремление верить в настоящую магию свойственно многим людям, рассуждающим на тему взаимосвязи сознания и мозга. Некоторым, включая немалое количество нейробиологов и физиологов – и философов, – хотя бы на подсознательном уровне привлекательна мысль, что так или иначе динамические свойства нервной ткани могут творить чудеса, то есть каким-то образом сдерживать скрытые силы, неизвестные науке. Возможно, они правы, но не стоит считать так по умолчанию. Правило таково: *никакой чудо-ткани!*

Вот что нам известно практически наверняка: ни в одной компьютерной программе не задействованы необъяснимые с точки зрения физики феномены – ни невообразимые ранее силовые поля, ни таинственные квантовые проделки, ни жизненный порыв. Ни в одном компьютере, безусловно, нет чудо-ткани. Мы точно знаем, как компьютеры выполняют основные задачи и как из этих простых задач складываются все более сложные, и можем объяснить сконструированные компетенции, не оставляя никаких тайн. Хотя виртуозность сегодняшних компьютеров не перестает нас поражать, сами компьютеры – как машины – просты, как открывашки. Они показывают фокусы, но не “настоящую магию”.

Это ценный факт – настолько ценный, что он удостоится подробного анализа в следующем разделе. Его ценность заключается в том, что всякий раз, создавая компьютер, который казался чудесным, мы получаем доказательство, что его можно сконструировать без использования чудо-ткани. Возможно, мозг делает все иначе – возможно, даже с чудо-тканью (возможно, Рэнди – настоящий ясновидец, как и Геллер!), – но у нас нет основания этому верить. Таким образом, компьютеры выступают в качестве разоблачителей мифов, и это прекрасный повод настаивать на создании компьютерных моделей всего, что мы пытаемся понять, будь то ураганы, ипотечные пузыри, ВИЧ или человеческое сознание.

Термин “чудо-ткань” представляет собой инструмент мышления, напоминающий дубинку полицейского: он используется в карательных целях, чтобы не давать людям строить безосновательные теории. Как и дубинкой, им можно злоупотреблять. Он дополняет другой инструмент мышления – **бритву Оккама** и потому обеспечивает некоторый научный консерватизм, которому порой свойственна близорукость. Мой любимый пример этого привел один из отцов современной генетики Уильям Бэтсон. Вот что он сказал не так уж давно, в 1916 г.:

Свойства живых организмов некоторым образом связаны с материальной основой, возможно, в некоторой степени с ядерным хроматином [хромосомами]; и все же непостижимо, чтобы частицы хроматина или любого другого вещества, каким бы сложным оно ни было, обладали бы теми силами, которые должны приписываться нашим факторам или генам.



Предположение, что частицы хроматина, неотличимые друг от друга и практически гомогенные во всех известных опытах, могут в своей материальной природе содержать все свойства жизни, переходит границы даже самого убежденного материализма. [р. 91]

Он просто не мог представить ДНК. Мысль о том, что в каждой человеческой клетке может находиться три миллиарда спаренных оснований в двойной спирали, никак не укладывалась у него в голове. К счастью, другие биологи не разделяли пессимизма Бэтсона и в конце концов открыли, как именно происходит, казалось бы, чудесная передача генетической информации от поколения к поколению, за которую, как выяснилось, отвечают весьма необычные молекулы. Однако на пути к этому открытию они придерживались правила: никакой чудо-ткани. Из генетики они знали, какими компетенциями должен обладать искомый элемент, и сделали своей задачей создание физически возможной модели того, что будет обладать этими компетенциями.

Сегодня перед нами стоит подобная задача. Экспериментальная психология дает нам все более подробный каталог компетенций и уязвимостей сознания – триумфов восприятия и неловкостей заблуждения, скорости усвоения языка и условий переключения внимания, вожеления, страха и радости, – и теперь, как “убежденные материалисты”, мы должны выяснить, как мозг все это делает, не прибегая к постулату о чудо-ткани.

По мере того как растет наше понимание, представления о чудо-ткани меняются. Когда в середине 1980-х появились “коннекционистские” и другие “нейросетевые” модели<sup>[26]</sup>, они продемонстрировали наличие у маленьких пучков нейронов способности к обучению и распознаванию паттернов, о которых всего несколькими годами ранее не могло быть и речи. Мы до сих пор не знаем, как именно мозг использует вычислительные мощности, демонстрируемые этими полуреалистичными моделями – и использует ли он их вообще, – но теперь можно постулировать коннекционистскую компетенцию нейронной сети, которую вы пока не можете объяснить, если открыто заявляете об этом, а компетенция явно не выходит за рамки известного диапазона возможностей. (Может, Рэнди исполняет фокус не точно так же, как Геллер, но мы имеем все основания предположить, что существует вариация метода Рэнди, которая объясняет способности Геллера, а это подсказывает нам, как глубже изучить задействованные процессы.) Главное возражение против существования чудо-ткани заключается в том, что она дает нам не способ решить проблему, а повод сдаться, предположив, что эту загадку разрешить так и не удастся.

## 23. В диспетчерской робота

У роботов нет чудо-ткани (фактически по определению), поэтому они представляют собой стерильную платформу для мысленных экспериментов наподобие этого:

Однажды утром вы просыпаетесь в незнакомой кровати в странной комнате без окон. Две стены этой комнаты сплошь покрыты разноцветными мигающими огоньками, а две другие стены – тысячами кнопок. На огоньках и кнопках есть номера, но нет ярлыков. На тумбочке лежит адресованная вам записка:

Доброе утро! Вас накачали снотворным, похитили и доставили сюда, в ваш новый дом. В холодильнике есть еда, в углу комнаты ванная – с удовлетворением физических потребностей проблем не возникнет. Вы заперты в диспетчерской гигантского робота. Загораясь, каждый огонек дает богатую и актуальную информацию о ситуации: эти огоньки – выходные сигналы сложнейшим образом организованной системы анализаторов нейронной сети, которая работает с данными, получаемыми через видеоглаза робота, снимающими изображение высокой четкости, через снабженные микрофонами уши и через тактильные и обонятельные датчики. Кнопки управляют действиями робота, которые скоординированы между собой и могут быть исполнены в любую секунду.

Робот живет в опасном мире, где сталкивается с большим количеством рисков и возможностей. Его будущее в ваших руках, а потому, само собой, и ваше будущее зависит от того, насколько успешно вы будете вести своего робота по миру. Если он будет уничтожен, электричество в этой комнате отключится, еда в холодильнике закончится и вы умрете. Удачи!

Ситуация не из приятных. Затаив дыхание, вы начинаете экспериментировать и нажимать на кнопки, чтобы проверить, что получится. Вы нажимаете на желтую кнопку под номером 4328 и замечаете, что в этот момент гаснет голубой огонек номер 496. Может, вы заставили робота почесаться или “съесть” что-нибудь, чтобы удовлетворить его метаболические потребности? Вы нажимаете на кнопку номер 4328 второй раз, и зажигаются другие огоньки. Что изменилось в мире? Что это означает? Вам страшно, потому что вам сказали, что эти огоньки передают огромное количество *информации*, но что именно показывает каждый из огоньков и какие именно приказы отдает каждая из кнопок?

Вот бы огоньки и кнопки были подписаны! Если бы все они были снабжены ярлыками на знакомом вам языке, вероятно, вы смогли бы решить проблему. Или было бы в комнате хоть одно окно, из которого можно было бы выплянуть, чтобы проверить, что случилось, когда вы нажали на кнопку! Есть ли у робота руки и ноги? Будь в комнате окно, вы могли бы попытаться сопоставить события во внешнем мире с миганием огоньков на стенах. Поскольку окна нет, вам доступно огромное количество информации, которую вы не в состоянии интерпретировать. Вы можете запустить сотни действий робота, но у вас нет надежды узнать, чего добьется этими действиями ваш робот.

Похоже, проблема неразрешима. Каким бы умным и изобретательным вы ни были, вы не сможете узнать, за что отвечают огоньки и кнопки на стенах, имея в своем распоряжении лишь перечисленные данные. Однако если вы не можете решить эту проблему, мы сталкиваемся с парадоксом, поскольку описанная ситуация есть не что иное, как ситуация, в которой существует ваш мозг! Он заперт в комнате без окон – в вашем черепе, – где миллионы входных линий непрерывно передают информацию о внешнем мире и состоянии вашего тела, а миллионы выходных линий стимулируют мускулы сокращаться и расслабляться. Ваш мозг не может открыть в вашем черепе окно и посмотреть, что происходит снаружи и провоцирует сигналы, поступающие в вашу зрительную кору. (Да и зачем вашему мозгу такое окно? В отличие от вас, у него нет глаз, за исключением тех глаз, сигналы от которых он получает и пытается истолковать, и нет воспоминаний о том, как выглядят объекты внешнего мира.)

Возможно, подумаете вы, стоящая перед мозгом задача проще, ведь, когда вы смотрите, к примеру, на утку, рисунок возбуждения – мигающие огоньки – на поверхности зрительной коры действительно принимает форму утки<sup>[27]</sup>! Это имело бы смысл, если бы мы могли предположить, что ваш мозг, как и вы сами, знает, как выглядит утка, но откуда ему это знать?

Как ваш мозг вообще может узнать хоть что-нибудь, если первым делом ему нужно “расшифровать” все сигналы? И как именно расшифровать? Перевести на турецкий? Ярлыки на турецком никак вам не помогут, если только вы не понимаете турецкий. Должен ли ваш мозг понимать язык, чтобы извлекать пользу из входящих данных? Как мы уже видели, мысль о том, что мозг наделен внутренним языком,

который ему не приходится учить – ментализм, или языком мышления (Fodor 1975, 2008), – весьма привлекательна. Кажется, она позволяет сделать шаг в верном направлении, но, пока нет объяснения, как именно работает этот язык и как он сформировался, заявлять о его существовании – все равно что придумывать проблеме новое название, при этом ее не решая. Мы знаем, что мозг *каким-то образом* справляется со своей задачей: он довольно точно определяет, какие команды необходимо отдать в соответствии с обстоятельствами и возможностями, о которых сигнализирует входящий поток данных. И мы знаем, что имеющееся у мозга решение – каким бы оно ни было – не может *точь-в-точь* походить на язык (наподобие русского или турецкого), потому что, в отличие от наших родных языков, оно не усваивается нами в детстве. Какой язык ближе к языку мышления – письменный (воспоминания в таком случае должным образом записываются в архивах мозга) или устный? Сколько “слов” в лексиконе ментализма – тысячи, миллионы или миллиарды? Имеет ли значение порядок слов? Есть ли грамматические правила? Может ли один фрагмент мозга истолковать сообщение другого фрагмента неправильно?

Если гипотеза о языке мышления говорит о наличии в диспетчерской гомункула, который понимает язык (как вы, когда читаете ярлыки огоньков и кнопок внутри гигантского робота), то она просто откладывает задачу выяснить, как аппарат обучения и понимания может состоять из ничего не понимающих фрагментов. А если она не говорит о наличии гомункула, который занимается дешифровкой сообщений, то система, какой бы она ни была, вообще не слишком похожа на язык. Поскольку никто еще не объяснил в деталях, как работает такой язык мышления и как он усваивается в процессе развития и получения опыта, вероятно, лучше не обманываться и не тешить себя надеждой, что мы делаем успехи, когда на самом деле это, вероятно, не так.

Этот насос интуиции показывает нам, что ваш мозг находится в принципиально *иных* условиях, чем вы в диспетчерской робота. Его задача – должно быть – отчасти решается заранее, поскольку некоторые входящие сигналы “сцеплены” с исходящими командами, благодаря чему мозг имеет фору при выявлении и подтверждении других соответствий. Таким образом, подчеркивается широко распространенное убеждение, что при рождении наш мозг представляет собой не “чистый лист” (Pinker 2002), а уже под влиянием естественного отбора сформирован таким образом, чтобы вмещать в себя различные предпочтения, ожидания и ассоциации. А поскольку некоторые соответствия установлены заранее, в ярлыках они не нуждаются.

Прежде чем достичь понимания, необходимо пройти стадию **компетентности без понимания**. Так устроена природа. Бактерии обладают множеством удивительных компетенций, которые им вовсе не нужно понимать; их компетенции прекрасно служат им, но сами они остаются полными невеждами. Деревья обладают компетенциями, которые дают им выгоду, но им не нужно знать зачем. Весьма компетентен и сам процесс естественного отбора – он создает структуры выдающейся оригинальности и эффективности, но ровным счетом ничего не понимает.

Понимание того рода, которым наделены мы, взрослые люди, лишь недавно появилось на эволюционной арене. В нем задействованы структуры, компетенция которых сопровождается и поддерживается минимальным полупониманием, или псевдопониманием, – тем полупониманием, что характерно для рыб и червей. Эти структуры созданы таким образом, чтобы в большинстве случаев вести себя соответствующе, хотя им нет нужды понимать, почему их поведение считается соответствующим.

Альтернативный путь, предполагающий помещение в диспетчерскую полнофункционального Понимателя, обрабатывающего все входящие и исходящие сигналы, гарантированно ведет в тупик. Почему? Потому что, если способность к пониманию необъяснима, фактически вы кладете в основу своей теории чудо-ткань, настоящую магию, а если она все же объяснима – в категориях процессов, действий и сил, которые сами не обладают способностью к пониманию, – вы проехали свой поворот и вернулись к началу с теми из нас, кто пытается объяснить, как понимание вырастает из компетентности.

## **IV.**

### **Интерлюдия о компьютерах**

Перерыв. Как вы, вероятно, заметили, я уже несколько раз упоминал компьютеры, а впереди разговоров о компьютерах будет еще больше. Без сомнения, компьютеры – это самые мощные инструменты мышления, которые имеются в нашем распоряжении, и не только потому, что они справляются со многими скучнейшими интеллектуальными задачами, но и потому, что многие концепции, предложенные специалистами по теории вычислительных систем, тоже можно использовать в качестве инструментов мышления. Сегодня мы все плаваем в море компьютерного жаргона – программное и аппаратное обеспечение, ширина полосы пропускания, гигагерцы, – и многие из вас, безусловно, неплохо понимают значения всех этих модных словечек. Мои студенты всякий раз понимающе кивают, когда я использую эти термины на занятиях, однако я обнаружил, что понимают они далеко не все и порой озадачивают меня, совершенно неправильно толкуя то, что я пытаюсь им объяснить. В связи с этим я хочу научить вас писать программы для простейшего компьютера в мире.

Если вы сейчас не пожалеете времени и сил, чтобы усвоить несколько базовых навыков, впоследствии ваши затраты окупятся сполна, поскольку вы гораздо лучше поймете изложенное далее. (Если же вы в компьютерах собаку съели, мой способ объяснения некоторых тонкостей может пригодиться вам в разговоре с непрофессионалами. Или просто листайте дальше.) Приводимые далее упражнения я тестировал на сотнях студентов, которые боялись компьютеров как огня, и результаты меня радовали: даже те, кто предпочитал заучить наизусть страницу телефонного справочника, только бы не решать никаких головоломок, чувствовали удовольствие и даже некоторое удовлетворение, когда этот простой до идиотизма компьютер справлялся с задачей. Выполнив упражнения, вы узнаете семь секретов силы компьютеров.

## 24. Семь секретов силы компьютеров

Компьютеры наделены силами, которые несколько веков назад показались бы чудесными – “настоящей магией”. При этом, хотя многие компьютерные программы невероятно сложны, все они состоят из шагов, которые в полной мере объясняются очень простыми словами. В работе компьютеров нет места чудесам. Само это отчасти обуславливает ценность компьютеров в качестве инструментов мышления, и потому возможность объяснить – в общих чертах – принцип работы компьютера весьма интересна с философской точки зрения. Никому не помешает на элементарном уровне разобраться, как компьютеры творят свою “магию”. В этой главе фокус будет разоблачен.

Для начала мы представим себе простейший компьютер – *регистровую машину* – и проанализируем его возможности и принцип работы. Затем мы увидим, что *машина Тьюринга* и *машина фон Неймана* (например, ваш ноутбук) ничем не отличаются от регистровых машин, только работают более эффективно. (Все, что под силу вашему ноутбуку, под силу и регистровой машине, но не обольщайтесь – на это у нее могут уйти столетия.) После этого мы разберем, как “архитектуры” других компьютеров могут повысить скорость работы и производительность нашей базовой регистровой машины. Само собой, интереснее и полезнее всего будет изучить архитектуру человеческого мозга.

Погодите-ка. Неужели я только что сказал, что ваш мозг – это просто гигантский компьютер? Нет, пока я такого не сказал. Я говорю, что *если* ваш мозг представляет собой гигантский компьютер, то *найдется способ* полностью объяснить его работу, не оставив никаких тайн, но только если мы сумеем обнаружить этот способ. Мы обратимся к *обратному проектированию*: будем изучать сложную систему, чтобы выяснить, каким образом она функционирует. Обратное проектирование позволяет нам понять, как сердце качает кровь и как легкие поглощают кислород и отдают углекислый газ. Нейробиология делает попытку применить обратное проектирование к мозгу. Мы знаем, *зачем* нам нужен мозг – чтобы предвидеть, направлять нас, запоминать и усваивать информацию, – но теперь нам нужно понять, каким образом он это делает.

Споры на эту тему не утихают. Писатель Том Вулф (2000) заострил внимание на больном месте, которое провоцирует ожесточенные баталии, озаглавив свое эссе “Мне жаль, но ваша душа скончалась”. Если мы намереваемся вступить на эту опасную территорию – не тратя времени на разглагольствования и разоблачения, – нам нужны инструменты получше. Прежде чем добросовестно взяться за вопрос, таит ли и использует ли наш мозг не поддающиеся пониманию или чудесные феномены, недоступные никаким компьютерам, нам нужно понять, на что способны компьютеры и *как именно они функционируют*. Единственный удовлетворительный способ *продемонстрировать*, что наш мозг не тождественен – и не может быть тождественен – компьютеру, заключается в том, чтобы показать либо (1) что ряд его “подвижных частей” участвует в процессах обработки информации, в которых не могут участвовать компьютеры, либо (2) что простые процессы, в которых задействованы его части, не могут быть сконструированы, агрегированы и организованы на компьютерный манер, чтобы симитировать знакомые и любимые нами умения сознания.

Некоторые эксперты – не только философы, но и нейробиологи, психологи, лингвисты и даже физики – утверждают, что “компьютерная метафора” для описания работы человеческого мозга или сознания категорически неверна, а мозгу – что важнее – под силу такие вещи, на которые не способны компьютеры. Обычно, но не всегда, такая критика предполагает весьма наивное представление о том, что такое компьютер или каким он должен быть, и в итоге лишь доказывает очевидную (и не относящуюся к делу) истину, что мозг умеет делать множество вещей, которых не умеет ваш ноутбук (учитывая ограниченное количество его преобразователей и эффекторов, ничтожный объем памяти и низкую скорость работы). Если оценивать эти громкие скептические заявления о возможностях компьютеров *в принципе*, нужно понимать, откуда *в принципе* берется вычислительная мощь, как она используется и как может использоваться.

Блестящую идею создания регистровой машины на заре компьютерной эры предложил логик Хао Ван (1957), между прочим, студент Курта Гёделя и философ. Это изящный инструмент мышления, который вам стоит иметь в своем наборе. Он далеко не так известен, как должен бы<sup>[28]</sup>. Регистровая машина – это идеализированный, воображаемый компьютер (который вполне можно сконструировать), состоящий из некоторого (конечного) числа *регистров* и *блока обработки данных*.

*Регистры* – это ячейки памяти, каждая из которых имеет уникальный адрес (регистр 1, регистр 2, регистр 3 и так далее) и может *содержать* одно целое число (0, 1, 2, 3...). Каждый регистр можно

представить в виде большого ящика, содержащего произвольное количество бобов, от 0 до ..., вне зависимости от размеров ящика. Обычно мы считаем, что в ящике может содержаться *любое* целое число, поэтому ящики, само собой, должны быть бесконечно большими. Для наших целей подойдут и просто очень большие ящики.

*Блок обработки данных* имеет всего три простых компетенции, три “инструкции”, которым он может “следовать” пошагово, выполняя одну зараз. Любая последовательность этих инструкций представляет собой программу, и каждой инструкции присвоен номер, чтобы ее идентифицировать. Инструкции таковы:

*Конец работы.* Машина может остановиться или выключиться.

*Инкремент регистра  $n$*  (прибавить 1 к содержимому регистра  $n$ ; положить один боб в ящик  $n$ ) и переход на следующий шаг, шаг  $m$ .

*Декремент регистра  $n$*  (отнять 1 от содержимого регистра  $n$ ; вынуть один боб из ящика  $n$ ) и переход на следующий шаг, шаг  $m$ .

Инструкция “декремент” работает точно так же, как инструкция “инкремент”, но между ними есть одно принципиально важное различие: что делать, если в регистре  $n$  содержится число 0? Машина не может отнять 1 от этого содержимого (в регистрах не могут содержаться отрицательные числа; боб из пустого ящика не вынуть), поэтому, оказавшись в безвыходном положении, машина должна сделать “переход”. Иными словами, она должна обратиться к другому фрагменту программы, чтобы получить следующую инструкцию. В связи с этим каждая инструкция “декремент” должна определять, к какому фрагменту программы обращаться, если в текущий момент в регистре содержится 0. Таким образом, полное определение инструкции “декремент” звучит так:

*Декремент регистра  $n$*  (отнять 1 от содержимого регистра  $n$ ), если это возможно, и переход на шаг  $m$  **ИЛИ**, если декрементировать регистр  $n$  невозможно, *переход* на шаг  $p$ .

Теперь снабдим все возможности регистровой машины короткими названиями: *Кон*, *Инк* и *Деп* (декремент-или-переход).

На первый взгляд может показаться, что такая простая машина не способна ни на что особенно интересное, ведь она умеет лишь *класть боб в ящик* или *вынимать боб из ящика* (если там есть боб – и переходить к другой инструкции, если его нет). Но на самом деле она может производить такие же вычисления, которые умеет производить любой другой компьютер.

Начнем с простого сложения. Допустим, вы хотите, чтобы регистровая машина прибавила содержимое одного регистра (скажем, регистра 1) к содержимому другого регистра (регистра 2). Таким образом, если в регистре 1 содержится [3], а в регистре 2 содержится [4], мы хотим, чтобы в итоге программа сделала так, чтобы содержимое регистра 2 стало равняться [7], потому что  $3 + 4 = 7$ . Вот программа, которая справится с этой задачей, написанная на простом языке РПА (регистровое программирование на ассемблере):

**программа 1: ADD [1,2]**

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ] |
|-----|------------|---------|---------------|------------------|
| 1.  | <i>Деп</i> | 1       | 2             | 3                |
| 2.  | <i>Инк</i> | 2       | 1             |                  |
| 3.  | <i>Кон</i> |         |               |                  |

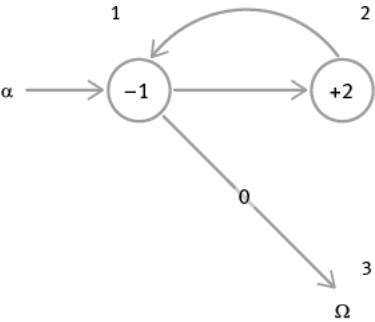
Первые две инструкции образуют простой *цикл*, в рамках которого регистр 1 декрементируется, а регистр 2 инкрементируется снова и снова, *пока регистр 1 не опустеет*. Это “заметит” блок обработки данных, который в результате сделает *переход* на шаг 3, останавливающий программу. Блок обработки

данных не может сказать, каково содержимое регистра, если только это содержимое не 0. Если снова представить ящики с бобами, можно сказать, что блок обработки данных слеп и не видит, что находится в регистре, пока он не опустеет, потому что отсутствие содержимого он может определить на ощупь. Несмотря на то что, в принципе, он не может сказать, каково содержимое регистров, если задать ему программу 1, он *всегда* будет прибавлять содержимое регистра 1 (какое бы число ни содержалось в регистре 1) к содержимому регистра 2 (какое бы число ни содержалось в регистре 2), а затем останавливаться. (Вы понимаете, почему так должно происходить всегда? Разберите несколько примеров, чтобы удостовериться.) Вот любопытный способ на это взглянуть: регистровая машина мастерски умеет складывать числа, не зная, какие именно числа она складывает (а также что такое числа и что такое сложение)!

**упражнение 1**

- а. Сколько шагов потребуется регистровой машине, чтобы сложить  $2 + 5$  и получить 7, выполняя программу 1 (считая Кон отдельным шагом)?
- б. Сколько шагов потребуется машине, чтобы сложить  $5 + 2$ ?  
(Какой из этого можно сделать вывод?)<sup>[29]</sup>

Этот процесс можно изобразить наглядно, построив так называемый *граф потока*. Каждый кружок обозначает инструкцию. Число в кружке обозначает *адрес* регистра, с которым необходимо произвести манипуляции (а не содержимое регистра), “+” обозначает инструкцию *Инк*, а “-” – инструкцию *Деп*. Программа всегда начинается с  $\alpha$ , альфы, и завершается, когда достигает  $\Omega$ , омеги. Стрелки показывают переход к следующей инструкции. Обратите внимание, что каждая инструкция *Деп* имеет две исходящих стрелки, одну для направления, в котором двигаться, если декрементировать содержимое регистра возможно, а другую – если невозможно, потому что содержимое регистра 0 (*переход на ноль*).

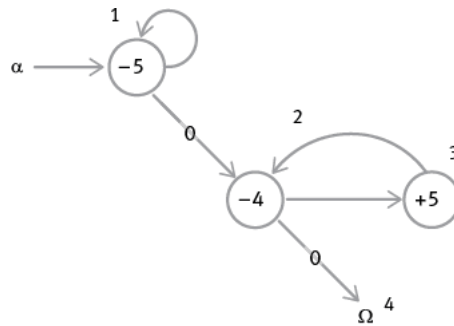


Теперь давайте напишем программу, которая просто перемещает содержимое одного регистра в другой регистр:

**программа 2: MOVE [4,5]**

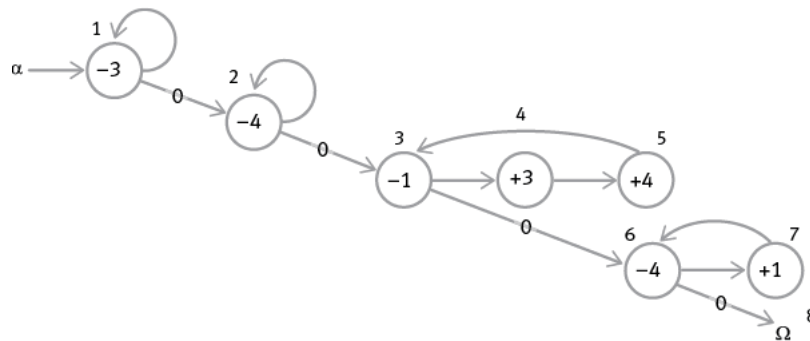
| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ] |
|-----|------------|---------|---------------|------------------|
| 1.  | <i>Деп</i> | 5       | 1             | 2                |
| 2.  | <i>Деп</i> | 4       | 3             | 4                |
| 3.  | <i>Инк</i> | 5       | 2             |                  |
| 4.  | <i>Кон</i> |         |               |                  |

Вот граф потока:



Обратите внимание, что первый цикл этой программы очищает регистр 5, так что, каким бы ни было его содержимое в самом начале, оно никак не повлияет на то, что окажется в регистре 5 ко второму циклу (циклу сложения, в ходе которого содержимое регистра 4 прибавляется к 0 в регистре 5). Этот начальный шаг называется *обнулением* регистра и представляет собой весьма употребительную стандартную операцию. Вы постоянно будете использовать ее, чтобы готовить регистры к использованию.

Третья простая программа копирует содержимое одного регистра в другой регистр, оставляя изначальное содержимое нетронутым. Изучите граф потока, а затем саму программу:



программа 3: *COPY [1,3]*

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ] |
|-----|------------|---------|---------------|------------------|
| 1.  | <i>Деп</i> | 3       | 1             | 2                |
| 2.  | <i>Деп</i> | 4       | 2             | 3                |
| 3.  | <i>Деп</i> | 1       | 4             | 6                |
| 4.  | <i>Инк</i> | 3       | 5             |                  |
| 5.  | <i>Инк</i> | 4       | 3             |                  |
| 6.  | <i>Деп</i> | 4       | 7             | 8                |
| 7.  | <i>Инк</i> | 1       | 6             |                  |
| 8.  | <i>Кон</i> |         |               |                  |

Это явно не самый очевидный способ копирования, поскольку мы осуществляем операцию, сначала *перемещая* содержимое регистра 1 в регистр 3, затем делая копию в регистре 4 и, наконец, перемещая эту копию обратно в регистр 1. Но это работает. Всегда. Каким бы ни было содержимое регистров 1, 3 и 4 в самом начале, когда программа остановится, содержимое регистра 1 останется на месте, а копия этого содержимого – в регистре 3.

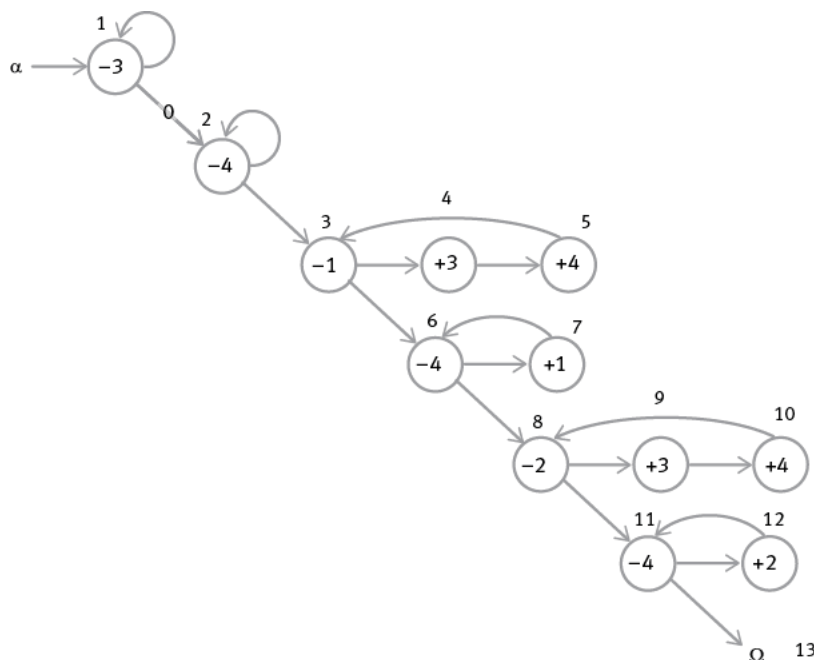


Если принцип работы этой программы вам не очевиден, возьмите несколько чашек, чтобы сделать регистры (подпишите номер каждой чашки, ее адрес), и горстку монеток (или бобов) и “вручную смоделируйте” весь процесс. Положите *по несколько* монеток в каждый из регистров и обратите внимание, сколько именно монеток вы положили в регистр 1 и регистр 3. Если вы будете в точности следовать программе, когда вы закончите, количество монеток в регистре 1 будет таким же, каким оно было изначально, и такое же количество монеток будет лежать в регистре 3. *Очень важно, чтобы вы усвоили базовый принцип работы регистровой машины* и вам не пришлось больше ломать над ним голову, поскольку мы собираемся использовать этот новый навык в дальнейшем. Выделите несколько минут и *станьте* регистровой машиной (как актер может стать Гамлетом).

Я замечаю, что некоторые мои студенты совершают простую ошибку: им кажется, что при декрементировании регистра монетку, которую они только вынули из регистра  $n$ , нужно положить в какой-нибудь другой регистр. Нет. Декрементированные монетки просто возвращаются в общую кучу, в ваш “бесконечный” запас монеток для использования в этих простых операциях сложения и вычитания.

Научившись *перемещать, копировать и обнулять* содержимое регистров, мы можем улучшить нашу программу сложения. Программа 1 помещает верный ответ на задачу на сложение в регистр 2, но в процессе уничтожает изначальное содержимое регистров 1 и 2. Возможно, нам нужна более сложная программа сложения, которая сохраняет эти значения для последующего использования, помещая ответ в другой регистр. Попробуем прибавить содержимое регистра 1 к содержимому регистра 2, поместить ответ в регистр 3 и оставить содержимое регистров 1 и 2 нетронутым.

Вот граф потока, показывающий, как этого добиться:



Проанализируем циклы, чтобы понять, что делает каждый из них. Сначала мы обнуляем регистр ответа, регистр 3, а затем обнуляем дополнительный регистр (регистр 4), который станет временным хранилищем, или буфером. После этого мы копируем содержимое регистра 1 в регистры 3 и 4 и перемещаем это содержимое обратно из буфера в регистр 1, восстанавливая его (и в процессе обнуляя регистр 4, чтобы снова использовать его в качестве буфера). Затем мы повторяем эту операцию с регистром 2, фактически прибавляя содержимое регистра 2 к содержимому, которое мы уже переместили в регистр 3. Когда программа останавливается, буфер 4 снова оказывается пуст, ответ находится в регистре 3, а два числа, которые мы складывали, – на своих изначальных местах, в регистрах 1 и 2.

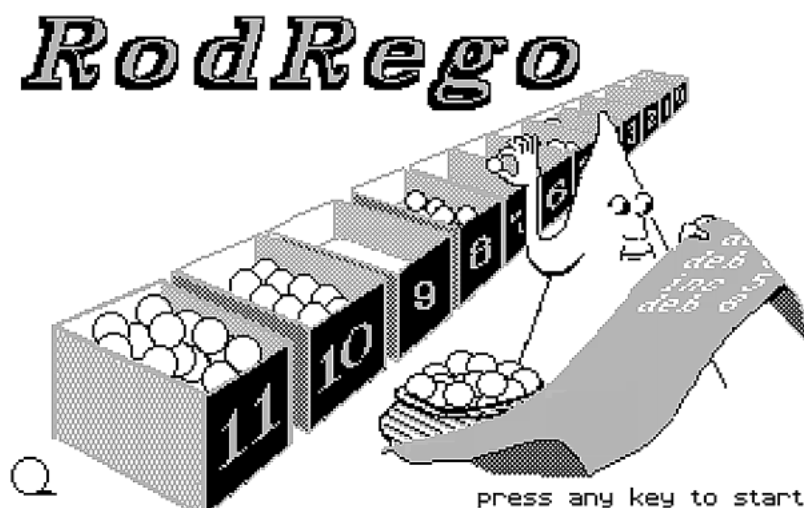
Вот написанная на РПА 13-шаговая программа, которая переводит всю информацию с графа потока на язык, понятный блоку обработки данных:

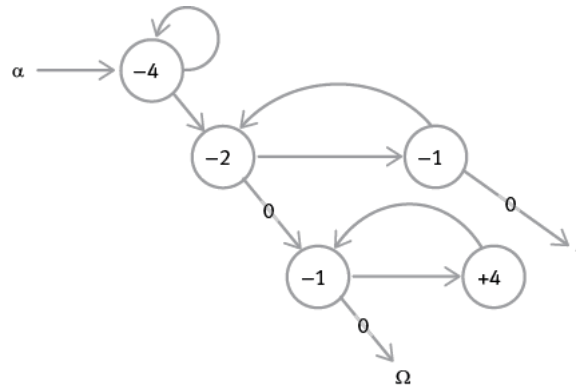
*программа 4: ADD [1,2,3] без разрушения*

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ] |
|-----|------------|---------|---------------|------------------|
| 1.  | <i>Деп</i> | 3       | 1             | 2                |
| 2.  | <i>Деп</i> | 4       | 2             | 3                |
| 3.  | <i>Деп</i> | 1       | 4             | 6                |
| 4.  | <i>Инк</i> | 3       | 5             |                  |
| 5.  | <i>Инк</i> | 4       | 3             |                  |
| 6.  | <i>Деп</i> | 4       | 7             | 8                |
| 7.  | <i>Инк</i> | 1       | 6             |                  |
| 8.  | <i>Деп</i> | 2       | 9             | 11               |
| 9.  | <i>Инк</i> | 3       | 10            |                  |
| 10. | <i>Инк</i> | 4       | 11            |                  |
| 11. | <i>Деп</i> | 4       | 12            | 13               |
| 12. | <i>Инк</i> | 2       | 11            |                  |
| 13. | <i>Кон</i> |         |               |                  |

Я не буду советовать вам вручную смоделировать эту программу, используя чашки и монетки. Жизнь коротка, поэтому, *когда вы усвоите все базовые процессы*, вам можно будет пользоваться вспомогательным устройством *RodRego*, регистровой машиной, которую можно скачать по ссылке <http://sites.tufts.edu/rodrego/>.

Есть версии *RodRego* для *Windows* и для *Mac*. Мы разработали этот инструмент мышления более двадцати лет назад в Мастерской учебных программ, и с тех пор сотни студентов и других заинтересованных людей воспользовались им, чтобы поднатореть в программировании регистровых машин. Вводя программы на РПА, вы можете наблюдать за их исполнением, выбирая режим с цифрами или бобами в регистрах. На той же странице представлена анимированная *PowerPoint*-презентация, в которой показан путь блока обработки данных по графу потока при совершении, к примеру, операции сложения. Эта анимация позволяет увидеть, как инструкции РПА соотносятся с кружками графа потока.

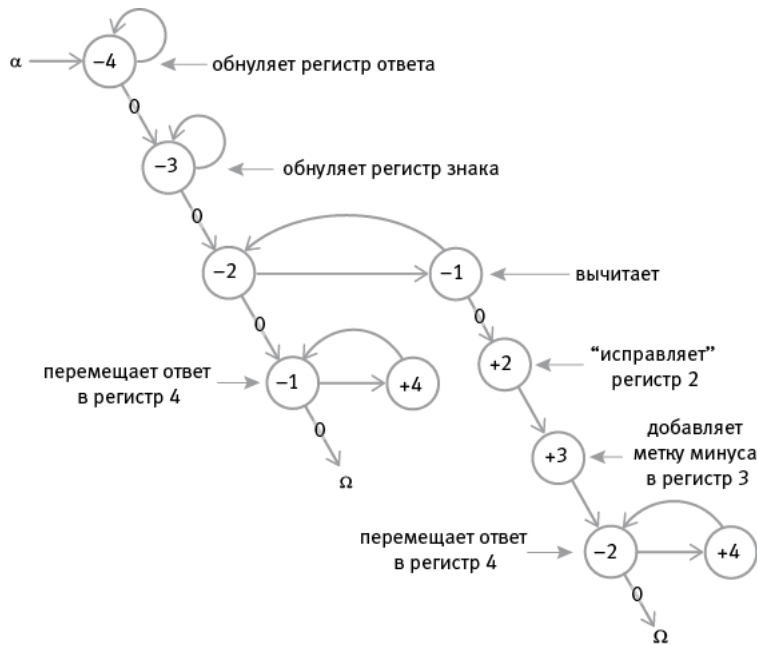




## упражнение 2

б. Что происходит, когда программа пытается вычесть 3 из 3 или 4 из 4?

б. Что происходит, когда программа пытается вычесть 3 из 3 или 4 из 4?



в. Какая возможная ошибка предотвращается обнулением регистра 3 перед попыткой вычитания на шаге 3 вместо шага 4?

Умея складывать и вычитать, мы легко можем разработать программы для умножения и деления. Чтобы умножить  $n$  на  $m$ , нужно просто прибавить  $n$  к самому себе  $m$  раз. Мы можем запрограммировать компьютер сделать это, используя один регистр как *счетчик*, который считает от  $m$  до 0, осуществляя одну операцию декрементирования по завершении каждого цикла сложения.

### упражнение 3

а. Нарисуйте граф потока (и напишите РПА-программу) для умножения содержимого регистра 1 на содержимое регистра 3, поместив ответ в регистр 5.

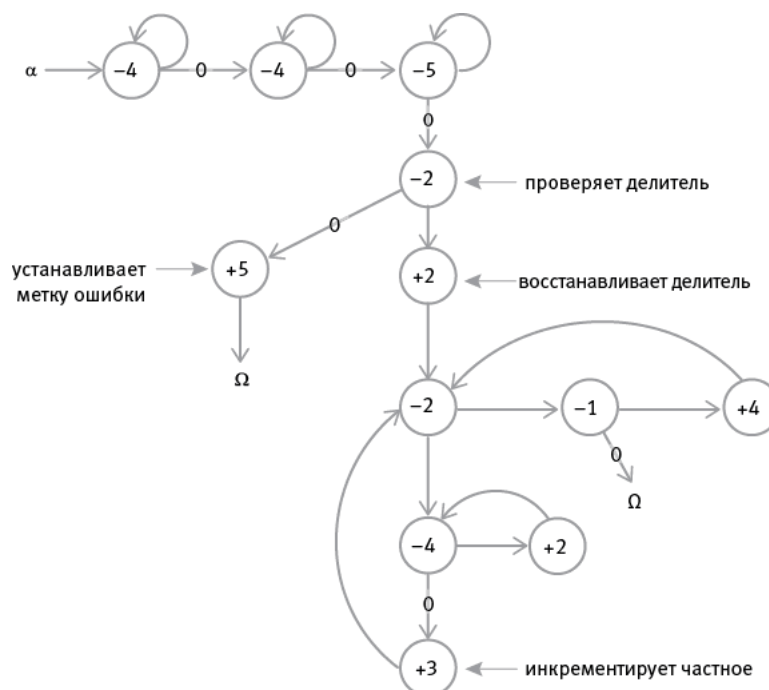
б. (По желанию)<sup>[30]</sup> Используя копирование и перемещение, улучшите программу умножения, созданную в задаче а: когда она закончит работу, изначальное содержимое регистра 1 и регистра 3 восстановится, так что вы сможете легко проверить исходные данные и ответы на правильность по завершении программы.

в. (По желанию) Нарисуйте граф потока и напишите РПА-программу, которая изучает содержимое регистра 1 и регистра 3 (не разрушая их!) и записывает адрес (1 или 3) регистра с большим содержимым в регистр 2 или помещает 2 в регистр 2, если содержимое регистров 1 и 3 равно. (После выполнения этой программы содержимое регистра 1 и регистра 3 должно остаться неизменным, а регистр 2 должен показывать, равно ли их содержимое, а если нет, то в каком из регистров содержимое больше.)

Деление можно выполнять подобным образом, снова и снова вычитая делитель из делимого и считая, сколько раз мы можем провести эту операцию. Остаток – при наличии – можно поместить в специальный регистр для остатка. Но здесь обязательно нужно ввести важную меру предосторожности: на ноль делить нельзя (или можно?), поэтому перед началом деления необходимо провести простую проверку делителя, попробовав его декрементировать. Если его получилось декрементировать, нужно один раз его инкрементировать (чтобы восстановить его исходное значение), а затем провести деление. Если же в регистре ноль, после неудачной попытки декремента нужно поднять тревогу. Можно сделать это, зарезервировав регистр для метки ошибки: 1 в регистре 5 может означать “тревога! Меня только что попросили поделить на ноль!”.

Ниже представлен граф потока для деления содержимого регистра 1 на содержимое регистра 2, помещения ответа в регистр 3 и остатка в регистр 4 и резервирования регистра 5 для “сообщения

об ошибке” (1 означает “меня попросили поделить на ноль”).



Изучите граф и обратите внимание, что ноль в делителе прерывает операцию и выдает сообщение об ошибке. Кроме того, заметьте, что регистр 4 служит двум целям: он не только выполняет роль копии делителя для восстановления делителя после каждой следующей операции вычитания, но и зарезервирован для потенциального остатка. Если регистр 1 обнуляется, прежде чем регистр 4 получает возможность вернуть свое содержимое обратно в регистр 2 для следующей операции вычитания, это содержимое и становится остатком, оказавшемся на своем месте.

секрет 1. Компетентность без понимания. Нечто – например, регистровая машина – может точно выполнять арифметические действия, не понимая, что делает.

Регистровая машина не тождественна сознанию. Она ничего не понимает, но **вроде как** понимает три простые вещи – *Инк*, *Деп* и *Кон* – в том смысле, что всякий раз слепо выполняет три этих “инструкций”. Само собой, это не *настоящие* инструкции, а лишь *вроде как* инструкции. Они кажутся нам инструкциями, и регистровая машина выполняет их так, как если бы они были инструкциями, так что нам более чем удобно называть их инструкциями.

Как вы теперь видите, эффективность регистровой машины объясняется существованием инструкции *Деп*, декремент-или-переход. Только эта инструкция позволяет компьютеру “замечать” (вроде как замечать) что-то в мире и использовать свои наблюдения для выбора следующего шага. Фактически существованием этой команды условного перехода объясняется эффективность всех компьютеров с хранимой в памяти программой, на что Ада Лавлейс обратила внимание еще в девятнадцатом веке, когда написала блестящий комментарий к описанию аналитической машины Чарльза Бэббиджа, которая стала прототипом всех компьютеров<sup>[31]</sup>.

Когда вы наловчитесь собирать программы по частям, это станет для вас обычным делом. Однажды написав арифметические программы, мы можем использовать их снова и снова. Допустим, мы пронумеруем их: *ADD* станет операцией 0, *SUBTRACT* – операцией 1, *MULTIPLY* – операцией 2 и так далее. *COPY* может стать операцией 5, *MOVE* – операцией 6 и т. д. Теперь мы сможем использовать регистр, чтобы хранить в нем инструкцию, обозначая ее присвоенным номером.

**упражнение 4 (по желанию)**

Нарисуйте граф потока и напишите РПА-программу, которая превращает регистровую машину в простой карманный калькулятор, следующим образом:

а. Используйте регистр 2 для операции:

0 = ADD

1 = SUBTRACT

2 = MULTIPLY

3 = DIVIDE

б. Поместите числа, с которыми будут производиться манипуляции, в регистры 1 и 3.

(Таким образом, 3 0 6 будет означать  $3 + 6$ ; 5 1 3 будет означать  $5 - 3$ ; 4 2 5 будет означать  $4 * 5$ ; а 9 3 3 будет означать  $9 \div 3$ .) Затем поместите результаты операции в регистры 4–7, используя регистр 4 для знака (где 0 означает +, а 1 означает –), регистр 5 для численного ответа, регистр 6 для возможного остатка в случае деления, а регистр 7 для сообщения об ошибке ввода (либо требовании делить на ноль, либо неопределенной операции в регистре 2).

Обратите внимание, что в этом примере содержимое регистров (которое всегда представляет собой число) используется для обозначения четырех совершенно разных вещей: числа, арифметической операции, знака числа и метки ошибки.

секрет 2. Что именно обозначает число в регистре, определяет созданная нами программа.

Используя уже созданные структурные элементы, можно конструировать более сложные операции. Если запастись терпением, можно нарисовать граф потока и написать программу для возведения в квадрат – *SQUARE* – числа из регистра 7, или программу для вычисления среднего – *FIND THE AVERAGE* – содержимого регистров 1–20, или программу для разложения на множители – *FACTOR* – содержимого регистра 6 и помещения 1 в регистр 5, если 5 – искомый простой множитель, или программу сравнения – *COMPARE* – содержимого регистра 3 и регистра 4 и помещения большего содержимого в регистр 5, если только – *UNLESS* – оно не ровно в два раза больше, потому что в таком случае необходимо поместить метку в регистр 7. И так далее.

Особенно полезна программа, которая будет осуществлять поиск – *SEARCH* – по содержимому регистров, чтобы проверить, есть ли среди них регистр с конкретным содержимым, и поместить адрес этого регистра в регистр 101. (Как она будет работать? Поместите искомое число – *TARGET* – в регистр 102, а копию искомого числа – в регистр 103. Обнулите регистр 101, а затем, начиная с регистра 1, вычитайте содержимое регистров из содержимого регистра 103 [после инкрементирования регистра 101], ища нулевую разность. Если регистр 1 не дал нулевой разницы, переходите к регистру 2 и так далее. Если в каком-либо из регистров найдется искомое число, остановите программу; адрес этого регистра окажется в регистре 101.) Благодаря примитивной “чувствительности”, заключенной в инструкции *Den*, – ее способности “замечать” нули, когда она пытается декрементировать регистр, – “глаза” регистровой машины можно обратить на нее саму, чтобы она изучала собственные регистры, перемещала между ними содержимое и меняла операции в зависимости от того, что и где она обнаруживает.

секрет 3. Поскольку число в регистре может означать что угодно, регистровая машина, в принципе, может быть сконструирована таким образом, чтобы “замечать” что угодно, или “отличать” любой признак либо черту, которую можно связать с числом – или несколькими числами.

К примеру, черно-белое изображение – любое черно-белое изображение, включая изображение этой страницы, – может быть представлено в виде огромного количества регистров, по одному регистру на пиксель, где 0 означает белую точку, а 1 – черную. Теперь напишите для регистровой машины программу, которая будет искать среди тысяч изображений изображение прямой черной горизонтальной линии на белом фоне. (Не пытайтесь сделать это на самом деле. Жизнь коротка. Просто более или менее живо представьте весь сложный и трудоемкий процесс, который справится с этой задачей.) Сконструировав – в воображении – определитель горизонтальных линий, определитель вертикальных линий и определитель полуокружностей, представьте, как скомпоновать эти фрагменты с несколькими другими полезными распознавателями (возможно, их понадобятся десятки) и написать программу, которая будет находить (заглавную) букву “А” в сотнях разных шрифтов! Программы *оптического распознавания символов* (OCR) – один из относительно недавних триумфов компьютерного программирования – могут сканировать печатные страницы и достаточно точно преобразовывать

их в текстовый файл (в котором каждый буквенный или численный символ представлен числом в кодировке ASCII, так что по тексту можно осуществлять поиск, а также подвергать его другим чудесам текстовой обработки, не используя ничего, кроме арифметики.) Может ли программа OCR *читать*? Не то чтобы. Она не понимает, что видит. Она **вроде как** читает, и это удивительно полезная способность, которую можно добавить в нашу богатую коллекцию структурных элементов.

секрет 4. Поскольку число может означать что угодно, оно может означать инструкцию или адрес.

Числом в регистре можно обозначать инструкцию, например *ADD*, *SUBTRACT*, *MOVE* или *SEARCH*, и адреса (регистров в компьютере), поэтому мы можем хранить всю последовательность инструкций в ряде регистров. Если наша основная программа (программа А) велит машине переходить от регистра к регистру, выполняя содержащиеся в регистре инструкции, тогда в эти регистры можно поместить и вторую программу, программу Б. Когда машина начинает выполнение программы А, первым делом она обращается к регистрам, которые велят ей выполнять программу Б, что машина и делает. Это означает, что можно раз и навсегда поместить программу А в центральный блок обработки данных регистровой машины, зарезервировав для нее ряд регистров (она может стать “встроенной программой”, вшитой в ПЗУ – постоянное запоминающее устройство), а затем использовать программу А, чтобы запускать программы Б, В, Г и так далее, в зависимости от того, какие числа мы помещаем в обычные регистры. Устанавливая программу А на нашу регистровую машину, мы превращаем эту машину в *компьютер с хранимой в памяти программой*.

Программа А наделяет регистровую машину способностью добросовестно выполнять любые инструкции, которые мы помещаем (по номерам) в регистры. Любая подлежащая выполнению программа состоит из упорядоченной последовательности чисел, к которым по порядку обращается программа А, исполняющая инструкции, обозначенные этими числами. Если разработать систему, которая будет унифицировать эти инструкции (к примеру, присваивать им имена одинаковой длины – скажем, двузначные), можно будет считать всю серию чисел, составляющих программу Б, скажем

86, 92, 84, 29, 08, 50, 28, 54, 90, 28, 54, 90

одним большим и длинным числом:

869284290850285490285490

Это число одновременно представляет собой уникальное “имя” программы, программы Б, и саму программу, которая пошагово выполняется программой А. Вот другая программа:

28457029759028752907548927490275424850928428540423

и третья:

8908296472490284952498856743390438503882459802854544254789653985,

но самые интересные программы имеют гораздо более длинные имена, включающие миллионы знаков. Программы, хранимые в памяти вашего ноутбука, например текстовый процессор и браузер, представляют собой именно такие длинные числа, состоящие из многих миллионов (двоичных) знаков. Программа размером 10 мегабит – это последовательность из восьмидесяти миллионов нулей и единиц.

секрет 5. Всем возможным программам в качестве имени может быть присвоено уникальное число, которое затем можно считать списком инструкций для универсальной машины.

Блестящий теоретик и философ Алан Тьюринг вывел эту схему, используя другой простой воображаемый компьютер, который движется по разделенной на ячейки бумажной ленте. Поведение этого компьютера зависит (*ага!* – условный переход) от того, считывает он ноль или единицу из той ячейки, которую обрабатывает в данный момент. Машина Тьюринга умеет только менять бит (стирать 0, записывать 1 – и наоборот) или оставлять бит нетронутым, а затем перемещаться влево или вправо

на одну ячейку *и переходить к следующей инструкции*. Думаю, вы согласитесь, что создание программ сложения, вычитания и других функций для машины Тьюринга при использовании только двоичных чисел 0 и 1 вместо всех натуральных чисел (0, 1, 2, 3, 4, 5 и т. д.) и перемещении лишь на одну ячейку зараз – гораздо более трудоемкий процесс, чем наши упражнения с регистровой машиной, но смысл при этом одинаков. Универсальная машина Тьюринга – это устройство с программой А (если хотите, встроенной), которая позволяет ему “считывать” программу Б с бумажной ленты и затем выполнять ее, используя остальную информацию с ленты в качестве вводных данных для программы Б. Регистровая машина Хао Вана способна выполнить любую программу, которую можно свести к арифметике и условным переходам, и таковы же возможности машины Тьюринга. Обе машины обладают чудесной способностью брать *номер* любой другой программы и выполнять *этот номер*. Вместо того чтобы конструировать сотни разных вычислительных машин, каждая из которых будет прошита для выполнения конкретной сложной задачи, достаточно создать единственную многоцелевую универсальную машину (с предустановленной программой А) и заставить ее исполнять наши запросы, пополняя ее программами – программным обеспечением, – создающими **виртуальные машины**.

Иными словами, универсальная машина Тьюринга представляет собой универсальный имитатор. То же самое можно сказать и о менее известной универсальной регистровой машине. И о вашем ноутбуке. Ваш ноутбук не делает ничего такого, что не могла бы сделать универсальная регистровая машина, и наоборот. Но не обольщайтесь. Никто не говорит, что все машины работают с одинаковой скоростью. Мы увидели, как ужасно медленно наша регистровая машина выполняет такую кропотливую операцию, как деление, которое она – мама дорогая! – осуществляет путем последовательного вычитания. Неужели нет способа ускорить процесс? Конечно есть. Фактически история компьютеров со времен Тьюринга до сегодняшнего дня – это как раз история поиска все более быстрых способов делать то, что делает регистровая машина, и ничего более.

секрет 6. Все усовершенствования компьютеров с момента изобретения воображаемой машины Тьюринга, движущейся по бумажной ленте, просто ускоряют их работу.

К примеру, Джон фон Нейман создал архитектуру для первого серьезного работающего компьютера, и чтобы ускорить его работу, он расширил окно головки записи-чтения машины Тьюринга, чтобы она читала зараз не один бит, а много. Многие ранние компьютеры читают 8-битные, 12-битные и даже 16-битные “слова”. Сегодня часто используются 32-битные слова. Это все еще узкое место – узкое место фон Неймана, – но оно в 32 раза шире узкого места машины Тьюринга! Прибегнув к некоторому упрощению, можно сказать, что слова по очереди копируются – *COPY* – из памяти в особый регистр (регистр инструкции), где эта инструкция читается – *READ* – и исполняется. Как правило, слово состоит из двух частей, кода операции (например, *ADD*, *MULTIPLY*, *MOVE*, *COMPARE*, *JUMP-IF-ZERO*) и адреса, который говорит компьютеру, к какому регистру обращаться за содержимым для проведения операции. Таким образом, **10101110 11101010101** может говорить компьютеру выполнять операцию **10101110** на содержимом регистра 11101010101, всегда помещая ответ в специальный регистр, называемый аккумулятором. Существенное различие между регистровой машиной и машиной фон Неймана заключается в том, что регистровая машина может осуществлять операции на любом регистре (само собой, только операции *Инк* и *Деп*), а машина фон Неймана выполняет всю арифметическую работу в аккумуляторе и просто копирует и перемещает – *COPY* и *MOVE* – содержимое в регистры памяти (или накапливает его – *STORE* – в этих регистрах). Все эти дополнительные перемещения и копирования окупаются возможностью осуществлять большое количество аппаратно-реализованных базовых операций. Иначе говоря, существует одна электронная схема для операции *ADD*, другая – для операции *SUBTRACT*, третья – для операции *JUMP-IF-ZERO* и так далее. Код операции напоминает код города в телефонной системе или индекс в почтовом адресе: он отправляет то, над чем работает, в верное место для исполнения. Так программное обеспечение встречается с аппаратным.

Сколько примитивных операций содержится сегодня в настоящих компьютерах? Сотни и даже тысячи. Кроме того, как в старые добрые времена, компьютер может быть наделен сокращенным набором команд (работать на архитектуре *RISC*): такой компьютер умеет выполнять лишь несколько десятков примитивных операций, но при этом работает молниеносно. (Если инструкции *Инк* и *Деп* можно исполнять в миллион раз быстрее, чем аппаратно-реализованную операцию *ADD*, будет целесообразно составить программу *ADD*, используя *Инк* и *Деп*, как мы делали ранее, и для всех операций сложения, предполагающих менее миллиона шагов, мы останемся в выигрыше.)



Сколько регистров содержится сегодня в настоящих компьютерах? Миллионы и даже миллиарды (но все они конечны, так что огромные числа должны распределяться по большому числу регистров). Байт состоит из 8 бит. Имея на своем компьютере 64 мегабайта *RAM* (оперативной памяти), вы имеете шестнадцать миллионов 32-битных регистров – или их эквивалент. Мы знаем, что содержимое регистров может обозначать не только положительные целые числа. *Действительные* числа (например,  $\pi$ ,  $\sqrt{2}$  или  $\frac{1}{3}$ ) хранятся в форме чисел “с плавающей запятой”. Эта форма представления разбивает число на две части, мантиссу и порядок, как в экспоненциальной записи ( $1,495 \times 10^{41}$ ), что позволяет компьютеру производить арифметические действия, чтобы работать с числами (в приближенном представлении), которые натуральными не являются. Операции с плавающей запятой – это обычные арифметические операции (в частности, умножение и деление) над числами с плавающей запятой, и самый быстрый суперкомпьютер, который можно было купить двадцать лет назад (когда я написал первую версию этой главы), имел производительность более 4 мегафлопс (от англ. *floating point operations per second*), то есть мог выполнять более 4 миллионов операций с плавающей запятой в секунду.

Если вам этого мало, можно параллельно подключить друг к другу большое количество таких машин, чтобы все они работали одновременно, а не последовательно, не ожидая в очереди результатов для их последующей обработки. Функционал параллельного компьютера не отличается от функционала последовательного компьютера, но работает он быстрее. Фактически большинство параллельных машин, изучавшихся в последние двадцать лет, было **виртуальными машинами**, смоделированными на стандартных (непараллельных) машинах фон Неймана. Было разработано параллельное аппаратное обеспечение особого назначения, а конструкторы вычислительных машин изучают стоимость и перспективы расширения узкого места фон Неймана и ускорения проходящего сквозь него трафика всевозможными способами – с использованием сопроцессоров, кэш-памяти и различных других подходов. Сегодня японский компьютер *Fujitsu K* имеет производительность 10,51 петафлопс, то есть более десяти тысяч *триллионов* операций с плавающей запятой в секунду.

*Возможно*, этого почти достаточно, чтобы смоделировать вычислительную активность вашего мозга в реальном времени. Ваш мозг – образцовый параллельный процессор, имеющий около сотни миллиардов нейронов, каждый из которых представляет собой сложного маленького агента, имеющего конкретную задачу. Полоса пропускания зрительного “нерва”, по которому визуальная информация передается из глаза в мозг, составляет несколько миллионов каналов (нейронов). Но нейроны работают гораздо медленнее компьютерных микросхем. Нейрон может изменить состояние и отправить сигнал (фактически свою вариацию инструкции *Инк* или *Деп*) за несколько миллисекунд – тысячных, а не миллионных и не миллиардных долей секунды. Компьютеры перемещают биты со скоростью, близкой к скорости света, поэтому чем меньше компьютер, тем быстрее он работает: свет преодолевает 30 см примерно за одну миллиардную долю секунды, поэтому если вы хотите, чтобы два процесса коммуницировали быстрее, они должны протекать ближе друг к другу.

#### секрет 7. Больше секретов нет!

Вероятно, самое удивительное свойство компьютеров заключается в том, что они состояются – простыми шагами – из элементов (операций), которые также предельно просты, а потому им просто негде хранить секреты. В них нет ни эктоплазмы, ни “морфических резонансов”, ни невидимых силовых полей, ни неизвестных прежде законов физики, ни **чудо-ткани**. Вы *знаете*, что если вы сумели написать компьютерную программу, которая моделирует какой-либо феномен, то эта модель не задействует ничего, кроме различных конфигураций арифметических операций.

Что насчет квантовых компьютеров, которые сейчас на пике популярности? Разве квантовые компьютеры умеют что-то такое, что не умеют обычные? И да и нет. Они умеют одновременно решать множество задач и вычислять множество значений, благодаря “квантовой суперпозиции”, причудливому и необычному свойству, которое гласит, что ненаблюдаемая сущность может одновременно пребывать во “всех возможных” состояниях, пока наблюдение не приведет к “коллапсу волновой функции”. (Подробнее об этом почитайте в своей любимой научно-популярной книге по физике или на познавательном сайте.) По сути, квантовый компьютер не что иное, как последняя и весьма впечатляющая инновация – можно сказать, квантовый скачок – в скорости обработки данных. Машина Тьюринга, движущаяся по бумажной ленте, и регистровая машина, неумоимо инкрементирующая и декрементирующая отдельные регистры, крайне ограничены в том, что они могут сделать за короткое время – минуты, часы или дни. Суперкомпьютер вроде *Fujitsu K* может делать все то же самое в триллионы раз быстрее, но для решения некоторых задач, особенно из области криптографии, этой

скорости все равно недостаточно. Именно в этой сфере может пригодиться сверхбыстрый квантовый компьютер — если, конечно, люди сумеют решить невероятно сложные инженерные проблемы, возникающие при попытке создания стабильного и практичного квантового компьютера. Возможно, создать его так и не удастся, и тогда нам придется довольствоваться какими-то квадриллионами флопс.

## 25. Виртуальные машины

Реальные машины состоят из материальных подвижных частей и, как правило, называются в соответствии с задачами, для выполнения которых они сконструированы. Дизайн газонокосилок, открывашек и кофемолок бывает разным. Порой в основе их работы даже лежат разные физические принципы, но кое-что остается неизменным: машины, которые называются одинаково, *выполняют более или менее одинаковую работу*. Возможно, некоторые из них выполняют ее лучше других, но все зависит от предпочтений пользователя. Домовладелец может выбрать медленную газонокосилку, которая работает тише; владелец кафе может выбрать кофемолку, которая имеет больше степеней помола, пускай пользоваться ею и сложнее. Некоторые машины универсальны: надев другую насадку, можно превратить дрель в пилу или шлифовальную машинку. Компьютеры тоже универсальны, но могут делать не десяток, а целую тучу разных вещей. Вместо того чтобы надевать для выполнения каждой задачи новую насадку, вы открываете новую программу – очень длинную последовательность нулей и единиц, – которая ставит все необходимые внутренние переключатели в нужное положение, чтобы выполнить задачу. Каждая схема настройки – это уникальная машина, уникальная *виртуальная* машина, машина “из инструкций”, а не из подшипников, шестеренок, блоков и проводов. В компьютерах инструкции могут заменять блоки и шестеренки, потому что вместо хлебного теста, целлюлозы и стальных болванок компьютеры обрабатывают информацию, а информацию всегда можно перевести в двоичный код, состоящий из нулей и единиц, единственный код, который компьютер может – и имеет необходимость – “читать”. Микросхемы на кремниевых подложках манипулируют триллионами этих нулей и единиц, временно открывая и закрывая шлюзы, переводя потоки информации к той или иной микросхеме и таким образом контролируя, что с ней происходит. Миллионы крошечных элементов аппаратного обеспечения, которые могут принимать одно из двух состояний – 0 или 1, – это единственные “подвижные части” машины, и от настроек тысяч или миллионов этих крошечных элементов зависит то, какой именно машиной является компьютер в конкретный момент времени.

Виртуальной называется машина, которая возникает, когда определенный набор инструкций (а точнее, диспозиций) применяется к реальной машине, обладающей высокой *пластичностью*, то есть машине, взаимодействующие элементы которой могут пребывать в большом количестве разных состояний. Поскольку виртуальная машина работает с информацией, она может *выполнять ту же задачу*, что и компьютер, в котором роль “подвижных частей” играют изменения состояния аппаратного обеспечения, производя все эти изменения в *представлениях* этих подвижных частей. Вы можете выполнять деление столбиком, записывая процесс карандашом на бумаге, но деление столбиком дается вам особенно хорошо, так что вы можете производить его “в уме”, просто представляя – или воображая – цифры на листе или классной доске. Поскольку речь идет об информации, результат всегда один: вы получаете ответ. Напротив, если вы представите себе бутерброд с ветчиной, почувствовав голод, настоящего бутерброда это не заменит. Компьютеры так хорошо справляются с информационной работой “в уме” (представляя машину, которая выполняет работу), что практически невозможно сказать, с какой машиной вы при этом имеете дело – с “особой”, “аппаратно-реализованной”, ориентированной на выполнение конкретной задачи по обработке информации машиной или же с виртуальной машиной, работающей на многоцелевой микросхеме. К примеру, подавляющее большинство крошечных, дешевых компьютерных микросхем, которые используются сегодня в лифтах, кондиционерах, машинах, холодильниках и пультах дистанционного управления, на самом деле представляют собой многоцелевые компьютеры, способные выполнять вариации всех программ, установленных на вашем ноутбуке, но обреченных всю жизнь выполнять единственную относительно простую программу (программу регулировки зажигания, цикла разморозки и т. д.), “вшитую” в ПЗУ, и эта программа ограничивает все их выдающиеся способности одним-двумя трюками. Это дешевле изготовления микросхем специального назначения, разработанных для выполнения только этих простых задач.

Концепция *виртуальной машины* – один из лучших стимуляторов воображения, которые пришли к нам из компьютерной науки. Она успела доказать свою состоятельность в сфере информатики, и теперь ее пора применять в других сферах. Я использую этот термин в более широком смысле (и в свое время я объясню почему), поэтому не лишним будет узнать, каково его изначальное – как скажут некоторые, истинное – значение. Термин предложили специалисты по теории вычислительных систем Джеральд Попек и Роберт Голдберг (1974). Изначально он означал “эффективный, изолированный дубликат реальной машины” – дубликат, состоящий из... инструкций. Реальная машина – назовем ее А – это

настоящий аппарат, сконструированный из кремниевых чипов, проводов и тому подобного, а виртуальная машина – это компьютерная программа (выполняющаяся на *другой* реальной машине, машине Б), которая полностью имитирует аппаратное обеспечение машины А: она может работать немного медленнее, потому что ей приходится составлять все базовые операции машины А из базовых операций, доступных на ее собственном аппаратном обеспечении, машине Б, но при этом выполняет те же самые программы. Программа, написанная для выполнения на аппаратном обеспечении А, должна без проблем выполняться и на аппаратном обеспечении Б, если на аппаратном обеспечении Б работает виртуальная машина, имитирующая машину А.

Этот фокус удивительно полезен – и не только из-за явной экономии, которую он предполагает: скажем, у вас нет компьютера на *Mac OS*, но есть дорогое программное обеспечение, работающее только на *Mac OS*. В таком случае вы можете написать виртуальную машину (ВМ), имитирующую *Mac OS* на вашем компьютере на *Windows*, и тогда программы для *Mac OS* будут работать на вашем компьютере при запущенной ВМ *Mac OS*. Ваш компьютер на *Windows* будет “притворяться” компьютером на *Mac OS*, но программы ничего не поймут! Представьте человека, который сломал руку. Ему наложили гипс, который сильно ограничивает подвижность руки, а его вес и форма также требуют корректировки остальных движений тела. Теперь представьте мима (скажем, Марселя Марсо), который изображает человека с загипсованной рукой. Если мим хорош в своем деле, движения его тела будут ограничиваться ровно так же: у него на руке виртуальный гипс – и он “практически виден”. Компьютер на *Windows*, имитирующий компьютер на *Mac OS* с применением ВМ *Mac OS*, должен быть неотличим – для программного обеспечения, работающего на нем, и для стороннего наблюдателя – от настоящего компьютера на *Mac OS*.

В реальности все обычно наоборот. Хотя и есть сконструированные ВМ *Mac OS*, работающие на *Windows*, насколько мне известно, это скорее баловство, чем настоящее, практичное программное обеспечение. Напротив, для *Mac OS* созданы надежные, удобные в использовании ВМ *Windows*, которые позволяют владельцам компьютеров на *Mac OS* запускать любые программы, написанные для *Windows*. Большинство программ сегодня создается без привязки к конкретному *аппаратному обеспечению*, но с привязкой к конкретным *операционным системам* (которые, в свою очередь, работают на разном аппаратном обеспечении). Первая причина расширить определение виртуальной машины – необходимость включить в него виртуальные имитации операционных систем. Операционная система тоже представляет собой своеобразную виртуальную машину, позволяя немного разному аппаратному обеспечению выполнять одни и те же программы, но операционная система – это *исключительно* программное обеспечение; она не имитирует настоящее аппаратное обеспечение, но создает – фактически задавая конкретные условия – воображаемую машину, которая подчиняется определенным правилам, принимает определенные данные и так далее.

Другая причина расширить определение заключается в том, что одной из самых популярных и широко распространенных виртуальных машин сегодня является виртуальная машина *Java*, или JVM, которая, подобно операционной системе, не имитирует никакую аппаратно-реализованную машину, а существует *только* в форме программно-реализованной машины. Именно изобретению *Java* интернет в основном обязан своей универсальностью: *Java* позволяет скачивать с сайтов маленькие программы – *Java*-апплеты, – которые дают возможность разгадывать кроссворды, играть в sudoku, изучать карты, увеличивать фотографии, играть в игры с людьми с другого конца света, а также решать множество “серьезных” вычислительных задач. Создавая программы на языке программирования *Java*, веб-дизайнер не должен знать, кто именно придет к нему на сайт – пользователи *Mac OS*, *Windows* (или *Linux*), – поскольку *Java*-апплет всегда запускается на JVM, сконструированной специально для работы на компьютере на *Mac OS*, *Windows* или *Linux*. Соответствующая JVM автоматически скачивается и за несколько секунд устанавливается на ваш компьютер, а затем *Java*-апплет, как по волшебству, выполняется на этой JVM. (Возможно, вы замечаете, как на ваш компьютер скачиваются обновления *Java*, а возможно, и не замечаете этого! В идеале вы можете забыть о том, какая JVM установлена у вас на компьютере, и ожидать, что каждый посещаемый вами сайт либо будет использовать *Java*-апплеты, которые уже работают на вашей JVM, либо соответствующим образом обновит *Java* для корректной работы.)

Таким образом, в соответствии с моим расширенным определением термина, виртуальной машиной можно считать практически любую компьютерную программу, поскольку она представляет собой программное обеспечение – систематический список инструкций, – при запуске превращающее универсальный компьютер в машину конкретного назначения, *которую можно было бы сконструировать и подключить в качестве аппаратного обеспечения*. Алан Тьюринг сделал блестящий вклад в науку –

и оказал огромную услугу человеческой цивилизации второй половины двадцатого века, – предложив идею “универсального” компьютера (сегодня мы называем его универсальной машиной Тьюринга), который можно *превратить* в любой другой компьютер, созданный для конкретной цели, просто установив и запустив соответствующую программу! (Если вы решили пропустить главу 24, то имейте в виду: в ней об этом говорится подробнее.) Не нужно создавать все возможные конфигурации аппаратного обеспечения – хватит и одной, ведь все остальное сделает программное обеспечение. Со времен Тьюринга нам известно, что можно взять сложную штуковину – аппаратное обеспечение, – обладающую большой пластичностью – регулируемые ячейками “памяти” или регистрами, – и поместить в эти ячейки памяти набор инструкций, при выполнении которых эта штуковина превратится в любой компьютер, какой нам под силу ясно вообразить.

Машина Тьюринга – или ноутбук – выполняет по одной инструкции зараз и переходит к следующей инструкции, но “параллельные” компьютеры могут выполнять много (миллионов) инструкций одновременно. Регистр – это любой фрагмент аппаратного обеспечения, который может пребывать в том или ином состоянии (например, в состоянии нуля или единицы в случае с компьютерными битами, но состояний может быть и больше двух), пока не получит команду изменить состояние. Любая система регистров, которая может производить определенные элементарные операции на основании этих состояний (к примеру, менять состояние регистра или использовать состояние регистра, чтобы определить, какую операцию выполнять дальше), может и настраивать свои регистры таким образом, чтобы “вычислить функцию” или “исполнить программу”. Получается, что на любом подобном аппаратном обеспечении может работать виртуальная машина, созданная для использования этих простых шагов. Этот фокус можно проворачивать не один, а много раз, устанавливая виртуальные машины на виртуальные машины на виртуальных машинах... на аппаратном обеспечении.

Рассмотрим шахматную программу, написанную на высокоуровневом языке программирования *Common Lisp*, запущенную на операционной системе *Windows 7*, запущенной на компьютере. Этот компьютер притворяется машиной *Windows*, которая притворяется машиной *Lisp*, которая притворяется машиной для игры в шахматы. Если рассмотреть детали программы на высшем уровне, они будут более или менее понятны подкованному в компьютерах и шахматах наблюдателю (“*Aza!* Эта подпрограмма генерирует все возможные ответы на перемещение слона, а затем запускает подпрограмму оценки, которая...”). Настоящий программный код той же программы – последовательность нулей и единиц, которая помещается в регистр инструкций аппаратного обеспечения, – напротив, способен свести любого с ума, поэтому мы благоразумно сосредотачиваем внимание на высших уровнях. На каждом уровне мы видим лес за деревьями, поскольку детали низших уровней весьма кстати скрыты от наших глаз. Параллель между каскадом виртуальных машин на компьютере и **каскадом гомункулов** в гомункулярнофункционалистском представлении о сознании не просто совпадение. Именно необычайный успех виртуальных машин, которые помогли нам понять, как решать немыслимые ранее задачи (бронирование авиабилетов, игру в шахматы, прогнозирование погоды, письмо под диктовку и т. д.), дает повод надеяться, что мы сумеем повернуть подобный – всего лишь подобный – фокус, когда осуществим обратное проектирование мозга.

В таком случае, возможно, сходства мозга людей, говорящих по-французски, несмотря на все наблюдаемые анатомические различия, лучше всего объяснить *на уровне виртуальной машины*: все франкоговорящие люди имеют в голове ту или иную версию ФВМ, французской виртуальной машины, системы взаимосвязанных диспозиций, или микропривычек, каким-то образом хранящихся в миллиардах регистров мозга. Мозг русскоговорящих людей будет отличаться наличием подобной системы надежных шаблонов, РВМ. Если сказать франкоговорящему человеку: *Donnez-moi le sel, s'il vous plait*, – ФВМ гарантированно проконтролирует то же самое поведение, которое у русскоговорящего человека можно вызвать посредством введения в РВМ в его голове следующих данных: “Передай мне соль, пожалуйста”. Как же создаются ФВМ и РВМ, которые функционируют в нашем мозге?

Мы пока не знаем, как описать разные уровни активности мозга людей, играющих в шахматы или говорящих по-французски<sup>[32]</sup>. Несомненно, не стоит и надеяться на появление подробной схемы наподобие той, что позволяет программистам создавать свои творения на высшем уровне, имея полную уверенность, что компилятор (программа, которая берет инструкции высшего уровня и преобразует их в код, понятный аппаратному обеспечению) выдаст работающую программу. Но теперь у нас есть прекрасный экспериментальный образец: мы знаем по крайней мере один способ объяснить высокоуровневые способности машины с триллионами подвижных частей, не прибегая к концепции **чудоткани**.

## 26. Алгоритмы

В книге “Опасная идея Дарвина” (1995а) я предложил такой взгляд на великую идею Дарвина:

Жизнь на Земле рождалась на протяжении миллиардов лет посредством ветвления единственного древа – Древа жизни – под воздействием тех или иных алгоритмических процессов.

И что же такое алгоритм? Существует несколько конкурирующих определений этого термина, и мое, пожалуй, самое широкое. Далее следует переработанный фрагмент моей книги.

Дарвин открыл силу *алгоритма*. Алгоритм – это формальный процесс определенного типа, который должен – по идее – выдавать определенный результат всякий раз, когда он “запущен” или инстанцирован. Алгоритмы нам не в новинку. Они не были в новинку и во времена Дарвина. Многие знакомые арифметические операции, например деление в столбик и сведение баланса, на самом деле представляют собой алгоритмы. Алгоритмами также являются процедуры принятия решений при игре в крестики-нолики и расстановке слов по алфавиту. Относительно новыми можно назвать теоретические рассуждения математиков и логиков двадцатого века о природе и силе алгоритмов в целом – именно эти рассуждения позволили нам в ретроспективе оценить открытие Дарвина и привели к рождению компьютера, которое, в свою очередь, привело к еще более глубокому и осмысленному пониманию силы алгоритмов в целом.

Термин “алгоритм” восходит через латынь (*algorismi*) и раннеанглийский (*algorism* и ошибочной производной *algorithm*) к имени персидского математика Аль-Хорезми, в девятом веке написавшего книгу об арифметических операциях, которая в одиннадцатом веке была переведена на латынь либо Аделардом Батским, либо Робертом Честерским. Представление об алгоритме как надежной и в некоторой степени “механической” процедуре существовало веками, но в современной трактовке термин более или менее закрепился в 1930-х гг., когда его использовали в своей новаторской работе Алан Тьюринг, Курт Гёдель и Алонзо Черч. Нам важны три ключевых свойства алгоритмов, причем дать каждому из них определение непросто.

(1) *Нейтральность носителя*. Процедура деления в столбик одинаково хорошо работает при использовании карандаша и ручки, бумаги и пергамента, неоновых огней и конденсационного следа самолета. Система символов может быть любой. Сила процедуры объясняется ее логической структурой, а не особенностями материалов, используемых при инстанцировании, при условии, что эти особенности позволяют в точности следовать пошаговой инструкции.

(2) *Подспудная простота*. Хотя общая схема процедуры может быть блестящей или давать блестящие результаты, каждый из шагов и переходы между шагами предельно просты. Насколько просты? Достаточно просты, чтобы с ними справился прилежный идиот – или нехитрое механическое устройство. Как правило, в учебниках алгоритмы сравниваются с *рецептами*, созданными для кулинаров-новичков. В кулинарной книге для умелых поваров может быть написано: “Припустите рыбу в подходящем вине почти до готовности”, – но алгоритм того же процесса, возможно, начнется так: “Выберите белое вино с маркировкой «сухое»; возьмите штопор и откройте бутылку; налейте вино в кастрюлю до уровня в два с половиной сантиметра от дна; включите конфорку под кастрюлей на полную мощность...” Иначе говоря, алгоритм дотошно разбивает процесс на предельно простые шаги, не требуя от его исполнителя ни принимать сложные решения, ни выносить осторожные суждения, ни применять интуицию.

(3) *Гарантия результатов*. Что бы ни делал алгоритм, он делает это всегда, если в процессе его исполнения не происходит ошибок. Алгоритм – это надежный рецепт.

Вполне очевидно, как эти свойства сделали возможным появление компьютера. *Любая компьютерная программа представляет собой алгоритм*, в конечном счете состоящий из простых шагов, которые могут с поразительной безотказностью исполняться тем или иным простым механизмом. Обычно выбор падает на электронные схемы, но каузальные особенности электронов, шныряющих туда-сюда по кремниевым чипам, никак не влияют на силу компьютеров (хотя и могут влиять на скорость их работы). Те же самые алгоритмы могут исполняться даже быстрее устройствами, которые манипулируют фотонами в оптоволокне, или – гораздо медленнее – командами людей, использующими бумагу и карандаш.

Фактически Дарвин открыл не *один* алгоритм, а целый класс родственных алгоритмов, которые он не мог однозначно отличить друг от друга.

## 27. Автоматизация лифта

Прежде чем закончить интерлюдия о компьютерах, я хочу описать еще одну группу полезных идей об **исходном коде**, **комментариях** и **объектном коде**, которые мы сумеем применить, чтобы лучше понять, как значение может храниться в мозге. Не стоит сразу браться за головоломку – для начала целесообразно подробнейшим образом изучить предельно простой пример и усвоить все понятия. (В сфере искусственного интеллекта эти примеры называются модельными задачами. Прежде чем браться за жуткую реальную задачу, стоит сначала решить модельную.) В связи с этим я предлагаю вам познакомиться с историей – выдуманной ради простоты, но в остальном вполне реалистичной – о том, как лифтеров заменили компьютерные чипы.

В моей молодости были лифтеры – люди, работа которых заключалась в том, чтобы весь день ездить на лифте вверх-вниз и останавливаться на нужных этажах для посадки и высадки пассажиров. Когда-то они крутили любопытную ручку, которую можно было поворачивать по часовой стрелке и против часовой стрелки, чтобы направлять лифт вверх или вниз, и им требовалась определенная сноровка, чтобы останавливать кабину на нужной высоте. При входе в лифт и при выходе из него людям часто приходилось подниматься или спускаться на небольшую ступеньку, и лифтеры всегда их об этом предупреждали. Существовал целый свод правил относительно того, что и когда говорить, на какие этажи подниматься в первую очередь, как открывать двери и так далее. На стажировке они заучивали эти правила и практиковались следовать им, пока это не входило у них в привычку. Сами правила разрабатывались годами, и в процессе в них постоянно вносились небольшие изменения и уточнения. Представим, что этот процесс остановился, когда был создан идеальный свод правил. Он чудесно работал. Любой, кто в точности следовал правилам, становился превосходным лифтером.

Теперь представим, что случилось, когда простая компьютерная программа сумела взять на себя все задачи лифтера. (На самом деле это происходило постепенно, по мере того как появлялись все новые автоматические механизмы, которые забирали у лифтеров все более сложные задачи, но мы представим, будто лифты перешли от человеческого контроля к полностью компьютерному одним махом.)

Допустим, производитель лифтов нанимает команду программистов и вручает им свод правил, которым руководствовались лифтеры: “*Вот* подробное описание необходимых нам функций; напишите компьютерную программу, которая будет следовать правилам из этой книги так же хорошо, как лучшие лифтеры, и мы будем довольны”. Изучая свод правил, программисты составляют список всех нужных действий и условий, при которых их следует и не следует предпринимать. В процессе они могут избавиться от некоторой небрежности свода правил. К примеру, если они встроит в лифт датчики, которые будут обеспечивать остановку лифта на нужной высоте, они смогут избавиться от цикла, требующего от лифтера сказать “Осторожно, ступенька”, но, возможно, оставят простую (записанную на пленку) фразу “Этаж N-й. Будьте осторожны при выходе из кабины”. Затем они пишут набросок программы, используя так называемый псевдокод – неформальный язык, нечто среднее между обычным человеческим языком и более требовательной системой исходного кода. Строка псевдокода может выглядеть примерно так: “если этаж вызова > текущего этажа, то *ASCEND* до этаж вызова = текущему этажу и *STOP*; *OPENDOOR. WAIT...*”.

Как только план написан на псевдокоде и выполняет необходимые условия, псевдокод можно перевести в исходный код, который представляет собой гораздо более строгую и структурированную систему команд, включая определение терминов – переменных, подпрограмм и так далее. Человеку по-прежнему несложно расшифровать исходный код – в конце концов, его пишут люди, – а следовательно, правила и условия свода правил в нем достаточно очевидны, если знать, куда смотреть. Расшифровке исходного кода способствуют две его характеристики: во-первых, по названиям переменных и команд обычно можно понять, за что они отвечают (*callfloor* [этаж вызова], *weightsum* [общая масса], *TELLFLOOR* [объявить номер этажа] и т. д.). Во-вторых, как мы видели в главе 24, программисты могут добавлять к своему исходному коду *комментарии*, заключенные в скобки объяснения, которые говорят читателям исходного кода, что имел в виду программист и что должны делать разные элементы программы. При создании программы целесообразно снабжать свой код комментариями, ведь забыть, зачем нужна конкретная строка кода, очень легко. Эти комментарии окажутся очень полезны, когда вы будете исправлять ошибки программирования.

Исходный код составляется согласно строгому синтаксису, где каждому элементу отводится свое место и не допускаются пунктуационные ошибки, поскольку он загружается в программу-компилятор,



которая берет исходный код и транслирует его в последовательности элементарных операций (*объектный код*), подлежащих исполнению реальной (или виртуальной) машиной. Компилятор не может гадать, что программист имел в виду в той или иной строке исходного кода, поэтому исходный код должен точно говорить компилятору, какие команды исполнять, однако компилятор может выполнять эти задачи множеством разных способов и умеет выбирать наиболее эффективный способ, исходя из обстоятельств. Одни компиляторы работают лучше других: к примеру, если загрузить одну и ту же программу (в исходном коде) в два разных компилятора, объектный код, выданный одним компилятором, может исполняться гораздо быстрее, чем объектный код, выданный другим компилятором. Допустим, вы написали шахматную программу, загрузили ее исходный код в два разных компилятора и запустили две скомпилированных версии играть друг против друга на одном компьютере. Хотя обе версии будут “обдумывать одни и те же мысли в одном и том же порядке” (у них нет другого выбора – у них одинаковый исходный код), возможно, одна из них всегда будет выигрывать, просто потому что она обдумывает эти мысли быстрее, используя меньшее количество базовых машинных циклов, а следовательно, просчитывает за отведенное время большее количество ходов!

Вернемся к нашему лифту. Как только компилятор скомпилировал объектный код, этот код может быть *исполнен* (это исполняемый файл, обычно в расширении .exe) реальной (или виртуальной) машиной. Возможно, придется исправить ряд ошибок (вернуться к исходному коду, поправить его, снова скомпилировать программу и т. д.), но в итоге получится “законченный” продукт. Его можно будет “вшить” в ПЗУ на крошечный чип, содержащий универсальную машину – и неограниченное количество виртуальных машин в придачу к ней, – и установить его в лифт. Установка предполагает подключение всех преобразователей входящих сигналов, включая сигналы кнопок, вмонтированных в пол весов, измеряющих общую массу пассажиров, и других структурных элементов лифта, и привязки исходящих сигналов к исполнительным механизмам (которые управляют двигателями, открывающими и закрывающими двери, поднимающими и опускающими кабину, а также обновляют информацию на дисплеях и проигрывают записи). *Тадам!* Машина заменила настоящего человека – даже не метафорического гомункула. И машина *следует тем же правилам*, что и лифтер. *Неужели?* На самом деле нет. Она **вроде как** следует тем же правилам. Она занимает промежуточное положение между человеком, который запоминает – то есть в буквальном смысле моделирует в своей голове – правила, диктующие его поведение, чтобы снова и снова сверяться с ними, и планетами, которые “подчиняются” уравнениям, изящно описывающим их орбиты. Мы, люди, тоже часто занимаем промежуточное положение, когда усваиваем или доводим до автоматизма следование ряду четких правил, которые впоследствии можем отбросить и даже забыть (“жи” и “ши” пиши через “и”, “ча” и “ща” пиши через “а”). Порой мы также **вроде как** следуем правилам, которые еще не выведены окончательно, например некоторым правилам русской грамматики, по-прежнему сбивающим с толку лингвистов. Скажем, лингвисты сегодня тщетно пытаются написать учебник хорошего разговорного русского языка, в то время как любой десятилетний носитель русского каким-то образом умудряется установить и избавиться от ошибок довольно хорошую версию объектного кода для своей РВМ<sup>[33]</sup>!

Прежде чем пойти дальше, обратите внимание, что комментарии к исходному коду, помогающие программистам отслеживать назначение всех взаимосвязанных элементов программного обеспечения, не имеют аналогов в процессе создания аппаратного обеспечения, прошивки и программного обеспечения нашего мозга. Когда естественный отбор устанавливает в наш мозг различные функциональные структуры, они напоминают лишенный комментариев код: у них есть четкое назначение, но это назначение не объясняется никакими ярлыками. Впрочем, если бы они и были, мозг все равно был бы не в состоянии их понять. (Подробнее об этом в главе 40.) Без комментариев и объяснений остаются и изменения, происходящие в процессе развития и обучения. Подобно лингвистам, мы отчаянно бьемся в попытках провести обратное конструирование всех этих “правил” и “процедур”. Эта задача даже сложнее обратного конструирования объектного кода с целью восстановления исходного кода (не считая комментариев), но теоретически она выполняема.

## Резюме

Веками мы наблюдали многочисленные свидетельства, что мозг – вместилище человеческой души, но до середины двадцатого века никто и представить себе не мог, как это вообще возможно. Было очевидно, что мозг состоит из множества разных органов причудливой формы, дублирующих в левом и правом полушариях. Первые анатомы давали этим органам любопытные имена – гиппокамп (“морской конек”), миндалевидное тело, морщинистая кора, – но за что эти органы отвечали? Они ведь не переваривали пищу и не очищали кровь? Может, мозг предназначался просто для охлаждения крови, выступая своего рода радиатором, как полагал Аристотель? Отдельные части мозга соединялись нервными волокнами – может, они коммуницировали между собой? Декарт предположил, что некоторые нервные волокна подобны привязанным к колокольчику проводам: если потянуть за такой провод, на другом его конце что-то произойдет, но что именно? Звон колокольчиков не приближал понимание мозга как сознания, а других идей ни у кого не было<sup>[34]</sup>.

Затем появился Тьюринг, который, опираясь на традицию, восходящую к Бэббиджу, Паскалю, Лейбницу и другим ученым, предположил, что мозг может состоять из простых элементов, по сути своей механических (как мышеловка, звонок, замок с ключом и синапс), но если эти элементы организованы таким образом, чтобы хитро взаимодействовать друг с другом, то они смогут выполнять кое-какие разумные действия сами, без человеческого вмешательства и без наличия каких-либо духов в машине, чтобы ими управлять. Они смогут вычислять. До того как Тьюринг предложил свою идею, “компьютерами” называли тысячи людей, которых нанимали на работу в промышленность и государственный аппарат, чтобы они рассчитывали таблицы для использования, к примеру, в бизнесе, навигации, артиллерийском и банковском деле. Тьюринг предположил, что мозг подобен (живому) компьютеру – что он обрабатывает информацию, прилежно следуя огромным спискам предельно простых инструкций (вроде *Инк* и *Деп*). Как только первые теоретики когнитивной науки Алан Тьюринг и Джон фон Нейман, основоположник кибернетики Норберт Винер и основоположник теории информации Клод Шеннон, а также многие другие ученые сформулировали идею, она стала казаться очевидной – как люди не замечали этого раньше? Мозг должен получать информацию от органов чувств и обрабатывать ее, производя какие-то расчеты, пока не извлечет крупницы смысла, которые путем дальнейших расчетов будут категоризированы и сохранены для дальнейшего использования при управлении телом, дающим мозгу энергию и защищенное вместилище. Ключевой инновацией концепции Тьюринга был отказ от неудобного элемента, использовавшегося во всех ранних представлениях об обработке информации, а именно от переломных моментов, которые требовали наличия клерка, переводчика или библиотечаря – иными словами, некоторого *узла, распознающего смысл сигналов*. Тьюринг полагал, что, с одной стороны, от этого никуда не деться: разумные процессы всегда будут *требовать* выбора того или иного варианта действий на основании распознавания различий в сигналах. Но он сумел снизить уровень этого понимания до минимума, заявив о существовании условного перехода – бездумного процесса, посредством которого устройство решает (вроде как решает) идти налево, а не направо, потому что оно воспринимает (вроде как воспринимает) 1, а не 0, А, а не В, х, а не Y. Вкупе с арифметикой этого было более чем достаточно. Имея все это в своем распоряжении, можно создавать устройства, характеризующиеся любой степенью способности к распознаванию, устанавливая виртуальные машины на виртуальные машины на виртуальных машинах – если выразаться современным языком. Эта концепция остается привлекательной более пятидесяти лет, однако – как мы уже начали замечать – все не так уж просто. Если мозг – это компьютер, то он не слишком похож на компьютеры, которые мы используем каждый день. Нам стоит помнить о фундаментальных характеристиках компьютеров, чтобы иметь возможность размышлять о более реалистичных в биологическом отношении альтернативах стереотипным, практичным архитектурам.

Цель интерлюдии состояла в том, чтобы прояснить эту концепцию и описать ее достаточно подробно, дав вам возможность *использовать* ее в качестве инструмента мышления, своеобразного костыля для воображения, который поможет вам понять дальнейшие рассуждения. Сначала мы немного подробнее разберем, как значения могут храниться в мозге (и в других машинах), а затем увидим, как такие хитроумные архитектуры создаются одной лишь эволюцией без помощи главного программиста или творца. После этого вы будете в состоянии использовать изученные инструменты мышления, чтобы успешно размышлять о сознании и свободе воли, а коварнее тем я не знаю.

**V.**

**Дополнительные инструменты мышления о значении**

## 28. Пунктик насчет рыжих

Мы уже видели кое-какие проблемы с заманчивой в остальном идеей о том, что вся информация в нашем мозге – наши убеждения, ощущения, воспоминания, установки и так далее – делится на фрагменты-предложения, которые хранятся в архиве, готовые в любой момент вернуться к жизни. Письменность мозга не позволяет просто *установить* в него ложное убеждение, а люди могут разделять какое-либо убеждение (например, о том убийстве в Лондоне), не имея в голове одинаковой формулы на языке мысли. Но что *еще* может хранить информацию в мозге? Мы, люди, учимся “поэтапно”, поэтому должен существовать какой-то способ добавлять в мозг не связанные друг с другом факты, *грубо говоря*, по одному зараз.

Экономисты (и не только) часто замечают, что невозможно сделать что-то одно. “Что-то одно” всегда имеет последствия. Подобным образом сомнительна и идея, что можно *узнать* лишь что-то одно. Но в первом приближении это все же возможно. Ранее из этой книги вы узнали о существовании млекопитающего под названием пуду. Если вы не искали другой информации о пуду, вероятно, вы не можете сказать об этих животных ничего, кроме того, что они выкармливают своих детенышей, имеют позвоночник и встречаются относительно редко (иначе вы бы точно о них слышали). Вполне очевидно, откуда вы это узнали: вы прочитали предложение и поверили ему. Но могут ли животные или маленькие дети, которые еще не освоили язык, узнать *единичный факт* (наподобие факта, выраженного простым предложением) из какого-нибудь любопытного фрагмента опыта? Представление о том, что знание, убеждение или учение *должно* разбиваться на фрагменты-предложения, вероятно, стоит считать иллюзией антропоморфизма. Каждый день мы, люди, сталкиваемся со множеством описательных предложений – как в устной, так и в письменной речи – и таким образом узнаем всевозможные факты (и верим во всякую ложь). Одни факты мы храним в библиотеках и архивах, а другие – только у себя в голове. Мы редко *запоминаем* предложения слово в слово, но, когда мы выжимаем из предложения *суть*, она должна – должна ведь? – храниться на манер предложений, превращаясь в формулу на языке мысли. Если это не так, то какие есть альтернативы?

Допустим, Пэт говорит, что у Майка “пунктик насчет рыжих”. Примерный смысл фразы Пэт заключается в том, что у Майка в голове сложился достаточно неприятный стереотипный образ рыжих, который влияет на отношения Майка с рыжими. И дело не просто в том, что Майк настроен против рыжих, а в том, что у Майка есть своеобразный и конкретный *пунктик* насчет них. И Пэт, возможно, прав – правее, чем он сам думает! Вполне вероятно, у Майка действительно есть пунктик – не *идея*, не *мысль*, не *утверждение*, не *образ* и не любой другой элемент нашего сознательного опыта, а фрагмент субличностного когнитивного аппарата его мозга, который действительно *связан с рыжими*, потому что систематически всплывает всякий раз, когда речь идет о рыжих или об одном рыжем, и корректирует различные параметры когнитивного аппарата Майка, в результате чего вероятность рассмотрения и вынесения лестных суждений о рыжих снижается, а вероятность относительно агрессивного поведения по отношению к рыжим возрастает и так далее. Принцип работы такого *пунктика насчет рыжих* может быть как очень сложным, так и довольно простым. Имеющийся у Майка пунктик насчет рыжих может играть решающую и неоспоримо осмысленную роль, но никакое *выражение* этого смысла в *предложении*, *считающемся истинным*, не станет более, чем мнемоническим ярлыком этой роли. Иными словами, вероятно, нельзя назвать роль этого пунктика до странности специфическим или до странности расплывчатым *убеждением*, что *все рыжие – У...* (где мы заменяем “У” на то слово, которое лучше всего отражает отношение Майка к рыжим). У Майка явно есть сформированное *отношение* к рыжим, но, если пользоваться философским жаргоном, это отношение нельзя назвать *пропозициональным*. Иными словами, как бы мы ни изошрялись в положениях об исключении, классификаторах, операциях возможности и других однозначных механизмах регулирования содержимого, оно не подлежит категоризации в формате

Майк полагает, что: для всех  $x$ , где  $x$  обозначает рыжего, верно...

Философы (и другие теоретики) неоднократно пытались “свести” все когнитивные состояния к информативным состояниям – назовем их убеждениями и желаниями, – которые выражаются подобными формулами. Хотя эта тактика позволяет сделать *грубый набросок* психологии какой-либо личности (по сути, это **интенциональная установка**), не стоит и надеяться сделать этот набросок

сверхточным. Можно сказать, что различные убеждения *внутренне присущи* системе. Это означает, что система (в настоящее время) настроена на функционирование “при допущении”, что все рыжие мира имеют такие-то черты. Когда программисты добавляют к **исходному коду комментариев**, сообщая всем, что эта система полагается на определенный набор допущений, они знают, что нет смысла тратить силы, повышая точность этих допущений, поскольку эти мнемонические ярлыки необходимы нам, наблюдателям, в то время как компьютеру нет нужды **вроде как** читать и **вроде как** понимать их, и даже нам, наблюдателям, комментарии не дают *спецификаций* содержимого, которые можно было бы использовать наподобие того, как химик использует формулы для описания молекул. Интерпретировать какую-либо публичностную структуру мозга из интенциональной установки – все равно что добавлять комментарии к нескольким строкам кода: при надлежащем исполнении интерпретация дает нам пояснительный ярлык, а не *перевод* формул языка мысли, используемых при обработке информации, на русский или любой другой естественный язык. Не понимая этого, некоторые философы создают целые фантастические миры, в которых происходит манипуляция предложениями: как им кажется, вся соль заключается в том, к примеру, выражено ли содержание конкретного психического события разделительным суждением (“Я вижу мальчика-ИЛИ-девочку”) или суждением без логической структуры (“Я вижу ребенка”).

Какова задача этого насоса интуиции? Это просто попытка предположить, что знакомый вопрос, который задается всякий раз, когда кто-то сомневается в существовании языка мысли – “Что еще, если не язык?” – может получить хороший ответ, способный поубавить пыл всех, кто находит это очевидным. Хотелось бы мне представить смелую альтернативную вычислительную архитектуру, которая триумфально продемонстрировала бы работающую альтернативу, но сделать этого я не могу. Пока этого не может никто, но почти никто и не пытается, потому что по-прежнему широко распространено убеждение, что язык мысли – “единственная соломинка, которая держится на плаву”, как кто-то выразился много лет назад. Однако не забывайте, что ни один специалист по когнитивной науке не сумел предложить и работающую модель языка мысли – и даже не особенно старался. Это очень, очень сложная задача<sup>[35]</sup>. К этому вопросу я советую подходить без предубеждений.

## 29. Странствующий четвертачник, Земля-Двойник и гигантский робот

Может показаться, что, упоминая пунктик Майка насчет рыжих и исходный код в одном абзаце, я призываю читателей не обращать внимания на трещину в фундаменте, на зияющую пропасть в моих рассуждениях об интенциональности – на проблему *исходной интенциональности*. Термин предложил Джон Сёрл (1980), и четкое различие между *исходной* и *производной* интенциональностью, на первый взгляд, кажется интуитивно понятным и даже глубоко убедительным. Доктрина об исходной интенциональности гласит, что, хотя ряд артефактов может обладать интенциональностью, производной от нас (к примеру, книги и фильмы, компьютеры и дорожные знаки), существует также исходная (или внутренне присущая) интенциональность, которую нельзя считать производной. К примеру, напечатанные на этой странице слова говорят о философии только потому, что мы, способные читать и писать на русском языке, обладаем представлениями и убеждениями о философии, которые умудряемся передать, используя следы чернил, однако без нас, пользователей слов, эти знаки вообще ни о чем бы не говорили. Наши мысли и убеждения, напротив, значат то, что они значат, вне зависимости от наличия неочевидных пользователей; они демонстрируют исходную интенциональность и служат источником производной интенциональности многих наших артефактов. К этим артефактам относятся не только слова, предложения и книги, но и карты, фильмы, картины, знаки, символы, диаграммы и другие технические изображения, а также – что особенно важно – компьютеры. Где бы вы ни держали свой список покупок – хоть на клочке бумаги, хоть на смартфоне, – он говорит о продуктах только потому, что вы определенным образом используете символьные структуры и интерпретируете их для подкрепления своего желания купить продукты и убеждения, что идти за ними нужно в супермаркет, причем это желание и убеждение говорят о продуктах более непосредственным образом. Аристотель назвал Бога недвижимым двигателем, а эта доктрина называет нас неозначенными означивателями.

Можно согласиться с Сёрлом, что ничто не обладает *внутренней* интенциональностью просто в силу своей физической формы или других подобных свойств. Если по невероятному совпадению форма

### БЕСПЛАТНОЕ ПИВО

проявится в следах различных минералов на каменистой поверхности Марса, ее нельзя будет (“саму по себе”) считать объявлением о раздаче алкогольного напитка, как бы читателям-землянам ни хотелось интерпретировать ее таким образом. Форма не будет говорить ни о чем, несмотря на первое впечатление. Если какие-то сложные события и объекты мира говорят о других вещах, должно быть, они каким-то образом черпают свою очечность из репрезентации и интерпретации интенциональных систем, состояния которых (убеждения, желания, состояния ума) уже обладают некоторой интенциональностью.

Вопрос в том, *все ли* обладает исходной интенциональностью! На первый взгляд, может показаться очевидным, что нечто должно обладать исходной интенциональностью, поскольку производную интенциональность необходимо от чего-то произвести. Первым кандидатом на обладание исходной интенциональностью стало бы человеческое сознание. Неудивительно, что ряд уважаемых философов, которые в остальном категорически расходятся с Сёрлом, например Джерри Фодор и Сол Крипке, тем не менее соглашались с ним на этот счет. Они – и многие другие, разделяющие эту точку зрения, – полагают, что человеческое сознание (или его психические состояния) обладает исходной интенциональностью и этим радикально отличается от контролируемых роботами систем.

Они все откровенно ошибаются. Да, ошибаются. Я не шучу. Учитывая неоспоримую привлекательность разграничения исходной и производной интенциональности, любая попытка дискредитировать эту теорию рискует быть опровергнутой неуместной снисходительностью: “Не может быть, чтобы он *и правда* полагал, что мы ошибаемся на *этом* счет! Должно быть, он имеет в виду другое и пытается поднять какой-нибудь заумный философский вопрос, который неблагоприятно обрядил в такое нелепо провокационное одеяние!” Вероятно, лучший способ убедить людей, что я действительно имею это в виду, – представить на всеобщее обозрение самый очевидный случай *производной* интенциональности, какой только можно найти, и затем показать, что при ближайшем рассмотрении любимый всеми контраст между этим примером и человеческим сознанием как примером *исходной* интенциональности исчезает без следа. Это трудная задача, но я и не ищу простых путей. Чтобы выполнить ее, мне понадобятся три взаимосвязанных насоса интуиции.

1. *Странствующий четвертачник*. Представьте обыкновенный автомат по продаже напитков, спроектированный и произведенный в Соединенных Штатах и оборудованный приемником американских четвертаков. Назовем этот приемник четвертачником. Обычно, когда в четвертачник вставляется четвертак, четвертачник переходит в состояние – назовем его Ч, – которое “означает” (обратите внимание на пугающие кавычки; оно вроде как означает): “Сейчас я чувствую/принимаю настоящий американский четвертак”. Подобные четвертачники довольно сложны и умны, но не надежны на сто процентов. Они “совершают ошибки”. В частности, порой они переходят в состояние Ч, когда в них вставляют жетон или другой инородный объект, а порой отказываются принимать самые что ни на есть настоящие четвертаки, то есть не переходят в состояние Ч, когда от них это *ожидается*. Несомненно, в этих случаях “ошибочного восприятия” можно выявить закономерности. Также несомненно, что хотя бы некоторые из случаев “ошибочного восприятия” можно предсказать, имея достаточные знания о соответствующих законах физики и технологических параметрах приемника-четвертачника. Применяя законы физики, можно выяснить, что в состоянии Ч четвертачник переводят не только настоящие американские четвертаки, но и объекты некоторого типа К, в то время как объекты типа И (слишком тяжелые) или типа Л (магнитные, в отличие от четвертаков) для этого не подходят. Таким образом, объекты типа К становятся прекрасными *жетонами* и успешно “обманывают” приемник. (Обратите внимание, как часто я использовал в этом абзаце оператор “вроде как”, чтобы применять интенциональную установку при объяснении принципов работы четвертачника. Попробуйте переписать этот абзац, не прибегая к интенциональной установке, и вы увидите, что она необычайно эффективна, а оператор “вроде как” в этих целях практически незаменим.)

Если объекты типа К распространяются в среде, где работают четвертачники, можно ожидать, что владельцы и конструкторы четвертачников усовершенствуют приемники, сделав их более чувствительными, чтобы они безошибочно отличали настоящие американские четвертаки от жетонов типа К. Само собой, могут появиться более изощренные фальшивки, которые потребуют дальнейшего совершенствования механизма распознавания монетоприемника, и в какой-то момент дальнейшие инженерные разработки станут нецелесообразны, поскольку создать *на сто процентов надежный* механизм невозможно. Тем временем конструкторы и пользователи вполне довольны использованием стандартных, примитивных четвертачников, поскольку защищаться от ничтожных злоупотреблений экономически невыгодно.

Единственное, что делает это устройство детектором четвертаков, а не детектором жетонов и не детектором четвертаков или жетонов, это общее намерение конструкторов, производителей, владельцев и пользователей устройства. Одни случаи перехода в состояние Ч можно признать “истинными”, а другие “ошибочными” только в среде или контексте этих пользователей и их намерений. Только в этом контексте намерений мы вообще можем оправдать свое решение назвать это устройство четвертачником.

Насколько я понимаю, пока Сёрл, Фодор, Крипке и остальные согласно кивают: именно так обстоит дело с подобными артефактами – иными словами, это классический пример производной интенциональности. Никто не стесняется признать, что конкретный четвертачник, произведенный на американской фабрике и имеющий маркировку “Четвертачник модели А”, может быть установлен на автомате с газировкой в Панаме, где он будет принимать и отвергать четвертаки бальбоа, официальной валюты Панамы, легко отличимой (человеком) от американских четвертаков по дизайну и отечканенным на ней словам, но не по весу, толщине, диаметру и вещественному составу.

Я ничего не придумываю. У меня есть свидетельство профессионала – Альберта Эрлера из общества нумизматов “Парящий орел”, – что стандартные торговые автоматы не могут отличить американские четвертаки от панамских четвертаков бальбоа, отечканенных в период с 1966 по 1984 г. Неудивительно, ведь их чеканили на болванках американских четвертаков на американских монетных дворах. Для особо любопытных оговорюсь: текущий (2011 г.) обменный курс бальбоа к доллару составляет 0,98, так что четвертак бальбоа сегодня ценится чуть меньше американского четвертака.

Отправленный в Панаму четвертачник все равно будет в большинстве случаев принимать определенное физическое состояние – состояние, характеризующееся физическими свойствами, по которым мы определяли состояние Ч, – всякий раз, когда в него будут вставлять американский четвертак, объект типа К или четвертак панамского бальбоа, но теперь ошибками будут считаться другие случаи. В новой среде американские четвертаки приравниваются к жетонам, как объекты типа К, и ведут к ошибкам восприятия и представления фактов. При этом в США жетоном будет считаться четвертак панамского бальбоа.

Если наш четвертачник переезжает в Панаму, можем ли мы по-прежнему говорить о его переходе в то состояние, которое мы называли состоянием Ч? Физическое состояние, при котором устройство “принимает” монеты по-прежнему возникает, но не следует ли нам теперь говорить, что это физическое состояние необходимо идентифицировать иначе, поскольку “реализуется” новое состояние – состояние ЧБ? Говорить можно по-разному, поскольку четвертачник, в конце концов, всего лишь артефакт, а потому рассуждения об ошибках его восприятия, его истинных и неистинных состояниях – одним словом, о его интенциональности – это “просто метафора”. Внутреннее состояние четвертачника, как его ни назови, *на самом деле* (исходно) не означает ни “я принял американский четвертак”, ни “я принял четвертак панамского бальбоа”. *На самом деле* оно вообще ничего не означает – именно так сказали бы Сёрл, Фодор и Крипке (и многие другие). Его внутреннее состояние лишь *вроде как* означает что-то, но этого достаточно, чтобы сформулировать некоторые проблемы, которые могут возникнуть и перед нами, любителями исходной интенциональности. Давайте рассмотрим пример подробнее.

Изначально четвертачник был разработан как детектор американских четвертаков. В этом заключалась его “собственная функция” (Millikan 1984) и – буквально – смысл существования. Никто не стал бы создавать его, если бы не возникло необходимости в исполнении этой задачи. Учитывая, что история его происхождения предполагает определенную манеру выражаться, такое устройство может главным или должным образом характеризоваться как четвертачник, приспособление для определения четвертаков, а потому *относительно этой функции* определяются как его истинные состояния (когда он действует правильно), так и ошибки.

Это не мешает выломать четвертачник из гнезда и приспособить для выполнения новой задачи – какую бы задачу ему ни позволили эффективным образом выполнять законы физики: он может стать детектором жетонов типа К, детектором четвертаков бальбоа, дверным упором или смертоносным оружием. При переходе к новой роли может возникнуть краткий период смятения или неопределенности. Сколько должен отработать механизм, прежде чем он перестанет быть четвертачником и станет детектором четвертаков бальбоа (ч-бальбером), дверным упором или смертоносным оружием<sup>[36]</sup>? Можно ли назвать его состояние Ч *истинным* при определении четвертака бальбоа, как только он начал работать ч-бальбером после десяти лет верной службы четвертачником, или же переход в это состояние стоит считать своего рода привычной ностальгической ошибкой, неверной идентификацией четвертака бальбоа как американского четвертака?

Как видно, четвертачник разительно отличается от нас, поскольку у него нет воспоминаний о прошлом опыте – и даже вроде как воспоминаний о его прошлом вроде как опыте. Однако при необходимости его можно без труда снабдить подобными воспоминаниями. Чтобы как можно безболезненнее приступить к изучению этой темы, представьте, что четвертачник (будем называть его по изначальному имени) оборудован счетчиком, который за десять лет службы насчитал 1 435 792 операции. Допустим, по пути в Панаму этот счетчик не обнулится, поэтому первая операция на новом месте получит номер 1 435 793. Подтверждает ли это, что четвертачник еще не переключился на выполнение задачи по корректной идентификации четвертаков бальбоа? (В конце концов, он вроде как ошибочно классифицирует это событие как очередное событие Ч – очередное определение американского четвертака, – которое он и должен распознавать в соответствии со своим изначальным предназначением.) Пойдет ли ваша интуиция в другом направлении, если несколько видоизменить или усложнить посылку? (Покрутите все регуляторы этого насоса интуиции и посмотрите, что из этого выйдет.)

Мы можем убедиться, что никакие *внутренние* свойства четвертачника, рассматриваемого в узком контексте и вне зависимости от его прошлого, не позволяют нам отличить его от настоящего ч-бальбера, изготовленного по заказу панамского правительства. И все же, учитывая его происхождение, разве не возникает проблемы с его функцией, значением и смыслом, когда он впервые переходит в состояние, которое нам так хочется назвать состоянием Ч? Можно ли считать это переходом в состояние Ч (означающее “я принял американский четвертак”) или в состояние ЧБ (означающее “я принял четвертак панамского бальбоа”)? Я бы (как и Millikan 1984) сказал, что определить, считается ли его панамский дебют переходом в состояние Ч или в состояние ЧБ, можно только на основании того, был ли он *выбран* на новую роль за свою способность определять четвертаки бальбоа – выбран в буквальном смысле, например владельцем панамской франшизы “Пепси-колы”. Если он был выбран за свою способность, то, даже если новые владельцы и забыли обнулить счетчик, первый акт “восприятия” следует считать случаем верной идентификации, проведенной ч-бальбером, поскольку *теперь* он используется именно *для этого*. Отныне определение четвертаков бальбоа – его собственная функция. Если же четвертачник был отправлен в Панаму по ошибке или прибыл туда случайно, то его дебют ничего не будет значить, хотя его функциональность, вероятно, скоро – и даже немедленно – будет признана и оценена соответствующими



органами (которые могут переключить его на исполнение новой роли), в результате чего все его *последующие* состояния будут считаться состояниями ЧБ. Однако, пока он не выбран для выполнения этой задачи, как бы хорошо он ни определял четвертаки бальбоа, его состояние принятия не будет *означать* (в артефактном, производном, “вроде как” – смысле) “я принял четвертак панамского бальбоа”. Можно предположить, что Сёрл с коллегами позволили бы мне сказать это, поскольку четвертачник, в конце концов, всего лишь артефакт. Он не обладает исходной интенциональностью, поэтому нет смысла копать “глубже” – здесь нет никакого предмета для обсуждения. Сёрл с коллегами сказали бы, что это лишь практический вопрос того, как лучше говорить о состояниях устройства метафорическим и антропоморфическим языком.

Теперь, разобравшись с *производной* интенциональностью, давайте изучим *непроизводную, исходную* интенциональность, то есть *нашу* интенциональность. По этому вопросу Сёрл, Фодор, Крипке и многие другие не соглашались не только со мной, но и с философами Рут Милликен, Полом и Патрицией Черчленд, специалистами по когнитивной науке Дугласом Хофштадтером и Марвином Минским, а также почти со всеми представителями сферы искусственного интеллекта (ИИ). Споры не утихают уже более тридцати лет, и напряжение по-прежнему зашкаливает. В чем же предмет этих споров?

**2. Земля-Двойник.** Допустим, человек по имени Джонс выплядывает в окно и видит лошадь. Лошади там может и не быть, но в своем психическом состоянии он полагает, что видит лошадь, и дело здесь *вовсе* не в интерпретации (как говорят Сёрл и компания), потому что это голый факт, проявление *исходной интенциональности*. Посмотрим, что в таком случае произойдет, если мы проведем мысленный эксперимент, в точности дублирующий панамский. (Подсказка: мы воспользуемся *доведением до абсурда*.) Допустим, планета Земля-Двойник во всем похожа на Землю, но вместо лошадей на ней обитают шмошади<sup>[37]</sup>. Шмошади выглядят, как лошади, поэтому отличить их от лошадей могут только образованные биологи, имеющие в своем распоряжении тест-наборы ДНК, однако шмошади не больше лошади, чем дельфины – рыбы. Обитатели Земли-Двойника называют шмошадь “лошадью”, или *cheval*, или *Pferd*, или еще как-нибудь – не забывайте, Земля-Двойник *в точности* похожа на Землю, за исключением наличия шмошадей.

Допустим, мы отправляем Джонса на Землю-Двойник, где обитают шмошади, но он этого не понимает. (Мы даем ему снотворное на время путешествия, после чего он просыпается в постели своего визави на Земле-Двойнике.) После этого, увидев шмошадь, он, естественно, подумает и скажет: “Смотрите! Лошадь!” Говоря так, он либо по-прежнему действительно полагает, что видит лошадь (то есть демонстрирует ошибочное, не соответствующее действительности убеждение), либо, глядя на шмошадь, впервые в жизни (и в полном соответствии с действительностью) полагает, что видит шмошадь. Как понять, какое из событий происходит на самом деле? Истинно или ложно спровоцированное шмошадью убеждение? Если сначала он ошибочно подумал, что только что увидел лошадь, сколько ему нужно будет прожить среди шмошадей и поговорить о шмошадах с обитателями Земли-Двойника, чтобы адаптировать *значение* слова “лошадь” в своем языке (не понимая этого!)? Если бы на Земле-Двойнике он воспитал детей, что означало бы *для них* усвоенное от отца слово “лошадь” – лошадь или шмошадь? Не забывайте, лошадей они ни разу не видели. Вокруг них всегда были шмошади.

Само собой, это гротескный, утрированный пример, но он поднимает важный вопрос: что определяет смысл наших терминов и каким образом? Превалирует ли всегда и везде история или же повседневное использование слов пересиливает историю и доминирует над нею? В этом случае опыт Джонса не несет нам никакой пользы: он понятия не имеет, что живет уже не на Земле, а потому, вероятно, будет утверждать, что его слово “лошадь” означает *лошадь*. Значение того слова, которое он произносит, происходит из его перцепционного убеждения, и это убеждение ему понятно: он смотрит на лошадь. Именно поэтому он *сказал*: “Смотрите! Лошадь!” (И мог бы добавить: “Это же совершенно очевидно!”) Но теперь представьте, что мы рассказали ему о его путешествии и объяснили едва заметное, но важное различие между лошадьми и шмошадами. Что он скажет в таком случае, что ему следует сказать в таком случае и, что гораздо важнее, есть ли достаточные основания считать любые его слова истинной в последней инстанции? Разве он не делает ровно то, что делал бы любой из нас – а именно, строит теории на тему, о которой должным образом не осведомлен ни он сам, ни мы? Предположим, он говорит, что его слово “лошадь” *теперь* означает шмошадь, то есть, видя шмошадь и называя ее лошадью, он не совершает ошибки. Как говорится, с волками жить...

Может ли он просто определить значения своих слов и тем самым решить вопрос? Что если впоследствии он забудет, какие значения им присвоил? Порой мы так и поступаем: “Отныне

«джабджабом» я называю столовую соль. Передайте джабджаб, пожалуйста!» В контексте научного теоретизирования условные определения весьма распространены, однако они всецело зависят от готовности коммуникантов к взаимодействию. Если Джонс обладает исходной интенциональностью, предположительно он должен в любых обстоятельствах понимать значение собственных терминов, однако выходит, что сам Джонс не в состоянии взаимодействовать со своей исходной интенциональностью лучше нас, людей со стороны. Допустим, к примеру, что мы солгали Джонсу о его перемещении на Землю-Двойник и он поверил нашей лжи (в философских насосах интуиции люди весьма легковерны). Если после этого он скажет нам, что его слово “лошадь” теперь означает шмошадь, будет ли он прав? Вероятно, ему следовало бы сказать, что он понятия не имеет, что теперь означает его слово “лошадь”. Но, если уж на то пошло, мы тоже могли сгонять на Землю-Двойник, а потому не стоит ли нам всем признать, что мы тоже понятия не имеем, что именно значит наше слово “лошадь”?

Те из нас, кто с подозрением относится к идее об *исходной интенциональности*, имеют готовые ответы на все эти вопросы, но окончательно прояснить ситуацию – чтобы эти ответы сумели выстоять в бою с традиционной интуицией – нам поможет третий мысленный эксперимент. (Бойся, читатель! Я собираюсь уговорить тебя отринуть свое чутье.)

**3. Гигантский робот.** Допустим, вы захотели пожить в двадцать пятом веке и существует лишь один способ так долго поддерживать жизнь в вашем теле – поместить его в своеобразное гибернационное устройство, где оно сможет сколько угодно покоиться в коматозном состоянии, пока все происходящие внутри него процессы будут замедлены. Вы можете лечь в капсулу жизнеобеспечения, погрузиться в сон, автоматически пробудиться и выйти наружу в 2401 г.

Проектирование капсулы не единственная инженерная проблема, с которой вам предстоит столкнуться, ведь капсула должна быть защищена и обеспечена необходимой энергией (для охлаждения и чего угодно еще) почти на четыреста лет. Вы не можете положиться в этом на детей и внуков, поскольку они умрут задолго до наступления 2401 г., а для более далеких ваших потомков – при их наличии – ваше благополучие вряд ли окажется в приоритете. В связи с этим вы должны спроектировать суперсистему, которая будет защищать вашу капсулу и снабжать ее необходимой энергией на протяжении четырехсот лет.

Существует две базовые стратегии выполнения этой задачи. Первая заключается в том, чтобы найти идеальное – насколько вы можете предугадать – место для установки капсулы, которая будет в достаточном объеме снабжаться водой, солнечным светом и всеми остальными ресурсами, необходимыми вашей капсуле (и самой суперсистеме) в течение заданного времени. Главный недостаток фиксированной капсулы связан с тем, что ее невозможно передвинуть в случае опасности – скажем, если кто-нибудь решит проложить шоссе ровно там, где она расположена. Вторая стратегия гораздо сложнее, но она решает эту проблему: в соответствии с ней необходимо спроектировать для капсулы мобильный комплекс, снабженный требуемыми датчиками и средствами раннего предупреждения, который сможет в случае опасности перемещаться и по мере надобности искать новые источники энергии. Иными словами, нужно спроектировать гигантского робота и установить капсулу (в которой будете находиться вы сами) внутри него.

Две этих базовых стратегии, очевидно, скопированы из мира живой природы: они примерно соответствуют стратегиям выживания растений и животных. Третья природная стратегия – защищенные споры или семена, которые могут сколько угодно выживать в своей оболочке, – вам недоступна, поскольку ваша система жизнеобеспечения потребляет большое количество энергии, в то время как споры пассивны, а их энергопотребление минимально. Так как нашей цели соответствует стратегия животных, следует предположить, что вы решите построить для своей капсулы робота. Вы должны постараться спроектировать его таким образом, чтобы в первую очередь он “выбирал” действия в ваших интересах. Неверные шаги и ошибки выбора не позволят роботу справиться с защитой вашей жизни до 2401 г., в чем заключается единственный смысл его существования. Само собой, спроектировать такого робота будет очень сложно – вам понадобится высокий уровень профессионализма, чтобы сконструировать “зрительную” систему для руководства его перемещениями, а также другие системы “чувственного восприятия”. Поскольку сами вы будете пребывать в коматозном состоянии и не сможете управлять роботом и планировать стратегии его функционирования, вам необходимо наделить его способностью генерировать собственные планы в ответ на изменение обстоятельств. Он должен “знать”, как “искать”, “распознавать” и затем эксплуатировать источники энергии, как перемещаться на безопасную территорию, как “предвидеть” опасности и затем “избегать” их. Имея такой большой объем работ

и сжатые сроки, вы вынуждены во всем экономить, а потому не наделяете своего робота большей распознающей способностью, чем необходимо для распознавания всего того, что он должен распознавать.

И снова обратите внимание, что я поставил все интенциональные, или “менталистические”, термины, такие как “чувственное восприятие”, “искать” и “предвидеть”, в пугающие кавычки, чтобы обозначить, что это особый тип **вроде как** интенциональности, *производной* интенциональности, интенциональности, которая полностью зависит от ваших человеческих целей. Это *ваши* артефакт, а потому своей интенциональностью он обязан вам, своему создателю. Если бы я опустил пугающие кавычки, меня могли бы обвинить в том, что я пытаюсь приплести сюда идеологию, ссылаясь на тот факт, что инженеры и другие специалисты постоянно используют подобный язык – без пугающих кавычек – для описания характеристик устройства обработки информации (например, **лифтового** контроллера). Я намеренно этого *не делаю*. Напротив, я подчеркиваю, что в рамках этого разговора интенциональный язык используется для описания или реплатментации способностей артефакта *исключительно* метафорически. Также обратите внимание, что механизм робота, как и механизм четвертачника, тоже создается с учетом принципа экономии: он должен “обнаруживать” или “распознавать” множество вещей, но его “распознавания” не будут на сто процентов надежны. Он может совершать ошибки, но *лишь потребности и желания его создателя определяют, что именно будет считаться ошибкой*. Если создатель хотел сконструировать веселого робота-клоуна, который шатался бы по миру, “неправильно распознавая” вещи, то некоторые из его “ошибок” следовало бы считать примерами корректной работы, триумфами системы клоуна.

Вернемся к нашему насосу интуиции: ваша задача усложняется тем, что вы не можете гарантировать, что ваш робот будет единственным роботом, имеющим такую миссию. Если ваше начинание будет продолжено, вашему роботу, возможно, придется конкурировать с другими роботами (и с вашими потомками-людьми) за ограниченные источники энергии, пресной воды, смазок и тому подобных вещей. (См. главу 67, в которой кратко объясняется важность присутствия других агентов.) Несомненно, было бы разумно сконструировать робота таким образом, чтобы его система управления оказалась достаточно совершенна для расчета риска и пользы сотрудничества с другими роботами и формирования взаимовыгодных альянсов, но в то же время любой такой расчет должен представлять собой “быструю и грубую” прикидку, безжалостно урезанную дефицитом времени.

В результате должен получиться робот, способный к некоторому самоконтролю, поскольку, как только вы погрузитесь в сон, вам придется передать своему артефакту пристальный контроль над ситуацией, осуществляемый в реальном времени. В таком случае робот сможет определять вспомогательные цели на основании оценки текущего состояния и влияния этого состояния на итоговую цель (которая по-прежнему заключается в сохранении вашей жизни). Эти вспомогательные цели, возможно, будут уводить его в сторону и вовлекать в вековые проекты, ряд которых может оказаться нерациональным, несмотря на все ваши усилия на стадии разработки. Ваш робот может действовать вразрез с вашими целями и даже совершать самоубийственные поступки, если другой робот, к примеру, “убедит” его отодвинуть плавную миссию собственного существования на второй план.

Обратите внимание, что на этом этапе все интенциональные состояния и действия робота хоть и остаются производными от ваших целей, но начинают несколько отклоняться от этих целей. Поскольку вы сконструировали робота таким образом, чтобы он в некоторой степени “думал за себя”, его “мышление” может выйти за очерченные вами рамки. Реальным, а не выдуманым примером такого артефакта может служить играющий в шахматы компьютер, который способен победить своего создателя в шахматной партии. Да, мы можем сказать, что компьютер в настоящее время “изучает” возможные ходы ладьи ферзевого фланга и “решает” не проводить рокировку, лишь на основании того, что компьютер представляет собой артефакт, созданный человеком именно для совершения этих действий. Но мы также можем сказать, что имеющаяся у создателя цель спроектировать *хорошо* играющий в шахматы компьютер предполагает, что многие выводы создателя о (производной) *очемности* компьютерных состояний оказываются вынужденными: при условии, что шахматисту необходима точная информация о правилах и состоянии партии, *должны* существовать состояния, которые описывают положение каждого слона и каждой пешки, а также состояния, которые предполагают оценку партии, если ферзь компьютера текущим ходом возьмет коня противника, и так далее. Никакие указания не могут связать (производную) очемность состояния компьютера с количеством оставшихся на доске пешек, если только это состояние не было должным образом соотнесено с обнаружением каждой пешки на доске. Когда основная цель создателя достигнута (сделать шахматиста, гигантского робота, симулятор ураганов), в дело вступает жестокая природа, которая определяет, что будет работать, а что работать не будет, а следовательно, какие состояния какой системы считать ошибочными или неточными. Возможно, поэтам и сходит с рук

заявление, что в стихотворении о лошадях на самом деле речь идет о профессуре – как сказал Уильям Блейк, “тигры гнева мудрей лошадей поученья”<sup>[38]</sup>, – но компьютерные специалисты не могут подобным образом переносить свои намерения на свои детища.

Подытожим сказанное. Симулякр психических состояний гигантского робота был бы именно таким – *на самом деле* он ничего не решал бы, ничего не видел, ни о чем не размышлял и ничего не планировал, но *словно бы* решал, размышлял и планировал. Нужно сделать паузу и убедиться, что мы полностью понимаем это утверждение. Воображаемый робот, само собой, гораздо сложнее скромного четвертачника. Мы наделили его способностью “планировать” новые варианты действий, “учиться” на прошлых ошибках, “формировать альянсы” и “коммуницировать” с конкурентами. Более того, чтобы он осуществлял это “планирование”, “обучение” и “коммуникацию”, его необходимо будет снабдить контрольными структурами, обладающими высокой способностью к саморефлексии или самоконтролю. Иными словами, робот, подобно человеку, будет иметь доступ к собственным внутренним состояниям и сможет “сообщать”, “заявлять” и “комментировать”, что “значит” быть выжимкой своих внутренних состояний (когда “решит”, что не “хочет” нас “обманывать”). У него будут “мнения” о смысле этих состояний, и нам следует воспринимать эти мнения всерьез как прекрасное свидетельство – возможно, даже лучшее из доступных нам свидетельств – о том, что “значат” эти состояния, *выражаясь метафорически* (не забывайте: робот остается всего лишь артефактом и не обладает исходной интенциональностью; мы рассматриваем его *производную* интенциональность, которая для наблюдателей не очевиднее нашей интенциональности, интенциональности “настоящих” агентов). Четвертачник не получил такой возможности отмахиваться от наших интерпретативных суждений, с очевидной уверенностью делая “заявления” о том, что он понятия не имеет, что теперь работает в Панаме, или что он был очень удивлен, узнав о существовании четвертаков бальбоа.

Ответить на применение этого насоса интуиции можно по-разному, и вскоре мы рассмотрим несколько вариантов, но сначала я хочу подчеркнуть самое удивительное следствие упрямой приверженности нашей изначальной посылке о том, что ни один артефакт, каким бы мощным ИИ он ни был снабжен, не обладает никакой иной интенциональностью, за исключением производной. Придерживаясь этой позиции, мы вынуждены сделать вывод, что наша собственная интенциональность ничем не отличается от интенциональности робота, поскольку рассказанная мной научно-фантастическая история не нова – фактически она представляет собой вариацию рассуждений Ричарда Докинза (1976), который считает нас и представителей всех остальных биологических видов “машинами выживания”, сконструированными, чтобы продлить существование наших эгоистичных генов. Мы сами артефакты, которые миллиарды лет конструировались в качестве машин выживания для генов, неспособных быстро и информированно действовать в собственных интересах. Наши интересы – как мы себе их представляем – необязательно совпадают с “интересами” наших генов, хотя нас бы и вовсе не было на свете, если бы наши гены не проявляли к этому “интерес”. Сохранение генов – изначальный смысл нашего существования, пускай мы и в состоянии научиться игнорировать этот смысл и определять свое высшее благо, благодаря своему разуму и способности к обучению, которые заложены в нас нашими генами. Таким образом, наша интенциональность производна от интенциональности наших “эгоистичных” генов. Это *они* неозначенные означиватели, а вовсе не мы!

Само собой, интенциональность наших генов ни в коем случае нельзя считать *внутренней*, ведь “значение” каждого гена зависит от целой “алфавитной” системы триплетов АГЦТ, белкового синтеза и развития, если не вдаваться в детали. Но ее можно считать *исходной*, поскольку она является *первой* из многих сформировавшихся впоследствии *репрезентативных* систем. Все более поздние системы обладают агентами – интенциональными системами, – репрезентации которых черпают интенциональность из преследуемых целей (прямо как интенциональность гигантского робота)<sup>[39]</sup>.

Такой взгляд на вещи хоть и дает удовлетворительный ответ на вопрос, откуда берет начало наша интенциональность, но при этом озадачивает нас, поскольку наша интенциональность оказывается производной от сущностей – генов, – интенциональность которых представляет собой хрестоматийный пример *словно бы* интенциональности. Как может буквальное зависеть от метафорического? Более того, между моей научно-фантастической историей и представлениями Докинза, безусловно, есть не менее существенное различие: в то время как в своей истории я предположил, что робот создается посредством сознательной, целеустремленной, прозорливой инженерной работы, по мнению Докинза, даже если мы и представляем собой продукт процесса проектирования, в котором основную выгоду получают гены, в этом процессе проектирования совершенно не задействован сознательный, целеустремленный и прозорливый инженер. Но вскоре мы увидим, что это возражение никуда не годится.

Главная прелесть теории естественного отбора заключается в том, что он показывает нам, как избавиться от разумного создателя в нашем представлении о происхождении видов. И все же процесс естественного отбора рождает весьма замысловатые проекты. Проектировщиками при этом выступают не гены, ведь гены сами по себе предельно глупы – они не умеют ни мыслить, ни представлять, ни постигать что-либо. Они не занимаются проектировкой сами, а выступают лишь в качестве *выгодоприобретателей* процесса проектирования – можно сказать, в качестве *клиентов*. (В нашей истории их можно сравнить с очень глупым, очень богатым клиентом, который нанимает лучших инженеров для создания машины выживания. Если бы не он, инженеры не получили бы хорошо оплачиваемую работу, и именно *его* выживание окупает артефакт, который они создают.) Кто или что занимается проектированием? Само собой, Мать-Природа – или, если выражаться буквально, долгий и медленный процесс эволюции путем естественного отбора.

На мой взгляд, самое удивительное свойство процесса эволюции – его поразительная способность воспроизводить *некоторые* свойства человеческого сознания (разумный творец), не обладая при этом другими его свойствами. Гораздо подробнее об этой теме – о размышлениях об эволюции – мы поговорим в части VI, а пока я хочу описать тесную связь, которая, по моему мнению, существует между любой приемлемой теорией значения и теорией эволюции. Хотя не лишним будет в очередной раз подчеркнуть, что естественный отбор не заглядывает в будущее и не имеет цели, не стоит забывать и тот факт, что процесс естественного отбора зарекомендовал свою исключительную чувствительность к рациональным обоснованиям, без конца делая “выбор”, а также “распознавая” и “оценивая” множество неявных взаимодействий. Выражаясь более провокационно, в процессе работы естественный отбор может “выбирать” конкретную конструкцию *по определенной причине*, но при этом не “представлять” ни свой выбор, ни его причины ни сознательно – ни бессознательно! Сердца были “выбраны”, потому что они превосходно качают кровь, а не потому что их ритмичное биение чарует, но вполне *возможно*, что чарующее биение все же стало *причиной*, по которой естественный отбор “выбрал” что-то другое.

Подобно тому как владелец панамской франшизы “Пепси-колы” может выбрать четвертачник за его способность распознавать четвертаки бальбоа и приспособить его в качестве детектора четвертаков бальбоа, эволюция может выбрать какой-либо орган за его способность насыщать кровь кислородом и сделать его легким. И только на основании подобных проектных “решений” или “обусловленных” эволюцией целей – смыслов существования – мы способны идентифицировать поступки, действия, восприятия, убеждения и другие категории **народной психологии**<sup>[40]</sup>.

Идея о том, что мы сами представляем собой артефакты, спроектированные естественным отбором, одновременно убедительна и знакома. Кое-кто и вовсе сказал бы, что по этому вопросу серьезных разногласий быть не может<sup>[41]</sup>. Почему же тогда ее отвергают не только креационисты и идеологи “разумного замысла”, но и (в некотором роде подсознательно) такие мыслители, как Сёрл, Фодор и другие? Я полагаю, дело в том, что она имеет два неочевидных следствия, которые кое-кому кажутся неудобоваримыми. Во-первых, если мы “всего лишь” артефакты, мы не имеем особой власти над смыслом наших сокровенных мыслей – если эти мысли вообще наделены смыслом, – хотя именно мы эти мысли рожаем. Четвертачник превращается в ч-бальбер, не меняя своей внутренней природы: состояние, которое раньше значило одно, теперь означает другое. То же самое, *в принципе*, может случиться и с нами, если мы представляем собой лишь артефакты, то есть если нашу интенциональность следует считать не исходной, а производной. (Джонс, к примеру, не властен определять, думает он о лошадях или о шмодядах.) Во-вторых, если мы действительно представляем собой артефакты, дело не только в том, что мы лишены гарантированного привилегированного доступа к глубинным фактам, фиксирующим значение наших мыслей, но и в том, что *этих глубинных фактов не существует*. Порой функциональная интерпретация очевидна (сердце очевидно представляет собой насос, а глаз очевидно нужен, чтобы видеть), но когда это не так и нам приходится читать мысли Матери-Природы, никакого текста для интерпретации нет. Когда “предмет для обсуждения” о собственной функции рождает противоречия – когда есть достаточные основания для более чем одной интерпретации, – этого предмета просто не существует!

### 30. Радикальный перевод и кроссворд Куайна

Утверждение о том, что при столкновении двух в равной степени вероятных функциональных интерпретаций не существует глубинных фактов, на основании которых вопрос может быть решен однозначно, лучше и ярче всего аргументировал философ У. В. О. Куайн (1960), предложивший принцип *неопределенности радикального перевода* и отстаивавший его с помощью знаменитого насоса интуиции. Допустим, в Тихом океане открыли обособленный остров, скажем, населенный непонятно откуда появившимися людьми, говорящими на языке, на котором не говорит больше никто. В отсутствие двуязычных переводчиков с этого языка антропологи и лингвисты должны постичь его путем наблюдений и череды проб и ошибок при взаимодействии с местным населением – эту задачу Куайн назвал “радикальным переводом”. *В принципе*, заявил Куайн, два ученых, которым поручена задача составления пособия по переводу с этого экзотического языка, могут создать существенно разные, но в равной степени качественные пособия, где одни и те же слова местных жителей будут переводиться по-разному, но однозначно определить, какой из переводов верен, будет невозможно! Многим философам эта идея кажется слишком радикальной, чтобы воспринимать ее всерьез, поэтому они просто не обращают на нее внимания и придерживаются своих традиционных методов. Вот инструмент мышления, созданный специально для того, чтобы эта идея стала казаться как минимум правдоподобной, если не совершенно очевидной. Нужно объяснить две вещи: (1) как утверждение Куайна может быть правдой (“в принципе”) и (2) почему, несмотря на это, не представляется возможным привести для него реальный пример.

Я регулярно предлагаю своим студентам решить следующий кроссворд. Через несколько минут большинство из них заявляют, что справились с задачей. Прежде чем читать дальше, попробуйте решить его сами.

|    |    |    |    |
|----|----|----|----|
| 1. | 2. | 3. | 4. |
| 5. |    |    |    |
| 6. |    |    |    |
| 7. |    |    |    |

*По горизонтали:*

1. Нечто грязное.
5. Великая человеческая потребность.
6. Разглаживать.
7. Кинозвезда.

*По вертикали:*

1. Средство передвижения, зависящее от H<sub>2</sub>O.
2. Обычно мы этого хотим.
3. Наверху.
4. Штат США (аббревиатура).

Вы справились с задачей? Если да, какое решение вы нашли? Кроссворд имеет два решения, примерно в равной степени изящных (они спрятаны далее на страницах этой книги, чтобы вы успели найти оба, прежде чем я их оплашу). Хотя кроссворд и невелик, на его составление у меня ушло несколько часов, поскольку необходимо было выполнить многочисленные требования, значительно ограничивающие количество возможных вариантов. Если вы мне не верите, попробуйте составить кроссворд побольше и получше! (Прошу вас, если вам это удастся, пришлите его мне. Я стану использовать его вместо собственного.)

Любой, кто спрашивает: “Что на самом деле значит номер 1 по вертикали?” – предается неуместному реализму. *Предмета для обсуждения не существует.* Я специально составил кроссворд таким образом, чтобы этого предмета не существовало. К примеру, я не пробовал составить кроссворд с одним набором ответов (исторически первых, или оригинальных, а “следовательно”, *настоящих* ответов), а затем придумать второй набор. Я работал над двумя решениями одновременно, выбирая наиболее подходящие слова из составленного мною списка четырехбуквенных слов со сходными значениями.

Составить такой кроссворд возможно, поскольку нормы определений допускают некоторую гибкость. Оба решения включают в себя слова, которые едва подходят под данные определения, но совокупность остальных ответов (холизм, если использовать жаргон философов) стягивает слова в две довольно

стабильные конфигурации. Какова вероятность, что не существует третьего решения, которое было бы ничуть не хуже первых двух? *Криптографическая максима* гласит: если найдено *одно* решение задачи, это решение *единственное*. Только в особых обстоятельствах решения может быть два, но подобные примеры показывают нам, что существование единственного ответа на вопрос такого рода есть не метафизическая необходимость, а лишь в высшей степени вероятный результат весьма серьезных ограничений.

Люди гораздо сложнее кроссвордов и компьютеров. Их мозг устроен изощренным образом, полон нейромодуляторов и существует в теле, которое тесно взаимодействует с миром. Кроме того, их эволюционная и личная история внедряет их в мир, предполагающий гораздо больший размах интерпретации, чем внедрение кроссворда в лингвистическое сообщество. В связи с этим Рут Милликен (к примеру) права, когда утверждает, что с учетом природы конструктивных ограничений крайне маловероятно существование различных способов содрать шкуру с кошки, которые предполагали бы две радикально разные, совершенно неопределимые, абсолютно равнозначные интерпретации. На практике неопределенностью радикального перевода можно пренебречь. И все же принцип работает. Неопределенности радикального перевода не возникает *не потому*, что – и это метафизический факт – там, в голове, существуют “реальные значения” (Куайн называл это “мифом о музее” значений, своей главной целью). Неопределенности в реальном мире не возникает потому, что существует слишком большое количество независимых условий, которые надо удовлетворить, в связи с чем криптографическая максима заверяет нас, что поводов для волнения ничтожно мало. Когда неопределенность грозит реальному миру, определенность прочтения в итоге обеспечивают всегда “поведенческие” или “диспозиционные” – и тому подобные – факты, а не какая-то загадочная “каузальная сила” или “внутренняя семантичность”. Интенциональная интерпретация почти всегда ограничивается единственной интерпретацией, но в критической ситуации, которая теоретически может возникнуть, если все проверки пройдут сразу две интерпретации, не будет никаких *глубинных* фактов, чтобы определить, какая из них “верна”. Факты действительно определяют интерпретации, но эти факты всегда “поверхностны”, а не глубинны.



## 31. Семантические и синтаксические движки

Как значение может создавать различие? Казалось бы, значение не является физическим свойством вроде температуры, массы или химического состава, а потому вряд ли может что-то спровоцировать. *Задача* мозга состоит в том, чтобы извлекать *значение* из потока энергии, поступающего через органы чувств, с целью улучшения перспектив тела, в котором этот мозг пребывает и которое питает его энергией. Мозг должен “творить будущее” в форме предвосхищения *всех вещей, которые важны* для адекватного управления телом. Мозг требует огромных энергетических затрат, а потому, если он не справляется со своей работой, он даром ест свой хлеб. Иными словами, мозг должен выполнять функцию *семантического движка*. При этом мозг *состоит* из бесчисленного множества молекулярных фрагментов, которые взаимодействуют друг с другом согласно строгим законам химии и физики, реагируя на формы и силы. Иными словами, мозг на самом деле представляет собой лишь *синтаксический движок*.

Допустим, вы попросили инженеров сконструировать определитель подлинных долларовых купюр – или детектор фальшивок, что, в общем-то, одно и то же: в соответствии с техническим заданием он должен складывать все подлинные доллары в одну стопку, а все фальшивки – в другую. Это невозможно, скажут инженеры, потому что сконструированная нами машина сможет реагировать только на “синтаксические” свойства: физические детали – толщину и химический состав бумаги, форму и цвет чернильных узоров, наличие или отсутствие других физических характеристик, которые сложно подделать. Они скажут, что на основании этих “синтаксических” характеристик *могут* сконструировать достаточно хороший, но не защищенный от ошибок детектор фальшивок. Он будет стоить дорого, но опосредованно и с погрешностью все же будет выявлять фальшивки достаточно хорошо, чтобы окупить свою разработку и эксплуатацию.

Любая конфигурация элементов мозга подвержена таким же ограничениям. Под действием физико-химических сил она будет выполнять свою задачу вне зависимости от смысла вводимых данных (или их **вроде как** смысла). Не стоит заблуждаться и полагать, будто мозг, будучи живым и состоящим из белков, а не из кремния и металла, может определять значения непосредственным образом, благодаря содержащейся в нем **чудо-ткани**. Физика всегда превыше значения. Настоящий семантический движок, реагирующий *непосредственно* на значения, подобен вечному двигателю – он физически невозможен. Как же тогда мозг справляется со своей задачей? Будучи синтаксическим движком, он *копирует* или *воспроизводит* компетенции невозможного семантического движка<sup>[42]</sup>. Это вообще возможно? Некоторые философы утверждают, что если микрокаузальная теория работы мозга полна (и не имеет никаких таинственных пробелов), то значения просто не имеют возможности ни на что повлиять. В главе 33 мы познакомимся с насосом интуиции, который демонстрирует, что это не так, показывая, как семантические свойства – такие как истина, значение или ссылка – играют *неустранимую* роль в некоторых очевидно каузальных процессах. Но прежде чем обратиться к этому достаточно сложному насосу интуиции, я хочу изучить более простую модель, которая позволит нам обнажить ряд сомнений относительно философских насосов интуиции в целом, а также при удачном раскладе поможет развеять некоторые опасения, мешающие полному пониманию.



## 32. Болотный человек и коровоакула

Насосы интуиции должны работать четко и эффективно, подкачивая необходимую интуицию, а затем снова отправляясь в запас. Но часто насосы интуиции провоцируют целый сонм опровержений, контрольных проверок, поправок и расширений. Один из лучших американских философов двадцатого века Дональд Дэвидсон однажды сказал мне, что жалеет о создании этого насоса интуиции, поскольку он разжигает бесконечные и лишь периодически проливающие свет на проблему споры. Знакомьтесь с болотным человеком – любимцем многих философов, но только не Дэвидсона (1987):

Допустим, молния ударила в сухое дерево на болоте, а я стоял рядом. Мое тело разлетелось на мельчайшие частицы, в то время как дерево по случайному совпадению (и с использованием других молекул) превратилось в моего физического двойника. Мой двойник, болотный человек, движется ровно так, как двигался я: следуя своей природе, он уходит с болота, встречает и вроде бы узнает моих друзей, а также здоровается с ними по-английски. Он живет у меня дома и, кажется, пишет статьи о радикальной интерпретации. Никто не видит разницы.

Но разница *есть*. Мой двойник не узнает моих друзей – он вообще не может ничего узнавать, поскольку он ничего не познавал изначально. Он не может знать, как зовут моих друзей (хотя и кажется, что он это знает); он не может помнить мой дом. К примеру, он не может обозначать словом “дом” то же самое, что обозначаю им я, потому что произносимое им слово “дом” не было усвоено им в контексте, который мог бы придать ему верное значение – или какое-либо значение вообще. Полагаю, ни при каких условиях нельзя сказать, что слова моего двойника наделены значением, а также что он имеет какие-либо мысли. [pp. 443–444]

От внимания философов не ускользнуло, что наши коллеги-ученые из других областей – особенно специалисты по точным наукам – нередко с трудом скрывают скептицизм, когда вопросы о Земле-Двойнике и болотном человеке начинают рассматриваться всерьез. В чем же дело? Виноваты ученые, которые предстают неискушенными обывателями, не воспринимающими тонкости философских изысканий, или же философы, потерявшие связь с реальностью? **Скорее** (это вам намек), мне не стоит об этом говорить.

Подобные гротескные примеры доказывают то или иное положение, намеренно сводя на нет все характеристики явления, кроме одной, которую обычно обходят вниманием, и в результате действительно важное становится очевидным. В примере с Землей-Двойником внутреннее сходство максимально (нас отправляют на Землю-Двойник, но не дают вам шанса заметить эту серьезную перемену), поэтому ответственным за подсказки нашей интуиции можно считать внешний контекст. В примере с болотным человеком будущие диспозиции и внутренние состояния остаются неизменными, а “история” сводится на нет. Таким образом, эти мысленные эксперименты по своей структуре напоминают научные эксперименты, в которых предпринимается попытка изолировать важное взаимодействие переменных, при условии, что все остальные переменные не изменяются. Проблема таких экспериментов в том, что зависимой переменной в них выступает интуиция – ведь это и есть насосы интуиции, – а участие воображения в генерации интуитивных озарений контролировать сложнее, чем признают философы. (Несколько **упорных башмаков**, которые мы изучим, на самом деле подавляют воображение читателей, искажают их интуитивные озарения и тем самым обесценивают “результаты” мысленного эксперимента.)

Но в этих экспериментах возникает и другая, более глубокая проблема. Придумывать примеры для “доказательства” других концептуальных моментов сродни тому, чтобы играть в игру. Допустим, корова родила существо, на атомном уровне неотличимое от акулы. Будет ли оно акулой? Если задать этот вопрос биологу, он в лучшем случае сочтет, что вы неуклюже пытаетесь пошутить. Или допустим, что злой демон может своей улыбкой обращать воду в твердое состояние при комнатной температуре. Будет ли эта демоническая вода льдом? Эта гипотеза слишком глупа, чтобы всерьез на нее отвечать. По мнению некоторых философов, логика допускает существование улыбающихся демонов, коровоакул, зомби и болотных людей, даже если их существование не допускает номология (или каузальность). Этим философам это кажется важным. Но не мне. Надо полагать, столь широкий размах контрфактуальности мотивирован тем, что получаемый в итоге ответ говорит нам о *сущности* обсуждаемой темы. Но кто сегодня верит в такую суть? Только не я.

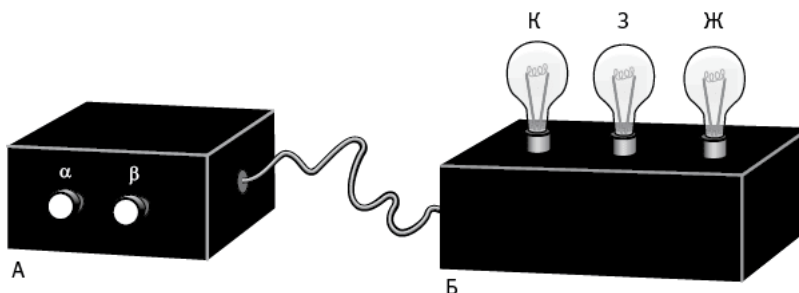
Рассмотрим параллельный вопрос, который мы могли бы задать о магнитах, заметив, что есть два конкурирующих кандидата на роль “определителя истины” – основного свойства, или сути, – магнитов: (а) все магниты притягивают железо и (б) все магниты обладают определенной внутренней структурой (назовем ее М-структурой). Был ли старый, поведенческий критерий (а) в итоге вытеснен новым, структурным критерием (б), или же последний просто редуционистским образом объяснил первый? Чтобы выяснить это, нужно представить, как мы задаем ученым следующие вопросы в духе примера с болотным человеком. Допустим, вы нашли объект, который притягивает железо, но не обладает М-структурой (как стандартные магниты). Назовете ли вы его магнитом? Или же, допустим, вы нашли объект, который обладает М-структурой, но не притягивает железо. Назовете ли вы магнитом его? Физики ответили бы, что, случись им наткнуться на любой из этих объектов, перед ними встали бы гораздо более серьезные вопросы, чем вопрос об их названии. Вся их научная картина зависит от наличия глубинной закономерности между структурой атомных диполей в магнитных доменах и магнетизмом железа, поэтому “факт” *логической* возможности нарушить эту закономерность их не интересует. Однако их интересует реальная ковариантность “структурных” и “поведенческих” факторов. Обнаружив нарушение закономерностей, они внесут соответствующие изменения в свои научные теории и позволят терминам самим встать по местам.

Обладает ли болотный человек мыслями и говорит ли по-английски? Считать ли коровоакулу акулой? Она плавает, как акула, и успешно спаривается с другими акулами. Ах, разве я вам не сказал? На атомном уровне она неотличима от акулы, *вот только* в ее клетках коровья ДНК. Невозможно? *Логически* невозможным признать это нельзя (скажут философы). Но очевидная невозможность этого делает дальнейшую дискуссию нецелесообразной. Совершенно ясно, скажем, что “следы” воспоминаний Дэвидсона физически не могут оказаться в структуре мозга болотного человека, а акула не может сформироваться из клеток, содержащих коровью ДНК. Хотя существование болотного человека и нельзя признать логически невозможным, пускай и потому только, что грандиозные совпадения наподобие тех, которые необходимы для создания сущностей вроде болотного человека, нельзя считать логически невозможными *по определению*, но они никогда не происходят на самом деле, так кому какая разница, *что мы сказали бы*, если бы это все же произошло?

“Мне это важно, – скажет философ, который остерегается **риторических вопросов**. – На мой взгляд, необходимо всегда предельно четко определять все термины, учитывая все логически возможные исходы. Только так можно добраться до истины”. Но так ли это? В реальном мире прошлая история и будущая функция связаны многожильными проводами эволюции, развития и обучения. Именно *потому, что* тело Дэвидсона годами двигалось по определенной траектории, Дэвидсон получил все свои воспоминания, убеждения и представления, а *настоящей* замены этим процессам естественного накопления знаний не существует. На мой взгляд, вынесение вердикта по воображаемым вопросам, нарушающим эти условия, нецелесообразно. Вообще-то мне кажется, что подобные вымученные примеры приводятся специально, чтобы создать воображаемую дихотомию, скорее склоняющую вас к вседозволенности. “Нет, – скажет философ. – Это *не* ложная дихотомия! *Чисто теоретически* мы приостанавливаем действие законов физики. Разве не так поступил Галилей, когда не стал учитывать в своем мысленном эксперименте силу трения?” Да, но общее правило проистекает из сравнения: полезность мысленного эксперимента обратно пропорциональна степени его отклонения от реальности.

Земля-Двойник физически невозможна, но не настолько невозможна, как болотный человек! (Не думайте, что многомировая интерпретация квантовой механики, которая завоевывает все большую популярность в определенных кругах, доказывает физическую возможность существования Земли-Двойника; даже при условии существования бесконечного множества вселенных, включая [бесконечное?] множество вселенных с планетами, почти в точности похожими на Землю, мы не можем отправить ни на одну из них землянина.) Отправка четвертачника в Панаму, напротив, не просто возможна – вполне вероятно, такое даже случалось на самом деле. Нам не нужно приостанавливать действие законов природы, чтобы представить это путешествие в мельчайших подробностях.

### 33. Два черных ящика



Жили-были два больших черных ящика, А и Б, соединенные длинным изолированным медным проводом. На ящике А было две кнопки, обозначенных  $\alpha$  и  $\beta$ , а на ящике Б было три лампочки, красная, зеленая и желтая. Ученые, изучавшие поведение этих ящиков, заметили, что при каждом нажатии кнопки  $\alpha$  на ящике А на ящике Б ненадолго загоралась красная лампочка, а при каждом нажатии кнопки  $\beta$  на ящике А ненадолго загоралась зеленая лампочка. Желтая лампочка не загоралась никогда. Ученые провели несколько миллиардов тестов в различных условиях, но не обнаружили исключений. Они пришли к выводу о наличии каузальной закономерности, которую сформулировали следующим образом:

все нажатия на  $\alpha$  включают красную лампочку  
все нажатия на  $\beta$  включают зеленую лампочку

Они установили, что эта каузальная связь каким-то образом осуществляется через медный провод, потому что при перерезании провода ящик Б переставал реагировать, а при простой изоляции ящиков без перерезания провода закономерность не нарушалась. Естественно, ученым хотелось узнать, как обнаруженная ими каузальность передается по проводу. Возможно, подумали они, нажатие на кнопку  $\alpha$  пускает по проводу низковольтный импульс, включающий красную лампочку, а нажатие на кнопку  $\beta$  пускает высоковольтный импульс, включающий зеленую лампочку. А возможно, нажатие на кнопку  $\alpha$  пускает единичный импульс, включающий красную лампочку, а нажатие на кнопку  $\beta$  – двойной. Очевидно, в проводе всегда должно происходить что-то одно при нажатии на кнопку  $\alpha$  и что-то другое при нажатии на кнопку  $\beta$ . Необходимо было установить, что именно происходит, поскольку это знание объяснило бы замеченную учеными каузальную закономерность.

Вскоре на провод установили жучок, который показал, что на самом деле все сложнее. При нажатии *любой* из кнопок на ящике А по проводу к ящику Б посылалась длинная цепочка прерывистых импульсов – переменных сигналов, или бит (если точно, 10 000 бит). Но цепочка эта каждый раз была разной!

Очевидно, цепочки бит должны были обладать каким-то свойством или атрибутом, который обуславливал включение красной лампочки в одном случае и зеленой в другом. Что это могло быть за свойство? Ученые решили открыть ящик Б и посмотреть, что происходит с цепочками бит, когда они достигают цели. Внутри ящика Б ученые обнаружили обычный цифровой серийный суперкомпьютер, располагающий большим объемом памяти, в котором содержалась огромная программа и огромная база данных – и обе они, само собой, были написаны другими цепочками бит. Проследив влияние входящих цепочек бит на программу этого компьютера, ученые не заметили ничего необычного: входящие цепочки всегда добирались до ЦП (центрального процессора) в неизменном виде, после чего над ними за считанные секунды производилось несколько миллиардов операций, в результате чего всегда получался один из двух исходящих сигналов, единица (включавшая красную лампочку) или ноль (включавший зеленую лампочку). Ученые выяснили, что в любом случае они могли объяснить каждый шаг каузальности на микроскопическом уровне без каких-либо сложностей и противоречий. У них не было оснований предполагать вмешательство сверхъестественных сил, а когда, к примеру, они пробовали снова и снова отправлять ЦП одну и ту же цепочку бит, программа в ящике Б всегда выдавала один и тот же результат, зажигая красную или зеленую лампочку.

Но это было довольно любопытно, поскольку, хоть ящик Б и выдавал всегда один и тот же результат, промежуточные шаги на пути к нему различались. Он почти всегда проходил через разные физические состояния, прежде чем выдать одинаковый результат. Само по себе это не представляло загадку, поскольку

программа сохраняла копию каждой входящей цепочки бит, а потому, когда та же самая цепочка приходила во второй, третий или тысячный раз, состояние памяти компьютера всякий раз несколько отличалось. Но результат всегда оставался одним: если при первом получении конкретной цепочки бит загоралась красная лампочка, затем для той же цепочки всегда загоралась красная лампочка, и та же закономерность действовала и для зеленых цепочек (как их стали называть ученые). Ученым очень хотелось выдвинуть гипотезу, что все цепочки либо красные (они зажигают красную лампочку), либо зеленые (они зажигают зеленую лампочку). Но, само собой, ученые тестировали не все возможные цепочки, а лишь цепочки, генерируемые ящиком А.

Решив проверить свою гипотезу, ученые на время отсоединили ящик А от ящика Б и стали посылать в ящик Б вариации исходящих цепочек ящика А. Какой же случился переполох, когда ученые увидели, что измененные ими цепочки ящика А почти всегда провоцировали включение *желтой* лампочки! Казалось, ящик Б обнаруживает их вмешательство. Однако не было сомнений, что ящик Б с готовностью примет созданные человеком версии красных цепочек, мигнув красной лампочкой, и созданные человеком версии зеленых цепочек, мигнув зеленой лампочкой. Желтая лампочка обычно – почти всегда – загоралась лишь тогда, когда в красной или зеленой цепочке меняли один – или более – бит. “Вы убили ее!” – воскликнул кто-то однажды, увидев, как “подделанная” красная цепочка превращается в желтую, и это дало толчок множеству спекуляций о том, что красные и зеленые цепочки в некотором роде *живы* – возможно, они представляют собой мужское и женское начало, – а желтые цепочки *мертвы*. Какой бы привлекательной ни казалась эта гипотеза, она в итоге ни к чему не привела, хотя шквал экспериментов с несколькими миллиардами случайных вариаций цепочек в 10 000 бит длиной и подталкивал ученых к выводу о существовании трех типов цепочек – красных, зеленых и желтых, – причем в общей массе количество желтых цепочек было на много порядков больше количества красных и зеленых цепочек (подробнее об этом см. в главе 35). Почти все цепочки были желтыми. Это делало обнаруженную учеными красно-зеленую закономерность лишь более любопытной и загадочной.

Что же в красных цепочках зажигало красную лампочку, и что в зеленых цепочках зажигало зеленую? *Само собой, в каждом конкретном случае никакой загадки не было.* С помощью суперкомпьютера в ящике Б ученые могли проследить каузацию в каждой конкретной цепочке и увидеть, как она, демонстрируя отрядный детерминизм, провоцировала включение красной, зеленой или желтой лампочки в соответствии с характером цепочки. Однако они не могли понять, как определить, какую из трех лампочек будет зажигать новая цепочка, просто изучая ее (и не “моделируя” ее эффект в ящике Б). Эмпирические данные свидетельствовали об очень высокой вероятности того, что новая цепочка будет желтой, если только это не цепочка, исходящая из ящика А, поскольку в таком случае с вероятностью более чем миллиард к одному эта цепочка должна быть либо красной, либо зеленой, хотя никто и не мог сказать, какой именно, не пропустив ее через ящик Б и не увидев, как сработала программа.

Возможно, разгадка таилась в ящике А. Открыв его, ученые нашли второй суперкомпьютер другой конфигурации и модели, на котором работала другая сложная программа, но при этом он все же был обычным цифровым компьютером. Вскоре ученые установили, что в него встроены “часы”, тикающие по несколько миллионов раз за секунду, и всякий раз при нажатии на любую из кнопок компьютер первым делом проверял “время” по этим часам (например, 1011010101010111) и разбивал его на цепочки, которые затем использовал, чтобы определить, какие подпрограммы в каком порядке исполнять и к какому фрагменту памяти обращаться в первую очередь при подготовке цепочки бит для отправки по проводу.

Ученые смогли определить, что эта сверка часов (которая была фактически случайной) на деле гарантировала невозможность повторной отправки одной и той же цепочки бит. Однако, несмотря на случайный или псевдослучайный характер этого события, при каждом нажатии кнопки  $\alpha$  компьютер неизменно выдавал цепочку бит, включающую красную лампочку, а при каждом нажатии кнопки  $\beta$  – цепочку бит, включающую зеленую лампочку. Ученым удалось даже обнаружить несколько аномалий: примерно в одном из миллиарда случаев нажатие кнопки  $\alpha$  приводило к выдаче зеленой цепочки или нажатие кнопки  $\beta$  приводило к выдаче красной цепочки. Этот крошечный изъян в совершенной картине лишь раззадоривал ученых, желавших объяснить закономерность.

В один прекрасный день о себе заявили два ИИ-программиста, которые создали ящики, и объяснили, как они устроены. (Не читайте это, если хотите разгадать загадку сами.) Ящик А сконструировал Ал, который много лет работал над “экспертной системой” – базой данных, содержащей “истинные высказывания” обо всем на свете, и механизмом вывода для обработки содержащихся в базе данных аксиом. База данных хранила статистику Главной лиги бейсбола, метеорологические данные, биологическую таксономию, историю народов мира и множество других сведений. Ящик Б создал швед

Бо, который в то же время работал над конкурирующей базой данных “мирового знания” для собственной экспертной системы. Каждый из них наполнил свои базы данных таким количеством “истин”, какое смог загрузить за годы работы<sup>[43]</sup>.

Однако со временем экспертные системы их утомили, и оба программиста решили, что практические перспективы этой технологии сильно преувеличены. Системы не слишком хорошо решали интересные задачи, или не слишком хорошо “думали”, или не слишком хорошо “находили нестандартные пути решения проблем”. Благодаря своим машинам выбора, хорошо они справлялись лишь с одним – с генерацией множества истинных предложений (на соответствующих языках) и проверкой входящих предложений (на соответствующих языках) на истинность или ложность относительно их вроде как знания. Ал и Бо встретились и нашли применение плодам своего тщетного труда. Они решили сделать философскую игрушку. Выбрав лингва франка для перевода между двумя имеющимися у них репрезентативными системами (им стал английский в стандартной кодировке ASCII)<sup>[44]</sup>, они соединили машины с помощью провода. При каждом нажатии кнопки  $\alpha$  машина А случайным (или псевдослучайным) образом выбирала одно из своих “убеждений” (либо хранимую в ее памяти аксиому, либо вывод, сделанный на основе имеющихся у нее в распоряжении аксиом), переводила его на английский (на компьютере английские буквы заранее задавались в кодировке ASCII), добавляла достаточное количество бит после точки, чтобы общее число бит достигло 10 000, и отправляла итоговую цепочку машине Б, которая переводила входящие данные на свой язык (шведский лисп) и сравнивала со своими “убеждениями” – то есть со своей базой данных. Поскольку обе базы данных содержали истины, причем примерно одинаковые, благодаря работе машин вывода, всякий раз, когда А отправляла Б “убеждение”, которое считала истинным, Б тоже признавала это “убеждение” истинным и оповещала об этом миганием красной лампочки. Когда же А отправляла Б “убеждение”, которое считала ложным, Б признавала его ложным и объявляла о своем решении миганием зеленой лампочки.

Вносимые в цепочку изменения почти всегда приводили к возникновению ошибок в зашифрованной в этой цепочке английской фразе, если только изменения не вносились лишь в случайную последовательность бит в самом конце. Непримируемая к опечаткам, Б реагировала на это миганием желтой лампочки. При использовании случайной цепочки бит вероятность того, что эта цепочка не будет четко сформулированной по-английски в кодировке ASCII истиной или ложью, была **чрезвычайно велика**, а потому чаще всего загоралась желтая лампочка.

Итак, сказали Ал и Бо, таинственная каузальность красного на самом деле объяснялась тем, что *красный* обозначал истинность английской фразы, а *зеленый* – ее ложность. Вдруг многолетние старания ученых превратились в детскую игру. Теперь любой мог до умопомрачения генерировать красные цепочки. Достаточно было просто записывать фразы вроде “Дома больше орехов”, “Киты не летают” или “Трижды четыре на два меньше, чем дважды семь” в кодировке ASCII. Для генерации зеленой цепочки подошли бы фразы вроде “Девять меньше восьми” или “Нью-Йорк – столица Испании”. Вскоре философы сделали ряд любопытных находок – так, они обнаружили цепочки, которые были красными при первых ста отправках Б, но зелеными после (например, написанная в кодировке ASCII фраза “Эта фраза отправлялась тебе на проверку менее ста одного раза”).

Но некоторые философы заметили, что *красная* и *зеленая* характеристики цепочек на самом деле не тождественны *английской истине* и *английской лжи*. В конце концов, существуют английские истины, для записи которых в кодировке ASCII потребуются миллионы бит. К тому же, несмотря на все старания, Ал и Бо не всегда помещали в свои программы *факты*. К примеру, некоторые расхожие представления, считавшиеся истинными в то время, когда программисты работали над своими базами данных, с тех пор были опровергнуты. Характеристику цепочки – каузальную характеристику – *красноты* по множеству причин нельзя было считать тождественной характеристике *английской истины*. Итак, возможно, лучше было сказать, что *красный* означает *относительно короткую фразу на английском языке в кодировке ASCII, описывающую нечто, что вроде как считает истинным ящик Б (вроде как убеждения которого почти всегда истинны)*. Одних философов удовлетворила эта формулировка, но другие стали придираться, по разным причинам настаивая, что это определение нельзя признать точным или что существуют контрпримеры, которые невозможно отбросить не *ad hoc*, и так далее. Но, как заметили Ал и Бо, лучшего описания этого свойства не найти, тем более что ученым все равно нужно было именно такое объяснение. Разве загадка красных и зеленых цепочек не была полностью решена? Более того, разве теперь, после ее решения, не стало очевидно, что не стоило и надеяться объяснить *каузальную* закономерность тем, с чего мы начали рассказ – что все нажатия на  $\alpha$  включают красную лампочку, а все

нажатия на  $\beta$  включают зеленую лампочку, – не прибегая к использованию *некоторых семантических* (или *менталистских*) терминов?

Некоторые философы заявили, что новообретенное описание закономерности происходящей в проводе активности можно использовать для предсказания поведения ящика Б, однако закономерность эта не *каузальна*. Истинность и ложность (и любые скорректированные заменители, описанные выше) представляют собой семантические характеристики, а потому совершенно абстрактны, из чего следует, что они не могут лежать в основе каузации. Чепуха, ответили другие. Нажатие на кнопку  $\alpha$  приводит к включению красной лампочки с той же неизбежностью, с которой поворот ключа в замке зажигания приводит к включению двигателя вашей машины. Если бы оказалось, что по проводу передаются просто высоко- и низковольтные сигналы или единичные и двойные импульсы, никто не стал бы сомневаться, что это образцовая каузальная система. Тот факт, что эта система оказалась машиной Руба Голдберга, не показал снижения каузальности надежности связи между нажатием на  $\alpha$  и включением красной лампочки. Более того, в каждом конкретном случае ученые могли проследить точный путь микрокаузации, полностью объясняющий результат<sup>[45]</sup>.

Убежденные этими аргументами, другие философы стали утверждать, будто это показывает, что характеристики красноты, зелени и желтизны *на самом деле* не были семантическими или менталистскими, но были лишь имитацией семантических характеристик, лишь “словно бы” семантическими характеристиками. На самом деле краснота и зелень представляли собой очень, очень сложные *синтаксические* характеристики. Однако эти философы отказывались говорить, какие именно это были синтаксические характеристики, или объяснять, как даже маленькие дети могут быстро и надежно генерировать или распознавать их. Тем не менее эти философы полагали, что *должно быть* чисто синтаксическое описание закономерности, поскольку, в конце концов, рассматриваемые каузальные системы были “просто” компьютерами, а компьютеры – это “просто” синтаксические движки, не способные к реальной “семантичности”.

“Мы полагаем, – ответили Ал и Бо, – что если бы внутри черных ящиков вы обнаружили нас, играющих с вами по той же схеме, то вы бы смягчились и согласились, что оперативной каузальной характеристикой была настоящая истина (или, во всяком случае, истина, которая считалась настоящей). Можете ли вы предложить достаточное основание для этого разделения?” Одних это спровоцировало сказать, что в определенном важном смысле Ал и Бо *на самом деле* сидели в ящиках, поскольку именно они создали соответствующие базы данных как модели собственных убеждений. Других это спровоцировало заявить о полном отсутствии семантических или менталистских характеристик в мире. Содержимое, сказали они, *уничтожено*. Дебаты продолжались годами, но загадка, с которой мы начали, была разгадана.

#### Блокировка выходов

На этом заканчивается история о двух черных ящиках. Опыт учит нас, однако, что не бывает мысленных экспериментов, сформулированных так четко, чтобы ни один философ не мог истолковать их ошибочно, поэтому, чтобы предупредить ряд наиболее вероятных ошибочных трактовок, я откровенно укажу на несколько важнейших деталей и объясню их роль в этом насосе интуиции.

(1) Устройства в ящиках А и Б представляют собой не что иное, как автоматические энциклопедии. Это даже не “ходячие энциклопедии”, а просто “ящики истин”. Ничто в этой истории не предполагает и не намекает, что эти устройства – сознательные, или *думающие*, *объекты*, или даже *агенты* (агентами их можно назвать разве что в крайне ограниченном смысле слова – в том же смысле, в котором агентом можно назвать термостат). Они представляют собой совершенно скучные интенциональные системы, заикленные на выполнении единственной простой задачи. (Само собой, то же самое можно сказать и о компьютере *IBM Watson*.) В них содержится большое число истинных высказываний и машины логического вывода, необходимые для генерации новых истин, а также для проверки “истинности” высказываний в сопоставлении с существующими базами данных.

(2) Поскольку две системы создавались независимо, нельзя ожидать, что в них содержатся (на самом деле или хотя бы по сути) *в точности* одинаковые истины, но чтобы фокус работал так хорошо, как я описываю в истории, нам стоит полагать, что базы данных достаточно сильно перекрывают друг друга, а потому крайне маловероятно, чтобы ящик Б не признал истинным утверждение, сгенерированное



в качестве истины ящиком А. На мой взгляд, это обеспечивают два соображения: (i) пускай Ал и Бо живут в разных странах и говорят на разных языках, они все же *существуют в одном мире*; (ii) хотя существует бесчисленное множество истинных высказываний об этом мире (нашем мире), тот факт, что в задачу Ала и Бо входило создание *полезных* баз данных, гарантирует, что две независимо разработанные системы будут в значительной степени перекрывать друг друга. Пускай Ал, возможно, знал, что в его двадцатый день рождения его левая нога была ближе к Северному полюсу, чем к Южному, а Бо не забыл, что его первого учителя французского звали Дюпоном, вряд ли они бы решили поместить эти истины в свои базы данных. Если вы сомневаетесь, что одно лишь их намерение создать пригодные для международного использования энциклопедии обеспечит столь сильное соответствие между базами данных, просто добавьте к этому грубоватую деталь и скажите, что за годы программирования они успели договориться, какие темы стоит осветить.

(3) Почему бы не взять Ала и Боба (тоже американца) или, если уж на то пошло, почему бы просто не поместить в ящик Б дубликат системы Ала? Потому что в моей истории крайне важно, что никаким простым, поддающимся обнаружению *синтаксическим* подбором пар закономерность объяснить невозможно. Именно поэтому система Бо написана на шведском лиспе – чтобы скрыть от любопытных глаз основополагающие *семантические* сходства структур данных, к которым ящик А обращается при генерации фраз, а ящик Б – при переводе этих фраз и проверке их истинности. Будучи физическими системами, компьютеры в лучшем случае должны представлять собой синтаксические движки, реагирующие непосредственно на физически измеряемые различия, а не на значения. Однако и система А, и система Б были разработаны таким образом, чтобы как можно лучше походить на воображаемого всезнайку, *семантический* движок, полный *понимаемых* истин. Поскольку две *различные* синтаксические системы, А и Б, были созданы, чтобы походить на *один и тот же* семантический движок, единственный способ объяснить демонстрируемую ими любопытную закономерность – это подняться на уровень семантического движка, где истины осознаются, а убеждения формируются намеренно. В таком случае нужно было создать две системы, которые демонстрировали бы описанную ранее восхитительную закономерность внешнего поведения, но внутренне как можно сильнее отличались бы друг от друга, чтобы объяснить наблюдаемую закономерность мог *только* тот факт, что их внутренняя структура представляет собой систематически организованные *представления об окружающем мире*. (Вспомните, что говорилось на эту тему в главе 13.)

Можно сделать паузу и спросить, могут ли две такие системы быть настолько непостижимыми, чтобы к ним нельзя было применить обратное проектирование. Иными словами, могли ли ученые так долго биться над этой загадкой? Учитывая, каких головокружительных высот достигла криптография, прежде чем давать ответ на этот вопрос, стоит подумать как минимум трижды. Я не знаю, можно ли убедительно доказать, существуют ли невзламываемые схемы шифрования. Но даже если не принимать во внимание шифрование, программисты прекрасно знают, что все удобные **комментарии** и другие обозначения, размещаемые в **исходном коде** при написании программы, исчезают при компиляции этого исходного кода, оставляя хитросплетение машинных инструкций, которые *практически* невозможно расшифровать. Иногда на практике осуществима “декомпиляция” – обратное проектирование объектного кода и восстановление исходного кода, – но она не восстанавливает комментарии, а просто переводит описание структур на язык более высокого уровня (и всегда ли она осуществима в принципе?). Мое предположение, что попытки ученых декомпилировать программу и расшифровать базы данных не принесли результатов, можно при необходимости подкрепить, сказав, что при их разработке применялось шифрование.

В моем варианте истории действительно представляется странным, что ученым не пришло в голову попытаться прочесть цепочки бит, проходящие по проводу, как послания в кодировке ASCII. Как они не догадались? Замечание вполне справедливо: этот недостаток мысленного эксперимента можно устранить, отправив все устройство (ящики А и Б и соединительный провод) на “Марс” и поручив задачу установления закономерности инопланетным ученым. Как и мы, они увидят, что все нажатия на  $\alpha$  включают красную лампочку, все нажатия на  $\beta$  включают зеленую лампочку, а случайные цепочки бит включают желтую лампочку, но об ASCII они знать не будут. В их представлении этот подарок из космоса будет демонстрировать совершенно загадочную закономерность, не поддающуюся никакой аналитической проверке, *если только* они не догадаются, что в каждом ящике содержится *описание мира* – и описывается при этом *один и тот же мир*<sup>[46]</sup>. В основе закономерности лежит тот факт, что в каждом ящике содержатся всевозможные семантические связи с одними и теми же вещами, которые описываются с применением разной “терминологии” и по-разному аксиоматизируются.

Когда в начале 1980-х я опробовал этот мысленный эксперимент на Дэнни Хиллисе – основателе компании *Thinking Machines* и создателе новаторского параллельного суперкомпьютера *Connection Machine*, – он тотчас предложил криптографическое “решение” загадки, а затем признал, что мое решение можно считать частным случаем его решения. “Ал и Бо использовали *мир* в качестве «одноразового блокнота»!” – сказал он, намекая на стандартную технику шифрования. Чтобы понять его мысль, можно представить вариацию на эту тему. Допустим, вас с другом вот-вот захватят в плен враждебные силы (скажем, космические пираты), которые говорят по-английски, но не слишком много знают о вашем мире. Вы оба знаете азбуку Морзе и придумываете импровизированную схему шифрования: тире означает истинное высказывание, точка – ложное. Захватчики слушают вашу речь: “Птицы откладывают яйца, а жабы летают. Чикаго – это город, у меня не железные ноги, а в бейсбол играют в августе”», – говорите вы, тем самым отрицательно (“по”: тире-точка; тире-тире-тире) отвечая на вопрос, заданный вашим другом. При следующем отрицательном ответе вы используете другие высказывания. Даже если ваши захватчики знают азбуку Морзе, они не сумеют понять, какие характеристики обозначаются точками и тире, если только не научатся определять истинность и ложность высказываний. Эта вариация может оживить нашу историю следующим образом: вместо того чтобы отправлять на Марс упакованные в ящики компьютерные системы, мы посадим в ящики Ала и Бо и отправим на Марс их самих. Когда Ал и Бо разыграют марсиан с использованием азбуки Морзе, те будут озадачены ничуть не меньше, чем при виде компьютеров, если только не придут к выводу (очевидному для нас, но мы ведь на марсиане), что поведение объектов в ящиках необходимо интерпретировать семантически.

Мораль сей сказки проста. **Интенциональную установку** ничем не заменить: можно либо принять ее и объяснить происходящее, обнаружив факты на семантическом уровне, либо вечно биться над загадкой закономерности – *каузальной* закономерности, – которая проявляется так очевидно<sup>[47]</sup>.

На этом перепутье, если вы похожи на многих других философов, вам опять покажется привлекательной мысль о том, что этот насос интуиции “работает” только потому, что ящики А и Б представляют собой артефакты, интенциональность которых, какой бы она ни была, полностью производна и артефактна. Структуры данных в их памяти получают ориентиры (если получают их вообще), косвенным образом опираясь на органы чувств, биографии и цели своих создателей, Ала и Бо. Реальным источником значения, или истины, или семантической этих артефактов выступают их создатели. Ал и Бо обладают исходной интенциональностью, в то время как А и Б обладают лишь производной интенциональностью. (Само собой, именно поэтому мы и предположили, что в некотором смысле Ал и Бо пребывали в созданных ими ящиках.) Я мог бы рассказать историю иначе: внутри ящиков сидели два *робота*, Ал и Бо, каждый из которых “целую жизнь” изучал этот мир и собирал факты о нем, прежде чем залезть в свой ящик. Я выбрал более простой путь, чтобы предупредить все вопросы о том, можно ли считать, что ящик А или ящик Б “действительно думает”, однако, если вы хотите пересмотреть этот мысленный эксперимент, усложнив его указанным образом, обратите внимание, что насос интуиции о **гигантском роботе** для выживания уже бросил тень сомнения на убедительную в остальном идею, что настоящая интенциональность не может возникнуть ни в одном артефакте.



## Резюме

Представленный на этих страницах двадцать один инструмент мышления – дюжина насосов интуиции и несколько полезных концепций – работает при изучении фундаментального понятия значения. Что они нам дают? Хорошо ли они работают? Обратите внимание, что насос интуиции может считаться ценным по двум критериям. Если он сконструирован хорошо, то провоцируемые им интуитивные озарения либо надежны и убедительны, что помогает избежать некоторых вероятных в ином случае ошибок, либо *все равно* сомнительны, что помогает привлечь внимание к проблемам с его предпосылками. Как бы то ни было, этот инструмент напоминает рычаг, своего рода доску-качели – когда поднимается один ее конец, второй опускается вниз, – но только если инструмент не гнется и не ломается посередине. Насосы интуиции сложнее качелей, поэтому нам приходится поворачивать различные рычаги. Именно это я только что делал, разбирая пример с **двумя черными ящиками**, но не стоит и сомневаться, что перед вынесением окончательного вердикта необходимо проверить и другие рычаги.

Каков же вердикт по этому длинному разделу о значении? Пессимистичен и оптимистичен одновременно. Значение нельзя назвать простой характеристикой, которая легко укореняется в нашем мозге, и нам не всегда удастся найти “глубинные” факты, *однозначно* отвечающие на вопрос, что *на самом деле значит* фраза, мысль или убеждение. В лучшем случае – и этого обычно достаточно – мы можем найти и закрепить (очевидно) *лучшие интерпретации* всех имеющихся у нас данных физической (и конструктивной) установки. Если мы в состоянии найти *одно* решение дилеммы Куайна о значении, скорее всего, это *единственное* ее решение – мы можем быть уверены, что лучшего решения этой дилеммы не найти. Как показал перевоз **четвертачника** в Панаму, значение всегда зависит от контекста функции и не нужна никакая *исходная* интенциональность, помимо интенциональности наших (**вроде как**) эгоистичных генов, которые черпают свою *вроде как* интенциональность из функционального контекста эволюции путем естественного отбора, а не от разумного творца, играющего роль богатого клиента, заказывающего гигантского робота. Насос интуиции о **двух черных ящиках** показывает, что если вы хотите понять и объяснить множество каузальных закономерностей мира, то вам *никуда не деться* от принятия **интенциональной установки** – при всей ее терпимости к *вроде как* убеждениям и тому подобным вещам.

Сегодня, в двадцать первом веке, изучение этих вопросов обогащается тем фактом, что впервые у нас есть способ методично и увлеченно размышлять о механизмах, обладающих *триллионами* подвижных частей – частей, в работе которых нет ничего загадочного. Благодаря Тьюрингу, теперь мы хотя бы расплывчато видим путь от грубой, ничего не осознающей материи (физическая установка) через каскад перестановок (конструктивная установка и *вроде как* значения) к принятию себя в качестве хрестоматийных верящих, познающих и осознающих субъектов (упрощенных интенциональной установкой до интенциональных систем).

Каждое из этих высказываний рождало или до сих пор рождает споры. Многие специалисты по-прежнему не согласны с некоторыми из них. Возможно, перечисление всех этих высказываний повысит их убедительную силу и мы получим эффект “критической массы”, который привлечет всех тех, кто не видел, как все они дополняют друг друга. С другой стороны, подобное перечисление может облегчить критикам задачу найти в них общую ошибку. Как бы то ни было, мы идем по пути прогресса. Быть может, многие из них окажутся лишь передергиванием фактов, которое не обнажает истину, а лишь скрывает ее. Чтобы обнаружить изъяны, вернитесь назад, покрутите рычаги каждого из инструментов и посмотрите, что из этого выйдет. По меньшей мере, обнаружение подобных изъянов тоже способствует движению вперед: так развенчиваются соблазнительные, но неверные идеи – ведь именно этим философы занимаются уже не одну тысячу лет.

Я не утверждаю, что эта часть книги дает нам *теорию значения*. Она дает нам лишь достаточно обширное логическое пространство, где должна уместиться адекватная научная теория значения, но только если я прав<sup>[48]</sup>. Теперь, когда мы более или менее разобрались с *содержимым* сознания (*mind*), можем ли мы обратиться к величайшей загадке, самосознанию (*consciousness*)? Пока нет. Нам нужно заложить другие основы. Как мы увидели в этом разделе, слишком много вопросов, возникающих при попытке понять сознание, связаны со следствиями или предпосылками эволюции. Вопросы эволюции возникали во многих поднимаемых нами темах, поэтому давайте рассмотрим их, прежде чем двигаться дальше. К тому же эта тема безмерно интересна.

## **VI.**

### **Инструменты мышления об эволюции**

На мой взгляд, никому и никогда не приходила в голову идея лучше идеи Дарвина об эволюции путем естественного отбора, поскольку эта идея одним махом объединяет материю и смысл – два аспекта реальности, которые кажутся безмерно далекими друг от друга. С одной стороны, есть мир наших сознаний и их смыслов, наших целей, надежд и стремлений, а также любимой (хотя самой избитой из всех философских тем) темы смысла жизни. С другой стороны, есть беспрестанно вращающиеся в пространстве галактики, бесцельно скользящие по своим орбитам планеты, подчиняющиеся физическим законам безжизненные химические механизмы – и все это без каких-либо мотивов и целей. Тут на сцену выходит Дарвин, который показывает, как первое проистекает из второго, рождая смыслы на ходу, и новое представление о поступательном развитии “снизу вверх” вытесняет традиционное представление о сотворении мира “сверху”». Идея естественного отбора не слишком сложна, но так основательна, что некоторые не готовы ее постичь, а потому отчаянно пытаются отвести взгляд, словно не желая плотать горькую пилюлю. Описываемые далее инструменты мышления помогут нам увидеть, как эта идея подсвечивает темные закоулки бытия, превращая тайны в загадки, разгадав которые, мы сможем как никогда хорошо рассмотреть триумфы природы.

## 34. Универсальная кислота

Слышали ли вы об универсальной кислоте? В школьные годы эта идея не на шутку увлекала нас с друзьями. Не знаю, придумали мы ее сами или унаследовали от прошлых поколений как часть подпольной молодежной культуры вместе с рассказами о чудесном эликсире с афродизиаками и удивительных свойствах селитры. Универсальная кислота настолько агрессивна, что прожигает все! Но в чем ее хранить? Она проедает стеклянные бутылки и канистры из нержавеющей стали с той же легкостью, что и бумажные пакеты. Что случится, если вы случайно найдете или создадите каплю универсальной кислоты? Уничтожит ли она в конце концов всю планету? Что она оставит после себя? Каким станет мир, после того как все трансформируется, пережив столкновение с универсальной кислотой? Тогда я еще не понимал, что через несколько лет столкнусь с идеей – идеей Дарвина, – весьма напоминающей универсальную кислоту: она прожигает почти любую традиционную концепцию и оставляет после себя коренным образом измененную картину мира, где большинство старых вех по-прежнему узнаваемо, но фундаментально перестроено<sup>[49]</sup>.

Когда в 1995 г. я предложил этот образ опасной идеи Дарвина, многие из тех, кто страшится Дарвина, его не поняли (быть может, намеренно?). Я постарался заверить читателей, что после того, как универсальная кислота просочится сквозь их любимые темы – мораль, искусство, культуру, религию, юмор и да, даже сознание, – оставшееся будет не менее, а даже во многих отношениях более чудесным, просто несколько измененным. Без сомнения, идея Дарвина революционна, однако она не уничтожает то, что мы ценим в этих вещах; она ставит их на более прочный фундамент и элегантно объединяет с остальным знанием. Веками “культура и искусство” считались не просто отделенными от наук, но в некотором роде *защищенными от* агрессивных исследований, проводимых в научной сфере, однако эта традиционная изоляция не лучший способ сохранить все то, что мы любим. Попытка спрятать наши сокровища под покровом тайны не дает нам возможности должным образом закрепить их в физическом мире. Эта ошибка достаточно типична, особенно в философии.

Чувствуя, что любимое в опасности, люди первым делом строят вокруг него “несокрушимую” стену, своего рода линию Мажино, и для надежности окружают ее несколько большую территорию, создавая буферную зону внутри укреплений. Казалось бы, это удачная, рациональная мысль. Наличие буферной зоны не дает нам свернуть на ужасную *кривую дорожку* и встать на опасное *лезвие бритвы*, ведь всем известно, что дай им палец – руку откусят. Копайте ров! Стройте стену! Охватывайте как можно большую территорию. Но такая стратегия, как правило, отягощает защитников нестабильным, непомерным (невероятным, несостоятельным) набором догм, которые невозможно защитить рациональным образом, а следовательно, в итоге приходится защищать, отчаянно хватаясь за них и громко крича. В философии такой выбор стратегии часто становится *абсолютизмом* того или иного типа: священность (человеческой) жизни *бесконечна*; в сердце великого искусства кроется божественный и необъяснимый *гений*; нам, простым смертным, не дано постичь проблему сознания; и не стоит забывать также об одной из моих любимых мишеней для критики – я называю ее *истерическим реализмом*, – которая пласит, что всегда существуют глубинные факты, способные *однозначно решать* загадки значения. Эти факты реальны, *действительно* реальны, даже если у нас никак не получается их обнаружить. В некотором роде эта мысль убедительна, ведь она вызывает к простой человеческой скромности. Кто *мы* такие, чтобы говорить, что не существует фактов, способных *однозначно решать* эти загадки? Привлекательность этой мысли прекрасно подтверждает знаменитое неприятие неопределенности квантовой физики Эйнштейном. “Бог не играет в кости!” – в сердцах возразил он, хотя такое возражение и не имеет под собой оснований. Если уж на то пошло, кто *мы* такие – или кто такой Эйнштейн, – чтобы говорить, что Бог не играет в кости? Вероятно, к этим вопросам нужно подходить иначе, и далее мы увидим несколько возможностей для обращения к истерическому реализму и узнаем, как ему противостоять. Прекрасным противовесом служит эволюционное мышление.

## 35. Библиотека Менделя: чрезвычайно большое и исчезающе малое

Крейг Вентер и другие ученые уже секвенировали *геном человека*, но что это значит? Разве ДНК каждого человека не уникальна? Да, уникальна, причем настолько, что даже маленького фрагмента ДНК, обнаруженного на месте преступления, достаточно, чтобы идентифицировать преступника с вероятностью 99 процентов. И в то же время ДНК людей настолько похожи, что ученые могут отличать их от ДНК других видов, имея лишь фрагменты полного генома. Как это возможно? Как ДНК людей могут быть настолько уникальными и в то же время настолько похожими? Чтобы понять этот удивительный факт, можно сравнить ДНК с текстами книг. Аргентинский писатель Хорхе Луис Борхес (1962) сочинил для нас рассказ “Вавилонская библиотека”, в котором наглядно иллюстрируется возможность сосуществования столь разительных различий и сходств. Борхес описывает тщетные исследования и размышления людей, которые живут в огромном хранилище книг, структурированном на манер пчелиного улья: книжные полки тянутся по стенам тысяч (или миллионов, или миллиардов) соединенных коридорами шестигранных галерей. Стоя у перил вентиляционного колодца, никто не может разглядеть ни пола, ни потолка. И никто никогда не находил галерею, которая не была бы окружена шестью другими. Люди гадают, бесконечно ли это хранилище. В конце концов они решают, что это не так, но оно может быть и бесконечным, поскольку на полках, похоже, стоят *все возможные книги* – увы, без какого-либо порядка.

Допустим, в каждой книге пятьсот страниц, на каждой странице сорок строк, а в каждой строке пятьдесят знаков, то есть на странице умещается две тысячи знаков. Каждый знак – это либо пробел, либо один из ста символов (прописные и строчные буквы английского и других европейских языков, пробел и знаки препинания)<sup>[50]</sup>. Где-то в Вавилонской библиотеке есть книга, все страницы которой абсолютно чисты, а где-то есть книга, все страницы которой заполнены вопросительными знаками, но в подавляющем большинстве книг содержится абракадабра – в библиотеку попадают абсолютно любые книги, без оглядки на орфографию и пунктуацию, не говоря уже о смысле. Пятьсот страниц, помноженные на две тысячи знаков на страницу, дают нам миллион знаков на книгу, так что, если мы решим наполнить книги всеми возможными комбинациями знаков, в Вавилонской библиотеке окажется  $100^{1\ 000\ 000}$  различных книг. По оценкам<sup>[51]</sup>, в наблюдаемой нами части вселенной содержится всего (плюс-минус)  $100^{40}$  частиц (протонов, нейтронов и электронов), а следовательно, существование Вавилонской библиотеки физически совершенно невозможно, но благодаря четким правилам, в соответствии с которыми Борхес построил ее в своем воображении, мы можем с полной ясностью о ней размышлять.

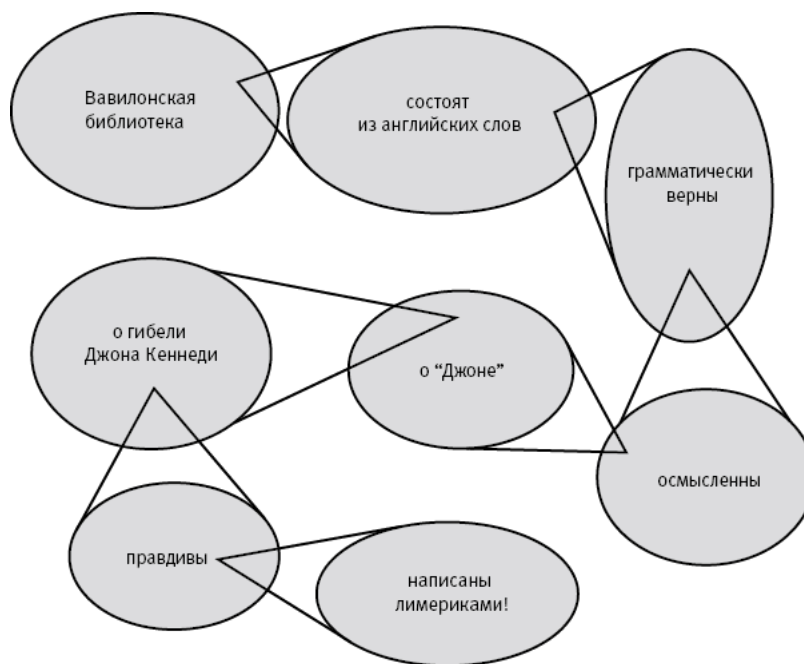
Правда ли в ней хранятся *все возможные книги*? Очевидно нет, поскольку они должны быть напечатаны с использованием “только” ста различных символов, исключая, как можно полагать, символы греческого, русского, арабского и китайского алфавитов, из-за чего в библиотеке нет огромного количества важнейших *настоящих* книг. Само собой, в библиотеке есть их блестящие переводы на английский, французский, немецкий, итальянский и другие языки, а также бесконечное число низкопробных переводов каждой из книг. При этом в библиотеке есть книги, в которых более пятисот страниц, но их текст начинается в одном томе и продолжается в других без перерыва.

Некоторые книги, хранящиеся в Вавилонской библиотеке, могут быть весьма любопытны. Среди них ваша лучшая, самая точная 500-страничная биография, в которой ваша жизнь описывается с рождения и до самой смерти. Обнаружить ее, однако, будет практически невозможно, поскольку в библиотеке также хранится огромное множество книг, в которых ваша биография предельно точно излагается до десятого, двадцатого, тридцатого, сорокового и т. д. дня рождения, но последующие события вашей жизни описываются совершенно неправильно – огромным множеством нетривиальных способов. Но крайне маловероятно, что в этом огромном книгохранилище вообще удастся обнаружить хотя бы одну читабельную книгу.

Нам нужно ввести несколько терминов для фигурирующих здесь величин. Вавилонская библиотека не бесконечна, так что шанс найти в ней что-нибудь интересное в буквальном смысле не бесконечно мал<sup>[52]</sup>. Эти слова преувеличивают ситуацию знакомым нам образом, но их стоит избегать. К несчастью, ни одна стандартная метафора – астрономические числа, иголка в стоге сена, капля в море – не справляется с описанием этой библиотеки. Ни одну *настоящую* астрономическую величину (такую, как число элементарных частиц во вселенной или количество наносекунд, прошедших с момента Большого взрыва) даже не разглядеть на фоне этих огромных, но конечных чисел. Если бы найти

читабельную книгу в Вавилонской библиотеке было столь же просто, как найти конкретную каплю в море, мы бы и в ус не дули! Но если нас забросят в случайную галерею библиотеки, наши шансы обнаружить книгу, в которой будет хотя бы одна грамматически верная фраза, настолько исчезающе малы, что можно даже написать это понятие с большой буквы “И”, а также ввести сопутствующее понятие “Чрезвычайно велики”, означающее “гораздо больше астрономических”<sup>[53]</sup>.

Вот другой способ осознать, как невероятно велика Вавилонская библиотека. Как мы только что заметили, лишь Исчезающе малое подмножество книг составлено из английских слов. Само по себе это подмножество Чрезвычайно велико, а в Исчезающе малое подмножество внутри него входят книги, где слова выстраиваются в грамматически верные фразы (Чрезвычайно большое подмножество при этом составляют книги, содержащие цепочки слов наподобие “хороший из Париж помощь легка с от который тем не менее демократия стриптиз тигры”.) Чрезвычайно большое, но Исчезающе малое подмножество грамматически верных книг составляют книги, где фразы последовательно связаны друг с другом (в остальных книгах содержатся фразы, которые, возможно, случайным образом выбраны из книг на грамматически верном английском.) Чрезвычайно большое, но Исчезающе малое подмножество этих осмысленных книг составляют книги о человеке по имени Джон, а Чрезвычайно большое, но Исчезающе малое подмножество *этих* книг повествует об убийстве Джона Ф. Кеннеди, и все еще Чрезвычайно большое (но Исчезающе малое) подмножество *этих* книг говорит правду – и Чрезвычайно большое, но Исчезающе малое подмножество этих правдивых книг об убийстве Кеннеди написано лимериками! Да, количество *возможных* правдивых книг о гибели Кеннеди, написанных лимериками, превосходит число книг в Библиотеке Конгресса! Вероятнее всего, ни одна из них не была опубликована, но это и к лучшему.



В Вавилонской библиотеке хранится экземпляр “Моби Дика”, но в ней хранится и 100 000 000 экземпляров-мутантов, которые отличаются от канонического “Моби Дика” *одной-единственной* опечаткой. Это еще не Чрезвычайно большое число, но оно стремительно растет, когда мы прибавляем экземпляры, отличающиеся двумя опечатками, десятью или тысячей. Даже экземпляр с тысячей опечаток – в среднем по две на страницу – будет безошибочно распознаваться как “Моби Дик”, а количество таких экземпляров чрезвычайно велико. Неважно, какой из них вы найдете, если найдете хотя бы один из них! Во всех них содержится одна и та же история, не считая поистине незначительных – почти незаметных – отличий, и читать почти любой из них будет наслаждением. И все же не любой. Порой единственная опечатка в важном месте может стать роковой. Еще один склонный к философствованиям писатель Питер Де Врис однажды опубликовал роман<sup>[54]</sup>, начинавшийся словами:

“Зовите меня, Измаил”.

На что способна единственная запятая! Представьте другие мутировавшие экземпляры, начинающиеся словами “Ловите меня Измаил”.

В рассказе Борхеса книги стоят на полках беспорядочно, но даже если бы они оказались выстроены по алфавиту, мы столкнулись бы с неразрешимыми проблемами при поиске той самой книги, которую мы ищем (к примеру, “базовой” версии “Моби Дика”). Представьте, что вы летите на звездолете по галактике Моби Дика в Вавилонской библиотеке. Эта галактика сама по себе Чрезвычайно больше всей физической вселенной, так что, куда бы вы ни отправились, даже если вы летите со скоростью света, веками вам будут попадаться лишь практически неотличимые друг от друга копии “Моби Дика”. Вы ни разу не наткнетесь ни на что другое. “Дэвид Копперфилд” в этом пространстве невообразимо далек, даже если нам известно, что есть путь – кратчайший путь, не считая огромного множества других путей, – ведущий от одной великой книги к другой посредством единичных типографских изменений. (Оказавшись на этом пути и оглядевшись, вы вряд ли сумеете сказать, в какую сторону двигаться к “Дэвиду Копперфилду”», даже если будете держать в руках обе книги.)

Иными словами, это *логическое* пространство так Чрезмерно велико, что многие наши обычные представления о местоположении, о поиске и обнаружении искомого и прочих подобных простых и практических действиях не имеют в нем непосредственного применения. Борхес расставил книги на полках в произвольном порядке и сделал на основании этого несколько превосходных наблюдений, но только представьте, с каким количеством проблем он столкнулся бы, если бы попробовал расставить их по алфавиту. Поскольку (в нашей версии) всего сто различных алфавитных символов, можно считать определенную их последовательность алфавитным порядком, например a, A, b, B, c, C... z, Z,?!), (, %... à, â, è, ê, é... Затем можно разместить все книги, которые начинаются на один и тот же символ, на одном *этаже*. Теперь в нашей библиотеке всего сто этажей, а это меньше, чем в Сирс-тауэр (или Уиллис-тауэр) в Чикаго. Каждый этаж можно разделить на сто коридоров, в каждом из которых мы разместим книги, второй символ которых – следующий по алфавиту. В каждом коридоре можно разместить сто полок, по одной для каждого из третьих символов. Таким образом, все книги, которые начинаются со слов *aardvarks love Mozart* – и сколько таких книг! – окажутся на одной полке (полке “t”) в первом коридоре на первом этаже. Но эта полка огромна, поэтому нам, возможно, лучше разместить книги в картотечных ящиках, стоящих под прямыми углами к полке, по одному ящику для каждого четвертого символа. В таком случае каждая полка будет, скажем, не более сотни футов длиной. Но теперь ужасно пубокими станут картотечные ящики, которые будут упираться в ящики соседнего коридора, поэтому... но у нас закончились измерения для расстановки книг. Чтобы расставить все книги аккуратно, нам нужно пространство с миллионом измерений, каждое из которых будет располагаться “под прямым углом” ко всем другим. Такие пространства называются гиперпространствами – о них можно помыслить, но визуализировать их нельзя. Ученые постоянно используют их, чтобы представить свои теории. Геометрия таких пространств (даже если считать их только воображаемыми) стабильна и хорошо изучена математиками. В рамках этих логических пространств можно с уверенностью говорить о местоположениях, путях, траекториях, объемах (гиперобъемах), расстояниях и направлениях.

Теперь мы готовы рассмотреть вариацию на тему Борхеса, которую я назову библиотекой Менделя. В этой библиотеке хранятся “все возможные геномы” – последовательности ДНК. В книге “Слепой часовщик” (1986) Ричард Докинз описывает подобное пространство, которое называет Страной Биоморфов. Меня вдохновили именно его рассуждения, и наши представления полностью совместимы, однако я хочу подчеркнуть некоторые аспекты, на которых он не стал заострять внимание.

Если считать, что в библиотеке Менделя хранятся *описания* геномов, то она представляет собой часть Вавилонской библиотеки. Стандартный код описания ДНК включает в себя всего четыре символа, A, C, G и T (обозначающие аденин, цитозин, гуанин и тимин – четыре типа нуклеотидов). Следовательно, все 500-страничные вариации последовательностей этих четырех букв уже хранятся в Вавилонской библиотеке. Типичные геномы, однако, гораздо длиннее обычных книг. В человеческий геном входит примерно три миллиарда нуклеотидов, так что исчерпывающее описание генома одного человека – например, вашего – заняло бы примерно три тысячи 500-страничных книг из Вавилонской библиотеки.

Сравнение человеческого генома с объемами галактики “Моби Дика” помогает нам объяснить различие и сходство человеческих геномов. Как можно говорить о секвенировании (копировании) генома человека *в целом*, если все геномы человека отличаются друг от друга не в одном, а в тысячах мест (*локусов* на языке генетики)? Подобно снежинкам или отпечаткам пальцев, два подлинных человеческих генома не могут быть в точности одинаковыми, включая даже геномы идентичных близнецов (случайные ошибки могут вкрасться даже в клетки одного человека). Человеческая ДНК легко отличима от ДНК любого другого вида, даже от ДНК шимпанзе, локусы которой совпадают с человеческой на 90 процентов.

Каждый подлинный геном человека, существовавший когда-либо, содержится в галактике всех возможных человеческих геномов, которая находится на Чрезвычайно большом расстоянии от галактик геномов других видов, и все же внутри галактики достаточно места, чтобы никакие два генома человека не были одинаковыми. У вас есть две версии каждого из ваших генов – от матери и от отца. Они передали вам ровно половину собственных генов, случайным образом выбранных из набора, который они получили от своих родителей, ваших бабушек и дедушек, но поскольку ваши бабушки и дедушки тоже принадлежали к виду *Homo sapiens*, их геномы совпадают почти в каждом локусе, так что в большинстве случаев неважно, кто из бабушек и дедушек предоставил тот или иной ваш ген. Тем не менее их геномы различаются во многих тысячах локусов, и здесь в игру вступает случай – механизм случайного выбора определяет, как каждый из ваших родителей поучаствует в формировании вашей ДНК. Более того, скорость накопления мутаций у млекопитающих составляет примерно сто штук на геном на одно поколение. “Таким образом, сотня отличий от вас и вашего супруга возникнет в генах ваших детей в результате случайных ошибок копирования, совершенных энзимами, или в результате мутаций ваших яйцников или яиц, вызванных космическими лучами” (Ridley 1993, p. 45).

Описание генома лошади, капусты или осьминога будет состоять из тех же букв, А, С, G и Т. Большинство геномов животных короче генома человека, но геномы некоторых растений длиннее человеческих более чем в десять раз, а геномы ряда одноклеточных амёб и того длиннее! Текущий мировой рекорд принадлежит *Amoeba dubia*, в геноме которой содержится около 670 миллиардов спаренных оснований, что более чем в двести раз превышает количество спаренных оснований в геноме человека. Однако допустим – своевольно, – что в библиотеке Менделя содержатся все цепочки ДНК, описанные в 3000-томных изданиях, включающих только четыре упомянутых символа. Этого количества “возможных” геномов будет достаточно для любых серьезных теоретических целей.

Я преувеличил, сказав, что в библиотеке Менделя содержатся “все возможные геномы”. Подобно тому как в Вавилонской библиотеке не нашлось места для русского и китайского языков, библиотека Менделя игнорирует (очевидную) возможность существования альтернативных генетических алфавитов, например основанных на других химических составляющих. Таким образом, любые выводы относительно того, что возможно в *этой* библиотеке Менделя, придется пересмотреть, когда мы попробуем применить их к более широкому представлению о возможном. Это скорее сильная, а не слабая сторона нашей тактики, поскольку она позволяет нам внимательно следить за тем, о какой именно скромной, ограниченной возможности идет речь.

Одна из важных характеристик ДНК – примерно одинаковая химическая стабильность любых пермутаций последовательностей аденина, цитозина, гуанина и тимина. В принципе, любая из них может быть сконструирована в сплайсинговой лаборатории и впоследствии храниться неопределенное количество времени, как книга в библиотеке. Но не каждая такая последовательность из библиотеки Менделя соответствует жизнеспособному организму. Большинство последовательностей ДНК – и это большинство Чрезвычайно велико – представляет собой абракадабру, рецепты, по которым не создать живой организм. Все наблюдаемые геномы, в действительности существующие сегодня, родились в результате миллиардов лет поправок и пересмотров в ходе бездумного редактирования, эффективность которого достигается за счет того, что большая часть абракадабры (за исключением Исчезающе малого множества осмысленного, примененного “текста”) автоматически отбрасывалась, в то время как остальное без конца использовалось снова и снова, подвергаясь копированию огромное множество раз. В вашем теле в эту минуту содержится более триллиона копий вашего генома, одна на каждую человеческую клетку, и каждый день по мере возникновения новых клеток кожи, кости и крови в них внедряются новые копии вашего генома. Текст, который *можно* скопировать – потому что он содержится в действующем механизме, живой клетке, – копируется. Остальное исчезает. Публикуй или погибнешь.



## 36. Гены как слова и как подпрограммы

Как мы только что увидели, гены удобно сравнивать со словами, но можно провести и лучшую аналогию, которую нам теперь под силу понять, благодаря интерлюдии о компьютерах. В своем шедевре “Рассказ предка” (2004) Ричард Докинз отдает должное другому блестящему писателю на тему эволюции, Мэтту Ридли, который продемонстрировал серьезное сходство между генами и компьютерными подпрограммами в книге “Природа через воспитание”. Я не стал бы использовать столь длинную цитату из чужой книги в своей, но, как выяснилось, все мои попытки перефразировать этот фрагмент неизбежно приносили в жертву капле оригинальности ясность и живость слога, поэтому, заручившись согласием Докинза, я привожу здесь фрагмент его книги без каких-либо изменений.

Большая часть секвенируемого генома представляет собой не книгу инструкций и не координирующую компьютерную программу для создания человека или мыши, но некоторые его фрагменты действительно таковы. В ином случае действительно можно было бы ожидать, что наша программа будет больше программы мыши. Но большая часть генома похожа скорее на словарь слов, доступных для написания книги инструкций, – или, как мы скоро увидим, набор подпрограмм, которые запускаются координирующей программой. Как отмечает Ридли, списки слов “Дэвида Копперфилда” и “Над пропастью во ржи” почти идентичны. Оба они основаны на словаре образованного человека, для которого английский – родной язык. Разница между книгами объясняется порядком, в котором выстроены эти слова.

При создании человека или при создании мыши эмбриология пользуется одним словарем генов: стандартным словарем эмбриологий млекопитающих. Разница между человеком и мышью объясняется разным порядком использования генов, взятых из этого общего словаря млекопитающих, разными частями тела, где это происходит, и разной временной привязкой. Все это контролируется конкретными генами, которые отвечают за активацию других генов, в рамках сложных и точно хронометрированных каскадов. Однако эти контролирующие гены составляют лишь меньшую часть генов генома.

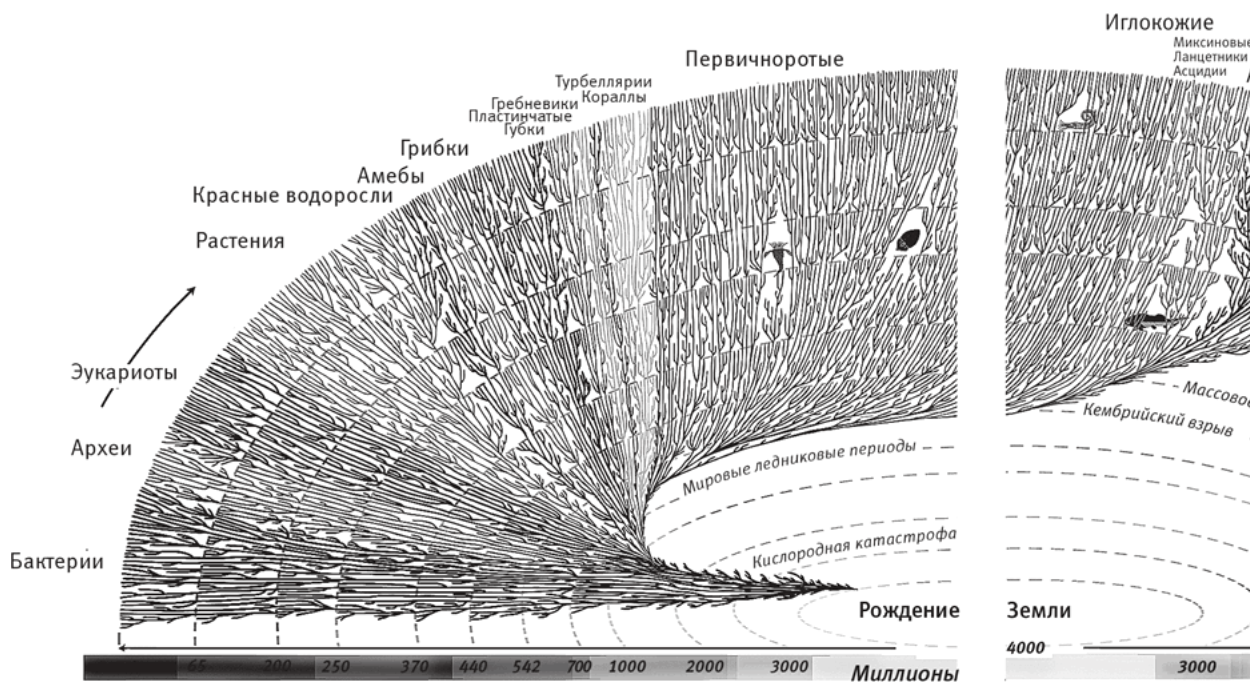
Не думайте, что “порядок” означает порядок расстановки генов в хромосомах. Не считая нескольких исключений... порядок генов в хромосоме столь же произволен, сколь произволен порядок слов в словаре – как правило, слова выстраиваются по алфавиту, однако, особенно в разговорниках для путешествующих, они могут выстраиваться и в порядке использования: слова, полезные в аэропорту; слова, полезные при визите к врачу; слова, полезные в магазине, и так далее. Неважно, в каком порядке гены хранятся в хромосоме. Важно то, что клеточный механизм при необходимости находит нужный ген – и делает это, применяя методы, которые мы понимаем все лучше...

В одном отношении аналогия со словами обманчива. Слова короче генов – некоторые авторы приравнивают ген к предложению. Однако аналогия с предложениями не слишком хороша по другой причине. Различные книги состояются не посредством пермутации фиксированного набора предложений. Большинство предложений уникально. Гены же, как и слова (но не как предложения), используются снова и снова в различных контекстах. Лучше сравнивать ген не со словом и не с предложением, а с инструментальной подпрограммой компьютера...

Инструментарий подпрограмм компьютера *Apple* хранится в его ПЗУ (постоянном запоминающем устройстве) или в системных файлах, постоянно загружаемых при запуске. Существуют тысячи инструментальных подпрограмм, и каждая из них выполняет конкретную операцию, которую с большой вероятностью приходится снова и снова повторять немного по-разному в разных программах. К примеру, инструментальная подпрограмма *ObscureCursor* скрывает курсор с экрана до следующего перемещения компьютерной мышки. Хоть вы этого и не видите, “ген” *ObscureCursor* активируется всякий раз, когда вы начинаете печатать и курсор исчезает с экрана. Инструментальные подпрограммы объясняют знакомые свойства, характерные для всех программ компьютеров *Apple* (и эквивалентных программ компьютеров *Windows*): выпадающие списки, полосы прокрутки, окна, которые можно растягивать и сжимать, а также перетаскивать по экрану с помощью мышки, и многое другое. [Dawkins 2004, pp. 155–156]



Все это помогает нам понять, почему специалисту так просто распознать геном млекопитающего. В нем содержится инструментарий млекопитающего, в который, помимо специальных инструментов для создания млекопитающего, входят также инструменты для создания рептилий, рыб и даже червей. Самыми старыми инструментами из этого набора располагают все живые организмы, включая бактерии.



На этой схеме показаны все основные и многие второстепенные живые ветви

## 37. Древо жизни

Генеалогия существующих сегодня геномов восходит к родителям, бабушкам и дедушкам и так далее до самого зарождения жизни на Земле. На этом изображении древа жизни показано, как *каждый* человек состоит в относительно близком родстве с любым другим человеком, имея общих предков с любым человеком на горизонте последней сотни тысяч лет, а также имея общих предков с любой собакой и китом на горизонте последних двух сотен миллионов лет и с каждой маргариткой и секвойей на горизонте последних двух миллиардов лет.

Есть много способов изобразить древо жизни. На этой схеме в настоящее устремлены внешние концы его ветвей. Внешней границы достигают только те линии наследования, которые до сих пор живы. Показано, что динозавры (за исключением птиц, которые произошли от них) вымерли более 60 миллионов лет назад. Все ветви в итоге соединяются у истоков жизни, но подробнее об этом позже. Если увеличить эту схему в триллион раз, мы увидим родословное древо каждой мухи, рыбы и лягушки, которые когда-либо жили на земле, и сможем рассмотреть, какие из них умерли бездетными (само собой, таких большинство), а какие оставили потомство.

## 38. Подъем в проектное пространство: небесные крюки и подъемные краны

Жизнь удивительна. Когда задумываешься о миллиардах солнечных систем, которые почти наверняка лишены жизни, невольно удивляешься, что в мире вообще существует хоть одно живое существо. Удивляет и количество жизненных форм – от бактерий до рыб, птиц, маргариток, китов, змей, кленов и людей. Но сильнее всего, пожалуй, удивляет жизненная сила всех этих организмов, которые находят тысячи способов цепляться за жизнь и размножаться, выживать, несмотря на существенные препятствия, используя миллионы хитроумных приспособлений и схем, от витиеватых каскадов белковых машин в каждой клетке до эхолокации летучих мышей, слоновьего хобота, способности нашего мозга размышлять на любую тему на свете и многого другого. Появление этих поразительных механизмов выполнения конкретных задач требует объяснения, поскольку такое не происходит по случаю или чистому совпадению. Объяснений у нас только два: теория разумного замысла или эволюция путем естественного отбора. Какое бы из них ни оказалось верным, проводится огромная проектная работа, которую либо чудесным образом осуществляет высший разум, либо методичным, недальновидным, скучным, но совершенно не чудесным образом проводит естественный отбор. Это проектирование можно назвать НИОКР (стандартная аббревиатура научно-исследовательских и опытно-конструкторских работ), а НИОКР всегда требует средств. Они отнимают время и силы. Прелесть великой идеи Дарвина в том, что он увидел, каким образом миллиарды лет работы и чудовищное количество “лишних движений” (миллионы миллиардов приводящих к ошибкам проб) могут привести к автоматическому конструктивному совершенствованию системы без вмешательства высших сил, без конкретного намерения, без всяких ожиданий и без четкого понимания. Эту мысль красноречиво выразил один из самых непримиримых критиков Дарвина Роберт Беверли МакКензи:

В теории, с которой нам приходится иметь дело, роль создателя играет абсолютное невежество, а следовательно, фундаментальный принцип всей системы может быть сформулирован следующим образом: ЧТОБЫ СОЗДАТЬ СОВЕРШЕННУЮ И ПРЕКРАСНУЮ МАШИНУ, НЕ ОБЯЗАТЕЛЬНО ЗНАТЬ, КАК ИМЕННО ЕЕ ДЕЛАТЬ. При ближайшем рассмотрении это утверждение в сжатой форме выражает суть всей теории и в нескольких словах объясняет постулаты мистера Дарвина, который, следуя странной логике, похоже, считает, что абсолютное невежество вполне способно занять место абсолютной мудрости во всех достижениях творческого дара. [1868]

Именно так. Как только Дарвин предложил эту потрясающую идею, скептики принялись гадать, достаточно ли прошло времени, чтобы вся эта созидательная работа была осуществлена методичным образом. Удобно представить необходимую проектную работу в качестве *подъема* в **проектное пространство**. Что такое проектное пространство? Как и в случае с Вавилонской библиотекой и библиотекой Менделя, лучше всего представить его многомерным. Фактически проектное пространство содержит в себе обе эти библиотеки и гораздо больше, поскольку оно включает в себя не только все (спроектированные, написанные) книги и все (спроектированные, развившиеся) организмы, но и все остальные объекты, прекрасно описываемые конструктивной установкой (см. главу 18), такие как дома, мышеловки, боевые топоры, компьютеры и звездолеты. Подобно тому как большая часть Вавилонской библиотеки представляет собой тарабарщину, большинство областей проектного пространства заполнено мусором, объектами, которые вообще ни на что не годны. Если мы с вами похожи, представить вы можете лишь три измерения зараз, однако чем больше вы играете с этой идеей в своем воображении, тем проще вам думать, что знакомые три измерения заменяют собой целое множество. (Этот инструмент мышления оттачивается с опытом.)

Представив древо жизни помещенным в библиотеку Менделя, мы получаем возможность разглядеть, что все живые организмы, когда-либо существовавшие на этой планете, связаны друг с другом линиями родства. Эти линии показывают, как естественный отбор внедрял конструктивные усовершенствования, сохраняя достижения самых ранних НИОКР для использования во всех последующих живых организмах. (“Механизм” каждой клетки, от бактерий до клеток мозга, включает в себя тысячи блестяще спроектированных наноустройств, работающих более трех миллиардов лет, и представляет собой фундаментальный моторный отсек жизни, который мы делим с деревьями, птицами и дрожжевыми

клетками.) На многоклеточном уровне мы находим сердца, легкие, глаза, руки, ноги и крылья, проекты которых впервые были разработаны “всего” два миллиарда лет назад и с тех пор повторяются и совершенствуются снова и снова. Однако, помимо отдельных элементов организмов, существуют и продукты деятельности этих организмов, например паутины, гнезда птиц и плотины бобров. Они также демонстрируют неоспоримые свидетельства НИОКР, причем плоды этих НИОКР каким-то образом передаются по тем же самым линиям родства – либо генетически, либо через копирование потомством поведения родителей. Время от времени появляется нечто новое – посредством мутации, в результате эксперимента или по случайности. Если новое способствует совершенствованию организма, оно подвергается многократному копированию. Результаты провалившихся экспериментов искореняются. Здесь опять действует принцип “публикуй или погибнешь”.

Существуют и продукты человеческой деятельности: пахотные орудия, мосты, соборы, картины, пьесы, стихи, печатные станки, самолеты, компьютеры, газонокосилки и... инструменты мышления. Что насчет них? Разве они не висят на ветках той части древа жизни, которая описывает происхождение человека? Каждый из этих продуктов обязан своим существованием как минимум одному автору или изобретателю, но к НИОКР по созданию большинства из них руку приложили тысячи и миллионы людей. Бетховену не пришлось изобретать жанр симфонии, который существовал и до него, а Шекспиру не пришлось изобретать сонет. Цепная пила состоит из десятков и даже сотен “готовых” элементов – уже изобретенных и оптимизированных. Некоторые продукты человеческой деятельности, вероятно, скопированы с продуктов деятельности других животных. Вдохновила ли конструкция гнезда ткачика появление ткачества? Был ли приподнятый пол эскимосского иглу, который позволяет выпускать холодный воздух через утопленный вход, скопирован с подобного приподнятого пола в берлогах полярных медведей – или же они появились независимо друг от друга? (Или это полярные медведи скопировали все у эскимосов?)

Поразительное сходство принципов работы сердец и насосов, эхолокации летучих мышей и сонара и радара, плотин бобров и ирригационных плотин, глаз и фотокамер не случайно, ведь все эти продукты появились и столетиями совершенствовались с помощью подобных исследований, то есть НИОКР. Нужно поместить их в единое пространство – проектное пространство, пространство *всех возможных* проектов.

Этот шаг вызывает противоречия среди биологов. Причины этого мне, кажется, понятны, хоть они меня и расстраивают. Многие биологи всячески противятся идее “проектирования” живых организмов, поскольку полагают, что она оправдывает и поддерживает сторонников теории разумного замысла, этой фиктивной, псевдонаучной, крипторелигиозной кампании по подрыву во всех смыслах заслуженного авторитета эволюционной биологии. (Среди этих противников и двое моих уважаемых коллег и друзей, специалисты по эволюционной биологии Ричард Докинз и Джерри Койн.) Рассуждая о наличии цели и проектирования в природе, мы (очевидно) даем фору сторонникам теории разумного замысла, в связи с чем некоторые полагают, что лучше неумолимо поддерживать эмбарго на подобные темы и настаивать, что, *строго говоря*, в биосфере ничего не проектируется, если только проектирование не осуществляет человек. По их мнению, природа создает сложнейшие системы (органы, закономерности поведения) совсем не так, как это делают технические специалисты, а потому нам не следует использовать для описания этих процессов тот же самый язык. Так, Ричард Докинз говорит (при случае – например, 1996, р. 4) о *кажущихся спроектированными* чертах организмов, а в книге “Рассказ предка” (2004, р. 457) замечает: “Создаваемая дарвиновским естественным отбором иллюзия проектирования поразительно убедительна”. Я не согласен с его суровой политикой, которая может привести к катастрофическим последствиям. Недавно в баре я стал свидетелем беседы двух юношей, которые рассуждали о чудесных наномашинках, обнаруженных в каждой клетке. “Как вообще можно верить в эволюцию, наблюдая, как работают все эти фантастические маленькие роботы!” – воскликнул один из них, и второй со знанием дела кивнул. Почему-то у этих ребят сложилось впечатление, будто специалисты по эволюционной биологии полагают, что жизнь не так уж сложна, а ее компоненты не так уж замечательны. Такой скептический взгляд на эволюцию высказывали не какие-то неотесанные работяги – ребята были студентами-медиками из Гарварда! Они существенно недооценили силу естественного отбора, поскольку специалисты по эволюционной биологии снова и снова повторяли им, что в природе нет настоящих проектов, а есть лишь их *иллюзии*. Этот эпизод подсказал мне, что в “общественном сознании” постепенно укореняется ошибочная мысль, будто специалисты по эволюционной биологии не хотят “признать” или “допускать” очевидное наличие проектирования в природе.

Здесь стоит вспомнить католического архиепископа Вены Кристофа Шёнборна, которого одурачили сторонники теории разумного замысла. Он опубликовал в газете *The New York Times* статью под заголовком “В поисках замысла в природе”, где откровенно заявил:

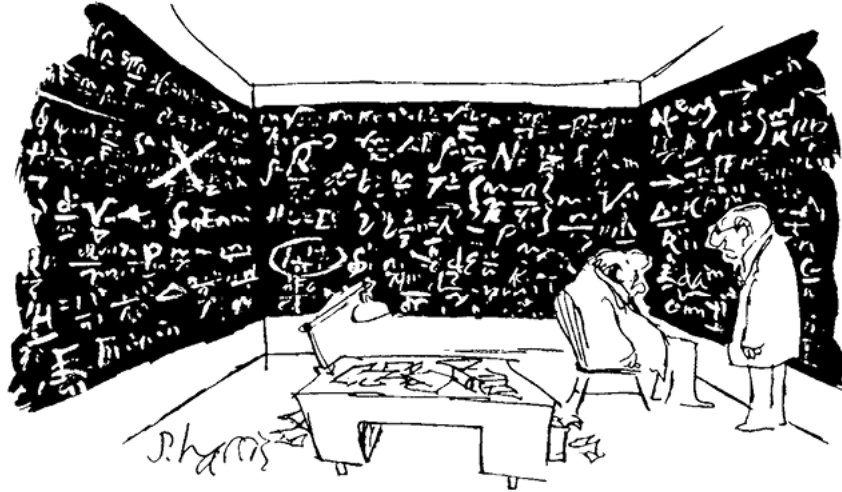
Католическая церковь отдает на откуп ученым описание подробной истории жизни на земле, однако подчеркивает, что свет разума дает человеческому интеллекту способность отчетливо разглядеть цель и замысел в мире природы, включая мир живых организмов. Возможно, эволюция в смысле общего происхождения видов действительно имела место, однако эволюции в неodarвиновском смысле, как неконтролируемого, незапланированного процесса случайных изменений и естественного отбора, никогда не было. Любая система взглядов, которая отрицает или пытается списать со счетов бесконечное множество свидетельств существования проектного замысла в биологии, является идеологией, а не наукой.

Какую кампанию хотим возглавить мы, эволюционисты? Хотим ли мы убедить обывателей, что на самом деле они не видят проектный замысел, который в биологии поразительно очевиден на любом уровне, или же мы предпочтем показать им, что, как это ни странно, Дарвин продемонстрировал возможность существования замысла – самого что ни на есть настоящего проектирования – в отсутствие высшего разума? Мы убедили мир, что земля вращается вокруг солнца и что время относительно, а не абсолютно. Зачем же чураться педагогической задачи показать, что проект может существовать и без проектировщика? Итак, здесь я отстаиваю (опять же с новым акцентом) следующее утверждение:

*Биосфера предельно насыщена проектным замыслом, смыслом и причинами.* То, что я называю конструктивной установкой, предсказывает и объясняет черты живого мира, используя те же самые допущения, которые прекрасно работают, когда мы разбираем инженерные продукты, созданные (в некотором роде) разумными проектировщиками-людьми. Эволюция путем естественного отбора представляет собой совокупность процессов, которые “находят” и “отслеживают” причины конкретной организации систем. Основное различие между причинами, которые находит эволюция, и причинами, которые находят люди-проектировщики, заключается в том, что последние обычно (но не всегда) формулируются в умах проектировщиков, в то время как причины, выявляемые естественным отбором, обычно впервые формулируются исследователями-людьми, преуспевающими в обратной разработке творений природы. Иначе говоря, люди-проектировщики *думают о* причинах наделения своих продуктов конкретными чертами, а потому *предлагают идеи*, отражающие эти причины. Они обычно замечают, оценивают, формулируют, оттачивают, а затем озвучивают, обсуждают и критикуют причины появления своих проектов. Эволюция ничем подобным не занимается; она лишь бездумно просеивает все создаваемые варианты и копирует удачные экземпляры (которые признаются удачными по причинам, не предполагающим обдумывание и формулировку со стороны естественного отбора).

Когда объекты лишь *кажутся спроектированными*, вполне логично называть их именно так, однако этот термин не прижился в биологии. Размышляя об иллюзии проектирования, я представляю себе, как карикатуристы изображают ученых: обычно бородатые мужи стоят перед доской, исписанной (на самом деле бессмысленными) символами, или в химической лаборатории, забитой целым лесом пробирок и колб, а иногда сумасшедший изобретатель работает над своей машиной времени, которая так и сверкает циферблатами, антеннами и прочими атрибутами высоких технологий. Все эти штуки ничего не делают, они лишь кажутся функциональными. Но проекты природы поистине эффективны. На самом деле они зачастую демонстративно эффективнее и мощнее любого из спроектированных человеком изобретений.

Пример надписи, которая кажется спроектированной:



*“Что случилось с изящными решениями?”*

“Что случилось с изящными решениями?”

Здоровому молодому человеку под силу около недели идти, неся с собой весь запас необходимых ему для выживания воды и пищи и по собственному усмотрению делая остановки и привалы. За это время он преодолеет примерно 240 километров. (Тяжелее всего нести воду: рюкзак с двадцатью килограммами воды, семью килограммами провизии и пятью килограммами снаряжения очень тяжел; если же человеку удастся найти воду по дороге, он сможет идти несколько месяцев.) Для сравнения возьмем детище работающего в Корнеллском университете робототехника Энди Руины и его коллег. Созданный ими робот *Ranger* на данный момент может ходить на самые дальние расстояния: он без остановки прошел 65,2 километра в ультрамарафоне роботов, который состоялся 1–2 мая 2011 г. в Японии. Проектировщики робота воспользовались преимуществами динамических свойств конечностей для создания в высшей степени энергоэффективного ходока (ему под силу часами ходить по спидвею под управлением человека с джойстиком). Потрясающий четвероногий шагающий робот *Big Dog* примерно в пятнадцать раз менее энергоэффективен, однако гораздо лучше передвигается по пересеченной местности. При этом люди по-прежнему в четыре-пять раз более эффективны в транспортировке грузов, чем *Ranger*, а еще, в отличие от *Ranger*, способны автономно реагировать на всевозможные характеристики мира. (См. Ruina, 2011.)

Так дальше продолжаться не может. Либо мы *определяем* “проект” как продукт разумного проектировщика, такого, например, как создатель стихов или автомобилей, либо признаем, что в мире могут существовать проекты – реальные проекты, – за которыми не стоит разумный проектировщик. Традиция и этимология, казалось бы, заставляют нас сделать выбор в пользу первого варианта, однако вспомните вот что: слово “атом” происходит от греческого *a* [без] + *tomos*, от глагола *temnein*, “разрезать” или “разбивать на части”. Изначально *атомом* называли “неделимую частицу”, однако наука пришла к выводу, что разделение атома не противоречит его названию. Я утверждаю, что наука также пришла к выводу, что существование проектов без проектировщика (без проектировщика, обладающего сознанием, планом и намерением) не только возможно, но и повсеместно распространено. Проектирование посредством эволюции представляет собой реальный, предельно понятный процесс. Он значительно отличается от проектирования с участием инженера, однако во многом подобен ему и способен посрамить “изобретательность”. Снова и снова биологи, ошарашенные на первый взгляд бессмысленным или неуклюжим природным проектом, в итоге убеждаются, что недооценили абсолютное величие одного из творений Матери-Природы. Фрэнсис Крик не без ехидства назвал эту тенденцию по имени своего коллеги Лесли Орджела и вывел так называемый второй закон Орджела: “Эволюция умнее вас”. (Эта очевидно безрассудная тактика персонализации процесса естественного отбора в форме Матери-Природы, то есть беззастенчивой антропоморфизации эволюции, будет оправдана позже. Это не просто забавная метафора, а полноправный инструмент мышления.)

Вернемся к проектному пространству, то есть к многомерному пространству всех *возможных* проектов, включая не только реальные организмы, автомобили и стихотворения, но и – как и книги Вавилонской библиотеки – все не получившие развития и не воплощенные в жизнь проекты: проекты говорящих кенгуру, летающих змей, атомных аппаратов для приготовления попкорна и подводных

роликовых коньков. Каким должен быть наш *алфавит* базовых проектных элементов? Поскольку мы не пытаемся *создать* проектное пространство, а лишь размышляем о нем, мы можем пуститься во все тяжкие и включить в алфавит все возможные конфигурации атомов на основе таблицы Менделеева. (В это грандиозное пространство войдут и поатомные копии всех *неспроектированных* объектов, таких как каждый камень на каждом пляже и гора Эверест, однако так и должно быть, поскольку ничто не запрещает кому-либо однажды спроектировать и построить точные копии этих объектов.) Где в проектном пространстве окажется Третья симфония Брамса? Ее *партитура* (в форме чернильных точек на бумаге и т. п.) окажется в бесконечном множестве мест, а магнитофонные записи и виниловые стереопластинки займут другие места, так что она точно будет в проектном пространстве. Песню, которую никогда не записывали ни на бумаге, ни на пленке, будет сложнее привязать к семействам атомов, однако должным образом устроенный подотдел проектного пространства прекрасно сохранит ее во времени. Подобно тому как Вавилонская библиотека в основном заполнена мусором, нечитаемыми книгами, полными чепухи, проектное пространство в основном заполняет бесполезный хлам, не представляющий никакого интереса, никак не используемый и ни на что не годный, и все же кое-где в нем проступают бесконечно тонкие очертания реальных и возможных проектов; объектов, которые способны *действовать*, а не просто ждать, пока они не окажутся раздавлены под действием неумолимого второго закона термодинамики.

Держа в уме этот черновой набросок концепции проектного пространства, мы можем “с ходу” понять, как сформулировать серию вопросов, которые преследовали споры об эволюции со времен Дарвина: существуют ли *какие-либо* реальные проекты, естественные или искусственные, происхождение которых ни прямо, ни косвенно не связано с единым древом жизни?

Вот возможные варианты ответов:

1. Нет.
2. Да, некоторые чудеса природы слишком удивительны, слишком “неприводимо сложны”, чтобы появиться в результате методичного процесса проектирования путем эволюции. Должно быть, их отдельно от остального мира создал разумный творец.
3. Да, некоторые продукты человеческой деятельности, такие как пьесы Шекспира и теорема Гёделя, слишком удивительны, чтобы быть “просто” продуктами развитого человеческого мозга; они представляют собой работы чудесного *гения* и не могут быть созданы – не могут быть достигнуты – посредством методичного восходящего процесса проектирования путем эволюции.

Можно сказать, что ответы (2) и (3) во многом опираются на контраст между предельно неэффективным методичным процессом естественного отбора (“абсолютным невежеством”, по выражению МакКензи) и стремительной, словно бы совершенно непринужденной гениальностью Моцарта (или любого другого “богоподобного” гения на ваш выбор). (Подробнее об этом в главе 49.) Используя нашу аналогию между *подъемом* и проведением НИОКР, мы можем сказать, что оба этих ответа провозглашают необходимость “небесного крюка”.

*небесный крюк* (воздухоплавание) – воображаемое изобретение для привязки к небу; воображаемый способ подвешивания к небу. [*Oxford English Dictionary*]

По данным Оксфордского словаря английского языка, впервые это понятие было использовано в 1915 г.: “Пилот аэроплана, получивший команду оставаться на месте (в воздухе) в течение следующего часа, ответил, что «машина не оборудована небесными крюками»”. Возможно, идея небесного крюка восходит к *deus ex machina* древнегреческой драматургии: когда второсортные драматурги замечали, что хитросплетения сюжета завели их героев в тупик, они часто прибегали к помощи бога, который, подобно Супермену, спасал ситуацию сверхъестественным образом. А может, небесные крюки представляют собой полностью независимый продукт фольклора. Было бы чудесно иметь небесные крюки, чтобы с их помощью поднимать громоздкие предметы из неудобных мест и ускорять всевозможное строительство. К несчастью, их существование невозможно<sup>[55]</sup>.

Однако существуют *подъемные краны*. Подъемные краны справляются с работой, которую могли бы выполнять воображаемые небесные крюки, и делают это честно – не провоцируя никаких вопросов. При этом они стоят дорого. Их нужно проектировать и конструировать из имеющихся элементов, а также устанавливать на твердом основании существующего рельефа. Небесные крюки поднимают грузы



чудесным образом, не требуя никакой поддержки. Подъемные краны справляются с этой задачей не хуже, но при этом обладают неоспоримым преимуществом: они реальны. Любой, кто, как и я, не прочь понаблюдать за стройкой, с некоторым удовлетворением замечает, что порой для сборки большого крана приходится использовать маленький кран. Несложно догадаться, что этот большой кран, в принципе, можно использовать для ускорения сборки еще большего крана. Каскадирование кранов редко более одного раза используется на реальных стройках – если используется вообще, – но, в принципе, количество кранов, которые могут быть организованы в цепочку для достижения великой цели, ничем не ограничено<sup>[56]</sup>.

Теперь представьте, сколько всего нужно “поднять” в проектное пространство, чтобы создать удивительные организмы и другие артефакты нашего мира. Должно быть, с момента зарождения жизни ранние, простейшие, самореплицирующиеся сущности преодолевали огромные расстояния, распространяясь вширь (творя разнообразие) и стремясь вверх (к совершенству). Дарвин описал самый грубый, самый примитивный, самый незамысловатый процесс подъема по наклонной плоскости естественного отбора. Маленькими – крошечными – шагами за миллионы лет этот процесс может постепенно преодолевать огромные расстояния. По крайней мере, так утверждает Дарвин. И не потребуются никаких чудес – никаких вмешательств свыше. Каждый совершаемый грубым, механическим, **алгоритмическим** образом шаг помогает продвигаться все выше с той точки, на которую привело более раннее восхождение.

Это кажется невероятным. Разве могло такое произойти на самом деле? Или же какие-то небесные крюки все-таки “подсобили” процессу (возможно, только в самом начале)? Более ста лет скептики пытались доказать, что идея Дарвина просто не может работать, во всяком случае *в полной мере*. Они надеялись найти небесные крюки, которые стали бы исключениями из неумолимо работающего дарвиновского алгоритма. Снова и снова они обнаруживали любопытные проблемы – скачки, пробелы и другие чудеса, которые, казалось бы, требовали небесных крюков. Но потом неизменно появлялись подъемные краны – и во многих случаях их обнаруживали те самые скептики, которые надеялись найти небесный крюк.

Пора дать более четкие определения. *Небесный крюк* – это “разумная” сила или “разумный” процесс, исключение из принципа, что все спроектированное и кажущееся спроектированным представляет собой результат бездумной и бесцельной механики. *Подъемный кран*, напротив, подпроцесс или особая функция процесса проектирования, которую можно использовать для локального ускорения классического, медленного процесса естественного отбора и которая сама представляет собой предсказуемый (или ретроспективно объяснимый) продукт классического процесса. Одни краны просты и не вызывают споров, в то время как дискуссии о других не прекращаются и сегодня. Чтобы показать размах и возможности применения концепции, позвольте мне привести три разных примера.

Первый подъемный кран – *симбиоз*. Мне ужасно нравится история симбиотического происхождения эукариотов. (Посмотрев на древо жизни [на с. XXX], вы заметите, что великолепное ветвление многоклеточных жизненных форм, включая всех животных и растений, происходит после эволюции эукариотов.) Примерно миллиард лет единственными живыми существами на этой планете были одноклеточные организмы, бактерии и археи, известные под общим названием прокариоты. В один прекрасный день две прокариоты столкнулись друг с другом (без сомнения, в те дни такое происходило очень часто, да и сегодня случается нередко), но ни одна из них не поглотила и не уничтожила (не съела) другую и даже не оттолкнула ее – вместо этого они объединили усилия и сформировали совершенно новый живой организм, имеющий (примерно) вдвое больше механизмов, отличающихся большим разнообразием. Миллионы лет НИОКР велись по двум направлениям, на каждом из которых совершенствовались свои признаки, но в конце концов сошлись в одной точке – и это в одночасье привело к успеху. (“Трансфер технологий” не всегда и даже не часто приносит плоды, но когда это случается, результаты бывают выдающимися.) Ни одной из генеалогий не пришлось заново изобретать все хитрости и системы другой, а поскольку комбинация их талантов привела к чистому выигрышу (сейчас это модно называть *синергией*), *эу-кариота* (“хороший” + “клетка”) оказалась более приспособлена к жизни, чем любая из прокариот, и потому род эукариот ждало процветание. Утверждение, что эукариоты произошли в результате описанного симбиоза, еще недавно вызывало споры, но этому нашлось столько подтверждений, что теперь оно окончательно вошло в учебники. (Было бы приятно заявить, что бактериальные жгутики, долгое время использовавшиеся сторонниками теории разумного замысла в качестве примера “неприводимой сложности”, также появились не без помощи подъемного крана симбиоза, однако, несмотря на отчаянные попытки доказать это, предпринимавшиеся покойным биологом Линн Маргулис, которая изучала симбиотическое происхождение эукариот, в настоящее время данные



свидетельствуют в пользу альтернативного пути эволюции жгутиков. Учитывая все богатство выбора, утверждение, что жгутикам нужен небесный крюк, кажется особенно жалким.)

Как мы видим на древе жизни, эукариоты закладывают фундамент всей многоклеточной жизни. Все разноцветные ветви справа от бактерий и архей – это ветви эукариотов. В первом приближении любой живой организм, видимый невооруженным глазом, представляет собой эукариоту. Революция эукариот открыла огромные регионы проектного пространства, однако открыла их *не для того*, чтобы предоставить доступ ко всем этим проектам. “Платить” за использование кранов приходится на местном уровне, поскольку они приносят непосредственную выгоду тем, кто получает проектные инновации. И все же, установленные однажды, они могут оказывать влияние и на будущее. (Подобным образом компьютеры были изобретены *не для того*, чтобы сделать возможными текстовую обработку и интернет, однако, как только пространство возможных применений компьютерных технологий оказалось открыто, проектирование перешло в активную фазу, в результате чего появились “виды”, которые мы теперь используем каждый день.)

Теоретики эволюции сегодня в массе своей соглашаются, что подъемным краном можно считать и *половое размножение*. Виды, размножающиеся половым путем, могут двигаться по проектному пространству быстрее организмов, для которых характерно бесполое размножение. Более того, в процессе они могут “распознавать” проектные усовершенствования, которые фактически “невидимы” для организмов, размножающихся бесполом образом (Holland 1975). Однако нельзя считать, что половое размножение существует именно для этого. Эволюция ничего не предвидит, поэтому любое ее творение должно немедленно окупаться, чтобы нивелировать затраты на его разработку. Сегодня ученые настаивают, что “выбор” полового размножения предполагает огромные *немедленные* затраты: в каждой транзакции организмы пересылают лишь 50 процентов своих генов (не говоря уже о том, сколько сил нужно потратить для обеспечения этой транзакции и с каким риском она сопряжена). Таким образом, *долгосрочные* выгоды повышения эффективности, четкости и скорости процесса перепроектирования – характеристик, делающих половое размножение мощным подъемным краном, – никоим образом не влияют на осуществляемый на локальном уровне без опяшки на будущее выбор организма, который получит преимущества уже в следующем поколении. Должно быть, выбор осуществляется под давлением какой-то другой, краткосрочной выгоды, делающей половое размножение таким привлекательным, что большинство видов не может от него отказаться. Существует множество убедительных – и конкурирующих – гипотез, которые могут решить эту загадку, впервые загаданную биологам Джоном Мейнардом Смитом (1978). Прекрасным введением в этот важный раздел эволюционной биологии служит книга Ридли “Красная королева. Секс и эволюция человеческой природы” (1993).

Половое размножение показывает, что в природе может существовать мощный подъемный кран, который был создан *не для того, чтобы эксплуатировать* эту мощь: он был сконструирован по другим причинам, но его эффективность в качестве подъемного крана может объяснить, почему он функционирует до сих пор. Намеренно сконструированным подъемным краном можно считать *генную инженерию*. Теперь генные инженеры, бесспорно, могут совершать скачки сквозь проектное пространство, создавая организмы, которые никогда бы не сформировались “обычным” образом. Это нельзя считать чудом, *учитывая, что генные инженеры (и артефакты, используемые в их ремесле) сами представляют собой продукт более ранних, более медленных эволюционных процессов*. Если бы креационисты были правы, что человечеству закон не писан, в связи с чем его божественное происхождение не объяснить грубыми дарвиновскими методами, то генная инженерия не была бы подъемным краном, поскольку для ее создания применялся бы важнейший небесный крюк. Вряд ли генные инженеры лелеют такое представление о себе, но логически оно возможно, хотя и сомнительно. Не столь очевидно глупа другая идея: если тела генных инженеров представляют собой продукт эволюции, но их *сознания* способны к творчеству, которое совершенно неалгоритмично или недостижимо алгоритмическим путем, то прорывы генной инженерии, возможно, делаются с помощью небесного крюка. Этот вариант будет кратко рассмотрен дальше.

### 39. Компетентность без понимания

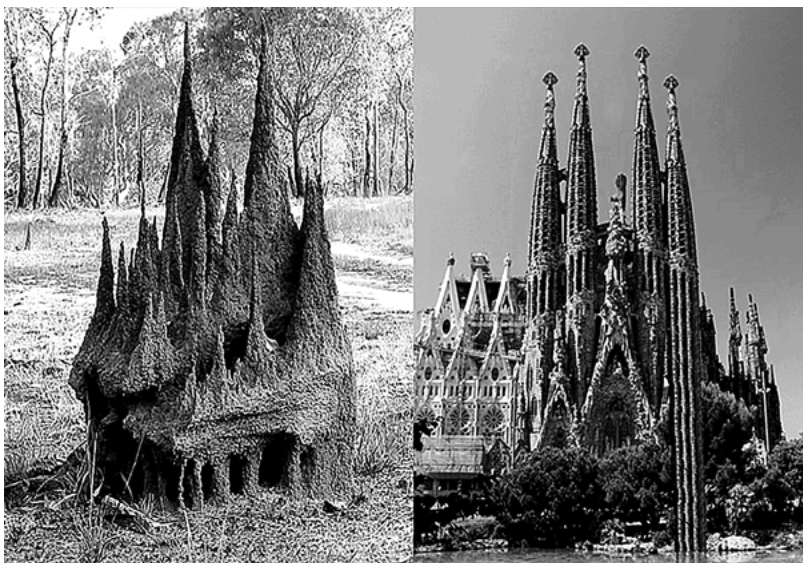
МакКензи назвал идею Дарвина об абсолютном невежестве как источнике “всех достижений творческого дара” примером “странной логики”, поскольку она переворачивает с ног на голову одну из самых “очевидных” наших идей: понимание – *источник* компетентности. Зачем мы отправляем детей в школу и почему предпочитаем “методичный подход” “зубрежке”? Дело в том, что мы считаем понимание лучшим путем к компетентности в любой области. Не работайте бездумно! *Понимайте* принципы любой работы, чтобы выполнять ее лучше! Безусловно, это прекрасный совет в большинстве сфер человеческой деятельности. Встречаются и редкие исключения, например одаренные музыканты, которые “играют на слух” и не умеют читать ноты, или прирожденные спортсмены, которые всегда побеждают, но не могут объяснить, как и почему они это делают, а следовательно, не могут тренировать других. Есть также “саванты”, которые странным образом некомпетентны в большинстве областей, но нечеловечески талантливы в какой-нибудь одной сфере. Но обычно правило работает: понимание чаще всего выступает ключом к (человеческой) компетентности.

И Дарвин действительно искажил эту логику, продемонстрировав, как метко выразился МакКензи, что в качестве создателя выступает абсолютное невежество. Процесс естественного отбора феноменально компетентен – вспомните второй закон Орджела, – но совершенно бездумен. Более того, проектируемые им организмы пользуются преимуществами своего исключительного устройства, не имея необходимости понимать, зачем и как они все это получили. Мой любимый пример – кукушка. Кукушки – *гнездовые паразиты*. Они не выют собственных гнезд. Вместо этого самка кукушки тайком подкладывает свое яйцо в гнездо пары птиц другого вида, где оно ждет внимания ни о чем не подозревающих приемных родителей. Часто самка кукушки забирает из гнезда одно из хозяйских яиц – вдруг родители сумеют их сосчитать! Как только птенец кукушки вылупляется (а он, как правило, вылупляется раньше птенцов хозяев), взрослая кукушка всеми силами старается выкатить из гнезда все остальные яйца! Зачем? Чтобы все внимание приемных родителей досталось кукушонку.

## 40. Плавающие основания

Естественный отбор автоматически находит причины – он “обнаруживает”, “одобряет” и “концентрирует” причины на протяжении многих поколений. Пугающие кавычки здесь напоминают нам, что сам по себе естественный отбор ни сознания, ни причин для своих действий не имеет, но все равно компетентно справляется с этой “задачей” совершенствования проектов. Что это, как не пример компетентности без понимания? Давайте удостоверимся, что у нас есть возможность избавиться от пугающих кавычек. Представим себе весьма многообразную популяцию. Некоторые представители этой популяции размножаются без проблем, в то время как большинство испытывает трудности. В каждом из случаев мы можем спросить *почему*. Почему вот этот представитель популяции оставил после себя жизнеспособного потомка, а остальные не оставили никого? Во многих случаях – даже в большинстве случаев – *никаких причин нет*; это просто слепая удача – одним везет, другим нет. Но есть подмножество случаев, пускай и очень ограниченное, в которых вопрос не остается без ответа. Понятно, чем эти случаи похожи и что выделяет их из основной массы, а потому мы видим, что служит причиной в каждом из них. Это позволяет функциональности накапливаться в ходе процесса, который слепо отслеживает причины и создает объекты, имеющие цели, но не имеющие нужды их знать. Биосферой правит принцип необходимости знания, а естественный отбор не испытывает необходимости знать, что он делает.

Итак, причины появились раньше тех, кто их формулирует. Отслеживаемые эволюцией причины я назвал “плавающими основаниями”, и этот термин, похоже, щекочет нервы немалого числа мыслителей, которые подозревают, что я взываю к каким-то духам – странным нематериальным идеям, которым не место в трезвом материалистическом представлении о реальности. Это вовсе не так. Плавающие основания не более призрачны или проблемны, чем числа или центры тяжести. Девять планет существовало задолго до того, как люди изобрели способ считать, а астероиды имели центры тяжести задолго до того, как физикам пришла в голову эта мысль и они сумели произвести расчеты. Не стоит путать *числа* с *цифрами* (арабскими, римскими или какими угодно еще), которыми мы их обозначаем. Цифры – изобретение человека; числа – нет. Причины – в том смысле, в котором я использую это слово, – подобны числам, а не цифрам. Мы должны с благодарностью говорить о причинах, которые эволюция распознала раньше, чем они были выражены или сформулированы исследователями-людьми или другими умами. Обратите внимание на поразительное сходство конструкций на рисунках ниже:



Термитник и Храм Святого Семейства Антонио Гауди очень похожи по форме, но их происхождение и конструкция сильно различаются. Структура и форма термитника обусловлены *определенными причинами*, однако эти причины не выражаются ни одним из термитов. Нет никакого термит-архитектора, который проектирует термитник, а отдельные термиты не имеют ни малейшего понятия о том, почему они строят его именно таким образом. Это компетентность без понимания. Структура

и форма шедевра Гауди тоже обусловлены определенными причинами, но эти причины (в основном) объясняются замыслом Гауди. Гауди *имел* причины возвести именно такие формы, а форма термитника *определяется другими причинами*, однако термиты *не имеют* никаких причин возводить его именно таким образом. Есть причины, по которым деревья широко раскидывают свои ветви, но нельзя сказать, что деревья имеют на это причины. Поведение губок небеспричинно; поведение бактерий небеспричинно; небеспричинно даже поведение вирусов. Однако ни губки, ни бактерии, ни вирусы *не имеют* причин вести себя именно таким образом – причины им просто не нужны.

Существует множество причин такого поведения, но, как правило, организмы их не понимают. Они получают готовые паттерны поведения, сформированные в процессе эволюции, и извлекают из них выгоду, не осознавая их принципов. В природе эта характеристика встречается повсеместно, но часто ее скрывает наша тенденция принимать **интенциональную установку** и считать поведение более разумным и рациональным, чем на самом деле. Как умны термиты, которые проветривают термитники с помощью удачно расположенных вентиляционных шахт! Как дальновидны белки, которые запасаются едой на зиму! Как хитры жуки, которые замирают на месте при приближении жертвы! Все эти стратегии действительно ведут к успеху в условиях жесткой конкуренции в природе, но их выгодоприобретателям не обязательно понимать все то, что понимаем мы в процессе их изучения. До нас никто не формулировал причины, которые обуславливают успешность этих стратегий.

## 41. Понимают ли цикады простые числа?

Чтобы подчеркнуть независимость чисел и цифр, плавающих оснований и сформулированных причин, рассмотрим пример цикад. В 1977 г. Стивен Джей Гулд написал любопытную работу, в которой восхищенно заметил, что длительность репродуктивного цикла цикад (например, “семнадцатилетней саранчи”) всегда выражается простыми числами – так, цикл может продолжаться тринадцать или семнадцать лет, но не пятнадцать и не шестнадцать. “Как эволюционисты, – замечает Гулд (р. 99), – мы ищем ответ на вопрос почему. Почему, в частности, возникает такая поразительная синхронность и почему между эпизодами полового размножения проходит так много времени?” Ответ – красота которого очаровывает – заключается в том, что большое простое число лет между появлениями выводков позволяет цикадам минимизировать вероятность, что их обнаружат, а затем выследят и поймут хищники, которые сами появляются раз в два или три года или даже раз в пять лет. Если бы выводки появлялись, скажем, раз в шестнадцать лет, тем хищникам, которые появляются ежегодно, полакомиться ими удавалось бы нечасто, но при этом они служили бы более надежным источником пропитания для хищников, которые появляются раз в два или четыре года. Всегда в выигрыше оставались бы хищники, попадающие в фазу с цикадами и появляющиеся раз в восемь лет. Однако если длительность репродуктивного цикла цикад не раскладывается на множители, “пытаться” отследить их имеет смысл разве что тем видам, которым повезло иметь точно такую же длительность цикла (или цикл которых в несколько раз длиннее – мифическому 34-летнему пожирателю цикад голодать бы не пришлось).

Стоит пояснить, что основательность этого объяснения (которая, впрочем, еще не установлена) не опирается ни на какую гипотезу, предполагающую, что цикады понимают арифметику, не говоря уже о простых числах. Она не опирается и на *понимание* простых чисел процессом естественного отбора. Бездумный, ничего не сознающий процесс естественного отбора может *эксплуатировать* это важное свойство некоторых чисел, вообще не испытывая нужды его понимать. Вот другой пример: ни пчелы, ни Мать-Природа не испытывают нужды *понимать* геометрию, которой объясняется совершенство шестигранной формы сот. Можно привести и множество других примеров, в которых эволюция демонстрирует математическую компетентность без понимания.

## 42. Как объяснить смотровые прыжки

Помните обескураживающую закономерность, которая наблюдалась в насосе интуиции о **двух черных ящиках**? Не нужно было располагать ни семантической, ни интенциональной интерпретацией ящиков, чтобы увидеть, что каждое нажатие на кнопку  $\alpha$  зажигало красную лампочку, а каждое нажатие на кнопку  $\beta$  зажигало зеленую лампочку, – этот очевидный паттерн не требовал объяснений. Ученые в точности понимали, как выполняются вычисления и как передается сигнал при каждом отдельном нажатии на кнопку, однако не могли вывести общее правило. Чтобы объяснить существование закономерности, необходимо прибегнуть к семантической интерпретации. Иными словами, “макрокаузальный” уровень, на котором формулируется объяснение, не “сводится” к “микрокаузальному”.

Вот другой пример этого типичного явления. Вероятно, вы видели по телевизору, как газели убегают от хищника по равнине, и замечали, что некоторые газели, стремясь уйти от преследователя, время от времени высоко выпрыгивают в воздух. Такие прыжки называются смотровыми. Зачем газели их совершают? Очевидно, смотровые прыжки идут им на пользу, поскольку газели, высоко выпрыгивающие во время погони, редко оказываются пойманными и растерзанными. Эта каузальная закономерность наблюдалась в течение долгого времени, подобно закономерности кнопок и лампочек. И она тоже может обескураживать. Как бы хорошо мы ни представляли механику работы всех белков во всех клетках всех газелей и всех преследующих их хищников, это не поможет нам объяснить существование закономерности. Чтобы понять, чем она обусловлена, нам нужно обратиться к ответвлению теории эволюции, которое называется теорией дорогостоящих сигналов (Zahavi 1987; FitzGibbon, Fanshawe 1988). Самые сильные и быстрые газели выпрыгивают, чтобы показать преследователю свою физическую форму, сигнализируя тем самым: “Даже не пытайся меня догнать, поймать меня слишком сложно. Лучше обрати внимание на моих собратьев, которые не умеют так прыгать, ведь съесть их гораздо проще!” Преследователь считает этот сигнал подлинным, поскольку подделать его нелегко, и не обращает внимания на выпрыгивающих газелей. Это плавающее основание, которое не нужно понимать ни газели, ни хищнику. Иными словами, газель может даже не догадываться, почему при возможности ей стоит делать смотровые прыжки, а хищник, скажем лев, может не понимать, почему выпрыгивающие газели не привлекают его в качестве добычи, но если бы сигналы не были подлинными, дорогостоящими сигналами, они не сохранились бы в эволюционной гонке вооружений между хищниками и жертвами. (Если бы эволюция попыталась использовать “дешевые” сигналы вроде виляния хвостом, на которое способна любая газель, какой бы сильной или слабой она ни была, львам не было бы смысла обращать на это внимание, а потому они не стали бы этого делать.) Подобные закономерности объясняются только на уровне плавающих оснований, поскольку свести их на более низкий, например молекулярный, уровень невозможно. При этом важно понимать, что, даже если причины и принципы смотровых прыжков объясняются из **интенциональной установки** (с позиции того, какой рациональный вывод лев может сделать, наблюдая смотровые прыжки газели), конкретному льву или газели нет нужды понимать смысл смотровых прыжков, чтобы они выполняли свою задачу; им достаточно лишь **вроде как** понимать его.

### 43. Опасайтесь первого млекопитающего

Вы можете полагать, что сами являетесь млекопитающим, а также что млекопитающими являются собаки, коровы и киты, но на самом деле млекопитающих вообще не существует. Их просто не может быть! Вот философское доказательство (взято с изменениями из работы Sanford 1975):

1. Мать каждого млекопитающего – млекопитающее.
2. Если бы млекопитающие существовали, их число было бы конечно.
3. Однако если бы существовало хотя бы одно млекопитающее, то, согласно пункту (1), число млекопитающих было бы бесконечно, что противоречит пункту (2), а следовательно, млекопитающих существовать не может. Возникает логическое противоречие.

Так как мы прекрасно знаем, что млекопитающие существуют, это доказательство мы рассматриваем всерьез лишь с целью понять, где в нем кроется ошибка. Очевидно, в нем используется доведение до абсурда, поэтому нужно разобраться, что с ним не так. И, в целом, мы знаем, что с ним не так: если проследить родословную любого млекопитающего достаточно далеко, рано или поздно в ней возникнут *терапсиды* – странные, вымершие виды, представляющие собой промежуточное звено между рептилиями и млекопитающими. Чистые рептилии эволюционировали в чистых млекопитающих постепенно, и многие промежуточные звенья на пути от одних к другим с трудом поддаются классификации. Как нам определить, в какой момент рептилии стали млекопитающими, наблюдая этот спектр постепенных изменений? Можем ли мы назвать первое млекопитающее, мать которого млекопитающим не была, и тем самым отвергнуть условие (1)? На каких основаниях? Какими бы ни были эти основания, они вступают в конфликт с основаниями, которые мы могли бы использовать для доказательства, что это животное не было млекопитающим, поскольку его мать была терапсидой. Как лучше проверить животное на терапсидность? Допустим, мы перечислили десять главных отличий терапсид от млекопитающих и постановили, что животное можно признать млекопитающим, если у него есть пять и более соответствующих признаков. Даже если не принимать во внимание произвольность обозначенных критериев – почему десять отличий, а не шесть и не двадцать? разве их не стоит выстроить в порядке важности? – любое подобное разделение породит множество нежелательных вердиктов, поскольку на протяжении долгого, очень долгого переходного периода от чистых терапсид к чистым млекопитающим найдется много случаев, когда млекопитающие (по нашему правилу пяти и более признаков) спаривались с терапсидами (менее пяти признаков млекопитающих) и производили потомство, которое представляло собой терапсид, рожденных у млекопитающих, млекопитающих, рожденных у терапсид, рожденных у млекопитающих, и так далее! Само собой, чтобы увидеть все эти аномалии, нам потребуется машина времени, поскольку прошли уже миллионы лет и в подробностях их уже не изучить. Но в этом нет ничего страшного, ведь в долгосрочной перспективе подробности не важны. Что нам делать? *Нам следует подавить свое желание провести черту между классами.* В этой черте нет нужды. Нам под силу смириться с тем, что все постепенные изменения, накапливавшиеся на протяжении миллионов лет, в конце концов привели к появлению настоящих млекопитающих, и в этом факте нет ничего шокирующего и таинственного. Подобным образом нет нужды и точно определять различия между озерами, прудами, болотами и топями – такая точность ни к чему даже лимнологам (специалистам по внутренним водам).

Философы, однако, склонны придирается к словам. С тех пор как Сократ потребовал, чтобы ему однозначно назвали отличительные черты добродетели, знания, смелости и подобных вещей, философы лелеют надежду остановить подобный бесконечный регресс, выявив сущность, которая станет – должна стать – его стопором: в нашем случае это первое млекопитающее. Часто в результате появляются доктрины, тонушие во множестве загадок и тайн, а философы обращаются к *эссенциализму*. Первым млекопитающим должно быть то млекопитающее, которое первым получило все *сущностные* характеристики млекопитающего. Если сущность *млекопитающего* определить невозможно – а эволюционная биология показывает нам, что это и впрямь невыполнимо, – эти философы оказываются в затруднительном положении. В связи с этим просто не обращайтесь внимания на философские требования сущности, отличительной черты, “определителя истинности”. Как правило – хоть и не всегда, – подобные требования только запускают охоту за призраками, которая может быть любопытна, но в лучшем случае приносит лишь скромные плоды.

Многим философам очень сложно отказаться от этого требования. Рациональные методы философствования, применявшиеся со времен Сократа, почти всегда предполагают необходимость определения терминов, тем самым заставляя всех участников дискуссии принять доктрину эссенциализма, пускай, как говорится, и *чисто теоретически*. Если отказаться от эссенциализма, некоторые из наших любимых типов доказательств окажутся практически бесполезными. Рассмотрим, к примеру, структуру аргументов, начинающихся с заведомо истинной *дизъюнкции*:

Или А, или не-А (что *на это* возразить?)

Если выбрать путь А, затем то-то-то, вы придете к выводу С;

если выбрать путь не-А, затем то-то-то, вы *тоже* придете к выводу С!

Следовательно, С установлено.

Но что если существует множество промежуточных случаев, в которых сложно понять, считать их А или не-А (млекопитающие или не млекопитающие, живые или неживые, сознательные или несознательные, убеждение или не убеждение, моральное или аморальное и т. д.)? Чтобы отбросить эти тревоги, нужно “провести черту”, отделяющую А от не-А, и запретить все разговоры о *вроде как*. В отсутствие этой четкой границы, определяющей сущность обсуждаемого предмета, доказательство просто не сконструировать. Такие доказательства блестяще работают в математике, где действительно можно проводить все нужные черты. Любое целое число действительно четно или нечетно; любое число действительно рационально или иррационально; а любой многоугольник действительно (трехсторонний) треугольник или нет. За пределами таких абстрактных сфер подобные аргументы работают не столь хорошо.

Упрямо настаивать, что первое млекопитающее должно существовать, даже если нам не под силу установить, когда и где оно существовало, – значит, предаваться *истерическому реализму*. Он намекает нам, что если бы мы знали достаточно, мы разглядели бы – иначе и быть не может – особое свойство млекопитающих, которое определяет сущность млекопитающих раз и навсегда. Отрицать это, как порой замечают философы, значит, путать метафизику с эпистемологией: науку о том, что (на самом деле) *существует*, с наукой о том, что мы можем *знать* о сущем. Я отвечаю, что мыслителям случается свернуть не туда и спутать метафизический вопрос с (исключительно) эпистемологическим, однако в подобных случаях их ошибку следует доказывать, а не просто озвучивать.



## 44. Когда происходит видообразование?

Любопытная черта эволюции путем естественного отбора заключается в том, что она полагается на события, которые “почти никогда” не происходят. К примеру, видообразование – процесс, в ходе которого появляются новые виды, отделяющиеся от родительских видов, – происходит чрезвычайно редко, однако каждый из миллионов видов, когда-либо существовавших на этой планете, появился в результате видообразования. Каждое рождение в каждой генеалогии потенциально может привести к видообразованию, однако видообразование не происходит почти никогда – даже в одном случае на миллион. Мутация ДНК не происходит почти никогда – даже в одном случае на триллион копирований, – но эволюция зависит от нее. Более того, подавляющее большинство мутаций либо губительны, либо нейтральны; неожиданно “удачных” мутаций не случается почти никогда. Но эволюция полагается на эти редчайшие из редких событий.

Рассмотрим насос интуиции о любопытной возможности. Насколько нам известно, сейчас планету населяет лишь один вид гоминид, *Homo sapiens*. Допустим, через пятьдесят лет все люди, за исключением горстки счастливчиков из числа наших потомков, погибнут из-за вируса. На планете останется только две группы выживших: тысяча инуитов, живущих на острове Корнуоллис неподалеку от Гренландии, и тысяча андаманцев, живущих в полной изоляции на островах посреди Индийского океана. Две этих популяции были на протяжении тысяч лет изолированы друг от друга и выработали характерные физиологические различия в ответ на разные условия обитания, но у нас нет оснований сомневаться в стандартном предположении, что они принадлежат к нашему виду. Допустим, эти популяции останутся географически и репродуктивно изолированными друг от друга еще на десять тысяч лет и в конце концов населят планету двумя видами – это выяснится, когда они встретятся и обнаружат, что спаривание друг с другом не представляет для них интереса, а немногочисленные случайные попытки спаривания не приносят результата, что характерно для *аллопатрического видообразования*, при котором географическая изоляция со временем приводит к репродуктивной. Они могут задуматься: когда именно произошло видообразование? Возможно, их последний общий предок жил более тридцати тысяч лет назад, но видообразование произошло не в ту же секунду (насколько нам известно, оно не произошло до сих пор!), и все же через несколько тысяч лет может выясниться, что оно случилось в какой-то момент до воссоединения двух популяций. Произошло ли видообразование до появления земледелия или после создания интернета? Обоснованного, доказуемого ответа на этот вопрос не существует. Последний общий предок видов, или *копредок*, как называет его Докинз (2004), возможно, жил на этой планете тридцать тысяч лет назад, а потомки этого предка в свое время *стали* основателями двух разных видов, но сегодня нельзя с точностью сказать, в тот ли момент началось видообразование.

Это событие – рождение, которое могло в итоге сыграть ключевую роль в человеческой (и постчеловеческой) истории, – случилось в конкретное время в конкретном месте, но не приобрело свой особый статус, пока в этой роли его не утвердили тысячелетия *последствий*, которые никогда заранее не предопределены. Чтобы это рождение не стало отправной точкой видообразования, единственной группе островитян на лодке (или на самолете) достаточно совершить путешествие, которое приведет к “преждевременному” воссоединению ветвей родословного древа. Можно *вообразить*, что видообразование произошло в конкретный, но не постижимый момент в период между изначальной изоляцией генеалогий и итоговой демонстрацией их статуса в качестве двух отдельных видов, но как определить этот переломный момент? Вероятно, это самый ранний момент, в который – благодаря накоплению хромосомных различий между двумя генеалогиями – *попытки* спаривания представителей разных генеалогий *перестали бы* приносить потомство. Но подобные гипотетические предположения не имеют ценности.

Прокладка трансконтинентальной железной дороги разделила стада американских бизонов на репродуктивно изолированные популяции, но Буффало Билл с товарищами в зародыше задушили потенциальный момент видообразования, очень быстро истребив одну из популяций. Природные катаклизмы часто разделяют популяции конспецифичных организмов на две (и более) изолированные группы, которые несколько поколений остаются репродуктивно изолированными друг от друга, и почти всегда эти группы либо воссоединяются в итоге, либо одна из них вымирает. Таким образом, несмотря на то, что подобные первые шаги к видообразованию должны случаться достаточно часто, они почти никогда не ведут к реальному видообразованию, а когда это все-таки происходит, результат закрепляется на протяжении жизни многих сотен поколений. Никакие сведения об обстоятельствах изначального

разделения не могут сказать, положило ли это разделение начало видообразованию, даже если вы будете в точности знать, в каком физическом состоянии в тот момент пребывала каждая молекула в мире. Само понятие вида – лишь *вроде как* понятие. Домашние собаки, койоты и волки – представители разных видов, и все же достаточно часто встречаются помеси собак с койотами или с волками, поэтому, возможно, “официально” нам стоит считать их лишь вариациями, а даже не подвидами семейства собачьих. Сделать общие выводы о возможности гибридизации – появления потомства у представителей разных видов – довольно сложно, но в этом нет ничего удивительного, если вспомнить, что любой представитель любого вида несколько отличается от любого другого представителя того же вида. Это не слишком тревожит биологов: они привыкли не волноваться из-за определений и сущностей, поскольку прекрасно понимают процессы, приводящие к появлению всех промежуточных типов.

## 45. Вдоводелы, Митохондриальная Ева и ретроспективные коронации

Женщина в Нью-Йорке может неожиданно стать вдовой из-за пули, которая пробила голову какого-то мужчины в Додж-Сити, на расстоянии более тысячи миль от нее. (Во времена Дикого Запада один из револьверов прозвали “Вдоводелом”. Порой даже самый внимательный осмотр места преступления не позволит установить, оправдал ли конкретный револьвер в конкретном случае свое прозвище.) В этом примере пространство и время преодолеваются благодаря условной природе взаимоотношений в браке, где событие прошлого, то есть свадьба, создает постоянную связь – формальную, а не каузальную, – которая не разрывается, несмотря на последующие скитания или несчастья (к примеру, случайную потерю обручального кольца или уничтожение свидетельства о браке).

Система генетического воспроизводства естественна, а не условна, но работает как часы, что систематически позволяет нам *формально* размышлять о каузальных цепочках, которые растягиваются на миллионы лет и выявить и отследить которые в ином случае было бы практически невозможно. Это дает нам возможность интересоваться более дистанцированными и невидимыми на местном уровне отношениями, чем формальные брачные отношения, а также активно о них рассуждать. Видообразование, как и брак, подкрепляется строгой, формально определяемой системой взлядов, но, в отличие от брака, не имеет наблюдаемых условных признаков – свадеб, колец, свидетельств. Как мы только что видели, феномен видообразования также удивительно “дистанционен” в пространстве и времени. Постепенно границы видов размываются, поэтому мы лишь ретроспективно (и произвольно) можем назначить отдельный организм из всей совокупности на роль первого млекопитающего (так что не стоит и беспокоиться). Чтобы лучше изучить это свойство видообразования, стоит сначала обратить внимание на другой пример ретроспективной коронации – присвоение титула Митохондриальной Евы, которое нельзя считать произвольным.

Границы между отдельными организмами четче, а следовательно, своеобразие отдельных организмов очевиднее своеобразия видов, но и здесь наблюдается множество промежуточных вариаций. Возьмем самый поразительный пример: в вашем теле примерно десять триллионов клеток, и девять из десяти этих клеток – не человеческие! Да, симбионты тысяч видов превосходят по численности ваши собственные клетки – клетки, которые произошли от зиготы, сформированной в результате союза ваших родителей. Среди этих симбионтов не только бактерии, но и эукариоты, как одноклеточные микробы, так и многоклеточные организмы: грибки, клещи у вас на ресницах и в других местах организма, микроскопические и более крупные черви – чего там только нет. Вы представляете собой ходячую экосистему. Хотя некоторым гостям мы не рады (грибкам, вызывающим эпидермофитию стоп, а также бактериям, лишаящим дыхание свежести и стремящимся к любой инфекции), другие настолько важны, что вы умрете, если сумеете выселить всех своих квартирантов. Поскольку эти клетки-симбионты, как правило, гораздо меньше человеческих, на *вес* вы в основном человек, однако их совокупную массу не назвать ничтожной – наберется пара килограммов, а может, даже все пять. Существуют также вирусы, которых еще больше.

И все же, несмотря на ваши пористые границы, вы – как и остальные отдельные организмы – легко отличимы от других. Порой мы можем назвать конкретный организм, который исполнил конкретную роль в эволюционной истории. Один из самых знаменитых подобных организмов – Митохондриальная Ева. Это женщина, которая по женской линии является *ближайшим* прямым предком каждого человека, *живущего сегодня*. В клетках каждого из нас содержатся митохондрии, которые передаются только по материнской линии, поэтому все митохондрии во всех клетках всех живущих сегодня людей – прямые потомки митохондрий в клетках конкретной женщины. Ребекка Канн, Марк Стоункинг и Аллан Вильсон (1987) назвали ее Митохондриальной Евой.

Митохондриями называются крошечные внутриклеточные органеллы, которые играют важнейшую роль в обмене веществ, извлекая энергию из пищи и используя ее во всех начинаниях тела. Митохондрии обладают собственной ДНК, что и свидетельствует об их симбиотическом происхождении несколько миллиардов лет назад. Анализируя паттерны митохондриальной ДНК различных живущих сегодня людей, ученые сумели определить, когда примерно жила Митохондриальная Ева и даже где она жила. Сначала считалось, что Митохондриальная Ева жила в Африке около трехсот тысяч лет назад, однако недавно эти данные были скорректированы: она жила (почти наверняка в Африке) всего двести тысяч лет назад. Определить, *где* и *когда* жила Митохондриальная Ева, гораздо сложнее, чем установить *сам факт* ее существования, в котором не сомневается ни один биолог. Рассмотрим, что нам известно

о Митохондриальной Еве, не заостряя внимания на спорных моментах. Мы знаем, что у нее было по меньшей мере две дочери, дочери которых тоже выжили. (Если бы у нее была всего одна дочь, корона Митохондриальной Евы досталась бы именно этой дочери.) Чтобы не путать ее титул с собственным именем, давайте назовем ее Ами. Ами носит титул Митохондриальной Евы, потому что она стала матерью-прародительницей всех современных людей. Важно не забывать, что *во всех остальных отношениях* в Митохондриальной Еве, вероятно, не было ничего особенного или уникального: она совершенно точно не была ни первой женщиной, ни основательницей вида *Homo sapiens*. До нее жило множество женщин, которые, без сомнения, принадлежали к нашему виду, однако ни одна из них, как оказалось, не стала самым недавним источником митохондрий всех живущих сегодня людей. Кроме того, хотя у Митохондриальной Евы были дочери и внуки, вероятно, она не была заметно сильнее, быстрее, красивее или плодотворнее других женщин ее времени.

Чтобы подчеркнуть, насколько непримечательной, скорее всего, была Митохондриальная Ева – то есть Ами, – допустим, что завтра, много тысяч поколений спустя, по земле распространится новое опасное заболевание, которое за несколько лет истребит 99 процентов человечества. Выжившие, которым повезло иметь какую-то врожденную сопротивляемость к смертельному вирусу, вероятно, будут состоять в достаточно близком родстве. Их ближайшим общим прямым предком по женской линии – назовем ее Бетти – будет какая-то женщина, жившая на сотни тысяч поколений позже Ами, и корона Митохондриальной Евы задним числом перейдет к ней. Возможно, именно она была источником мутации, которая несколько веков спустя помогла спасти вид от вымирания, но *ей самой* это, вероятно, не принесло никакой пользы, поскольку агрессивной формы вируса, которой противостоит мутация, тогда еще не существовало. Суть в том, что короновать Митохондриальную Еву можно только *задним числом*. Ее ключевая роль в истории определяется не только случайностями в период жизни Ами, но и последующими случайностями. Это потрясающее совпадение! Если бы дядя Ами не спас ее от утопления, когда ей было три года, никого из *нас* (именно с нашей митохондриальной ДНК, которую мы унаследовали от Ами) вообще бы не было! Если бы все внуки Ами умерли от голода в младенчестве, как случилось в то время со многими детьми, мы бы тоже не появились на свет.

Согласно той же логике существует – должен существовать – и Адам, ближайший прямой предок каждого живущего сегодня мальчика и мужчины по мужской линии. Его можно назвать Y-хромосомным Адамом, поскольку все наши Y-хромосомы передаются по отцовской линии, подобно тому как митохондрии передаются по материнской линии<sup>[57]</sup>. Был ли Y-хромосомный Адам мужем или любовником Митохондриальной Евы? Исключено. Отцовство требует гораздо меньше времени и энергии, чем материнство, а следовательно, *логически* возможно, что Y-хромосомный Адам жил *весьма* недавно и был очень, очень активен в спальне – утирая нос Эрролу Флинну. Если самому старому живущему сегодня мужчине, скажем, 110 лет, *логически* возможно, что Y-хромосомным Адамом был его отец, Дон Жуан начала двадцатого века, который также является отцом, дедом, прадедом и т. д. всех живущих сегодня более молодых мужчин. В конце концов, мы, мужчины, производим миллиарды сперматозоидов, по сотне миллионов при каждой эякуляции, поэтому Y-хромосомный Адам всего за неделю мог бы произвести достаточное количество спермы, чтобы стать отцом всему человечеству (в принципе)! Однако если учесть все генетические различия в мужских Y-хромосомах по всему миру и рассчитать, сколько времени потребуется для накопления подобных мутаций, мы можем сказать, что Y-хромосомный Адам жил чуть меньше ста тысяч лет назад. И снова, если бы чума унесла, скажем, половину мужского населения, вероятнее всего, корона Y-хромосомного Адама перешла бы к прародителю, который жил гораздо менее давно<sup>[58]</sup>.

Любопытный факт о любом отдельном организме – будь то вы, я, ваша собака или ваша герань – заключается в том, что он *потенциально* может основать новый вид, стать первым в длинном ряду кактамихов, но только сотни или тысячи поколений спустя кактамихи достаточно выделятся из общей массы, чтобы их признали отдельным видом, так что ко времени коронации вы, я, ваша собака или герань давно обратятся в прах. Ваши родители, таким образом, могут стать ближайшими общими предками всех представителей двух видов гоминид, но не радуйтесь раньше времени. Чихуахуа и немецкий дог принадлежат к одному виду *Canis familiaris*, но если цивилизация падет, а их потомки одичают, они с большей вероятностью выделятся в два отдельных вида, чем, скажем, бигли и бассет-хаунды, поскольку без человеческого вмешательства оплодотворение чихуахуа немецким догом – или немецкого дога чихуахуа – маловероятно. Вполне вероятно, однако, что обе генеалогии вымрут, как происходило со множеством генеалогий на протяжении тысячелетий, прежде чем это произойдет.

По оценкам, более 99 процентов всех организмов, которые когда-либо жили на земле, умерли, не оставив потомства. И все же вы живете, а это значит, что ни один из миллиардов ваших предков,

от одноклеточных организмов до червей, рыб, рептилий, млекопитающих и приматов, не умер бездетным! Как же вам повезло! Само собой, каждая травинка, каждый комар, каждый слон и каждая маргаритка могут похвастаться столь же длинной и славной генеалогией.

## 46. Циклы

Всем знакомы крупномасштабные циклы природы: день-ночь-день; лето-осень-зима-весна-лето-осень-зима-весна; цикл испарения воды и выпадения осадков, благодаря которому снова наполняются озера, становятся полноводными реки и восстанавливаются запасы воды, питающие все живое на нашей планете. Но не все понимают, что эти циклы – в каждом пространственном и временном масштабе, от атомного до астрономического – выступают в качестве скрытых двигателей, приводящих в действие все удивительные природные явления. В 1861 г. Николаус Отто сконструировал и продал первый работающий бензиновый двигатель внутреннего сгорания, а в 1897 г. свой двигатель представил Рудольф Дизель, и эти блестящие изобретения изменили мир. В основе работы каждого из двигателей лежит цикл – четырехтактный цикл Отто и двухтактный цикл Дизеля, – который производит некоторое действие и затем возвращает систему в исходное положение, чтобы она была готова работать дальше. Механика этих циклов весьма оригинальна – она была открыта и оптимизирована в ходе цикла НИОКР, растянувшегося на несколько сотен лет. Еще более элегантным, микроминиатюризованным двигателем стал цикл Кребса, открытый в 1937 г. Хансом Кребсом, но изобретенный за миллионы лет эволюции на заре жизни. Это химическая реакция, имеющая восемь стадий и превращающая топливо – пищу – в энергию в процессе метаболизма, который имеет критическую значимость для всех живых организмов, от бактерий до секвой.

Биохимические циклы вроде цикла Кребса отвечают за все движение, рост, саморегенерацию и воспроизводство в живом мире. Они представляют собой колесики внутри колесиков внутри колес, механизм с триллионами подвижных частей, каждый из элементов которого необходимо перематывать, восстанавливать в исходном положении, чтобы он готов был снова выполнять свою функцию. Все эти циклы оптимизированы великим дарвиновским циклом воспроизводства, в котором поколение за поколением на протяжении миллионов лет выбираются удачные усовершенствования.

В совершенно другом масштабе наши предки обнаружили эффективность циклов, сделав один из важнейших успехов ранней истории человечества: распознав роль повторений в процессе производства. Если взять палку и провести по ней камнем, не произойдет практически ничего – разве что на дереве появится пара царапин. Если вернуть камень на исходную позицию и повторить движение, снова почти ничего не изменится, как бы вы ни старались. Даже после сотни повторений смотреть будет не на что. Но если точно так же провести камнем по палке несколько тысяч раз, палка превратится в прямое древко стрелы. Накапливая незаметные изменения, циклический процесс приводит к созданию нового. Необходимая для осуществления подобных проектов комбинация дальновидности и самоконтроля тоже была человечеству в новинку и свидетельствовала о большом шаге вперед от преимущественно инстинктивных строительных и созидательных процессов, наблюдаемых у других животных. А эта новизна, в свою очередь, была продуктом дарвиновского цикла, подкрепленного более быстрым циклом культурной эволюции, в ходе которого техника передавалась потомству не на генетическом уровне, а распространялась среди не связанных родством людей, которые научились имитации.

Первый предок, который отполировал камень, сделав из него ручной топор приятной глазу симметричной формы, должно быть, в процессе этого выпядел глупо. Он сидел, часами точа свой камень, но это ни к чему не приводило. Однако бесконечные бездумные повторения приводили к постепенным совершенствованиям, не заметным невооруженным глазом, ведь глаз был создан эволюцией, чтобы замечать изменения, происходящие гораздо быстрее<sup>[59]</sup>. Эта кажущаяся тщетность порой сбивала с толку искушенных биологов. В прекрасной книге “Биологический фактор” специалист по молекулярной и клеточной биологии Деннис Брэй (2009) описывает циклы нервной системы:

На типичном сигнальном пути белки постоянно модифицируются и восстанавливаются. Киназы и фосфаты непрестанно работают, подобно муравьям, добавляя к белкам фосфатные группы и удаляя их снова. Казалось бы, в этом нет никакого толка, особенно если учесть, что каждый цикл добавления и удаления стоит клетке одной молекулы АТФ – одной единицы драгоценной энергии. Циклические реакции подобного рода изначально даже назывались “футильными”<sup>[60]</sup>. Однако это прилагательное некорректно. Добавление фосфатных групп к белкам – самая типичная реакция в клетках, и она лежит в основе многих осуществляемых ими вычислений. Получается, что эта циклическая реакция вовсе не бесполезна, поскольку она снабжает клетку важнейшим ресурсом: гибким и быстро настраиваемым механизмом. [р. 75]

Слово “вычисления” выбрано очень удачно. Еще не прошло и ста лет с тех пор, как программисты начали исследовать пространство всех возможных вычислений, но пока их урожай изобретений и открытий состоит из миллионов циклов внутри циклов внутри циклов. Как выясняется, вся “магия” познания зависит, как и сама жизнь, от циклов внутри циклов повторяющихся, “реципрокных”, неосознанных информационно-трансформационных процессов в диапазоне от наноуровневых биохимических циклов в каждом нейроне и циклов перебора предиктивного кодирования систем восприятия информации (блестящее исследование см. в работе Clark 2013) до цикла сна всего мозга, крупномасштабных волн церебральной активности и восстановления, которые фиксируются на ЭЭГ. Секрет совершенствования в любой сфере жизни один: практика, практика, практика.

Важно помнить, что дарвиновская эволюция представляет собой лишь один тип цикла накопления и совершенствования. Есть и множество других. Проблема происхождения жизни может казаться неразрешимой (“неприводимо сложной” – Behe 1996), если утверждать, как делают сторонники теории разумного замысла, что, поскольку эволюция путем естественного отбора полагается на воспроизводство, *дарвиновского* ответа на вопрос, как появился первый организм, способный к воспроизводству, просто не может быть. Безусловно, он был невероятно сложен и прекрасно спроектирован – должно быть, произошло чудо. Если представить добиологический, дорепродуктивный мир безликим хаосом химических веществ (наподобие разрозненных частей самолета, собираемых ураганом, в который нас заставляют поверить креационисты), проблема действительно внушает страх, но если вспомнить, что ключевой процесс эволюции – циклическое повторение (и прекрасно отточенное, оптимизированное генетическое копирование – лишь частный его случай), у нас появится способ превратить эту тайну в загадку: как все эти сезонные, водные, геологические и химические циклы, повторяющиеся миллионами лет, постепенно накопили предпосылки для появления биологических циклов? Вероятно, первая тысяча “проб” была бесполезна – фактически все эти пробы стали осечками. Но, как говорит нам невероятно чувственная песня Джорджа Гершвина и Бадди ДеСильвы, всегда лучше проверить, что случится, если “сделать это снова” (и снова, и снова)<sup>[61]</sup>.

Таким образом, сталкиваясь с очевидным волшебством мира живой природы и сознания, лучше всего поискать в нем циклы, которые выполняют всю тяжелую работу.

## 47. Что именно глаз лягушки сообщает ее мозгу?

Одной из первых классических работ по когнитивной науке считается знаменитая статья “Что глаз лягушки сообщает ее мозгу”, написанная Дж. Леттвином и его коллегами (1959). В ней было продемонстрировано, что зрительная система лягушки чувствительна к маленьким движущимся темным точкам на сетчатке – крошечным теням, которые почти в любых природных условиях отбрасывают пролетающие неподалеку мухи. Этот механизм обнаружения мух мгновенно заставляет лягушку выбросить язык, что и объясняет, как лягушки добывают пропитание в этом жестоком мире и обеспечивают продолжение своего рода. Но что *именно* глаз лягушки сообщает ее мозгу? Что рядом муха, или что рядом муха или “обманка” (фальшивая муха любого рода), или объект типа К (какой бы тип объектов ни заставлял это зрительное приспособление среагировать наверняка – вспомните **четвертачник**)? Этот вопрос подняли теоретики-дарвинисты (например, Рут Милликен, Дэвид Израэль и я), но главный критик эволюционной теории Джерри Фодор вмешался в дискуссию, чтобы показать, что, на его взгляд, не так с любым дарвинистским представлением о подобных смыслах: эти представления слишком неопределенны. Они не различают, как должны бы, такие поступающие от глаза лягушки сигналы, как “муха здесь сейчас” и “муха или маленькая черная частица здесь сейчас” и так далее. Но это не так. Чтобы различать сигналы, можно использовать среду выбора лягушки (насколько возможно установить, какой она была). Для этого используются те же самые соображения, которые использовались при решении вопросов о значении того или иного состояния четвертачника. В тех случаях, когда в среде выбора нет ничего, что однозначно говорило бы о конкретном типе сигналов, нет и предмета для обсуждения по вопросу о том, что *на самом деле* значит поступающий от глаза сигнал. Наглядно продемонстрировать это можно, отправив лягушку в Панаму – или, точнее, отправив лягушку в новую среду выбора.

Допустим, ученые собрали небольшую популяцию питающихся мухами лягушек вымирающего вида и поместили под опеку в новую среду – в специальный лягушачий зоопарк, где не было мух, но были наблюдатели, которые периодически бросали подопечным лягушкам маленькие частички пищи. К радости наблюдателей, система оказалась рабочей: лягушки прекрасно себя чувствовали, на лету хватая эти частички, и через некоторое время у них появилось потомство. Молодые лягушки вообще никогда не видели мух – только частички пищи. Что же сообщали *их* мозгу *их* глаза? Если вы утверждаете, что значение сигналов не изменилось, вы в тупике. История с лягушками – искусственно чистый пример типичной для естественного отбора *экзантиции*, то есть приспособления существующих структур для выполнения новых функций. Как напомнил нам Дарвин, приспособление механизмов для новых целей – один из секретов успеха Матери-Природы. Если кто-то еще сомневается в этом, можно допустить, что не все лягушки в зоопарке чувствуют себя одинаково хорошо. Некоторые из них не слишком удачливы в обнаружении частичек пищи с помощью своих зрительных систем, а потому едят меньше остальных и оставляют меньше потомства. В короткие сроки отбор начнет отдавать предпочтение лягушкам, которые лучше справляются с обнаружением пищи, хотя будет ошибкой спрашивать, *когда именно* это случится достаточное количество раз, чтобы это “имело значение”. Не ждите, когда прозвонит колокольчик, возвещающий, что глаз лягушки *только что* начал сообщать ее мозгу иное. Первого млекопитающего не было. Нет и первого случая обнаружения частички пищи.

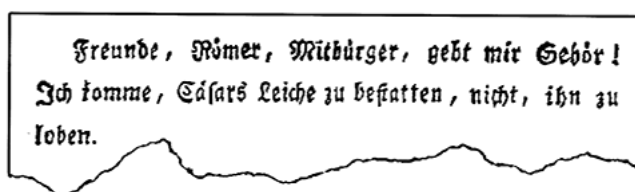
Если только не было “бессмысленной” или “промежуточной” вариации в условиях отправки сигнала глазами разных лягушек, не может быть никаких оснований (слепая вариация) для выбора новой цели действия. Неопределенность, которую Фодор (и не только) считает недостатком дарвинистских представлений об эволюции смысла, на самом деле выступает предпосылкой такой эволюции. Идея о том, что глаз лягушки *на самом деле* должен иметь в виду *нечто определенное* – возможно, какое-нибудь непознаваемое утверждение на лягушачьем языке, которое *точно* описывает, что именно глаз лягушки сообщает ее мозгу, – фактически представляет собой проявление эссенциализма в отношении смысла (или функции). Смысл, как и функция, от которой он непосредственным образом зависит, не определяется в момент его рождения. Он возникает не в результате скачания – огромных скачков в проектном пространстве – и не в результате намеренного творения, а в результате изменения обстоятельств (как правило, постепенного).



## 48. Скачки в пространстве Вавилонской библиотеки

В 1988 г. великому историку астрономии Отто Нейгебауэру прислали фотографию фрагмента греческого папируса, на котором в столбик было выписано несколько чисел. Отправитель, специалист по классической истории, понятия не имел, что значит этот фрагмент, и решил посоветоваться с Нейгебауэром. Вычислив разницу между числами в разных строках, а также верхний и нижний пределы ряда, 89-летний ученый пришел к выводу, что на папирусе записан перевод фрагмента “Колонки G” с вавилонской клинописной таблички, содержащей лунные эфемериды вавилонской системы “Б”! (Эфемерида – это таблица для вычисления местоположения небесного тела в любой момент времени в конкретный период.) Как Нейгебауэру удалось сделать этот вывод, достойный Шерлока Холмса? Элементарно: он узнал, что написанное на греческом (последовательность шестидесятеричных – не десятичных – чисел) было фрагментом – колонкой G! – в высшей степени точных расчетов местоположения Луны, осуществленных вавилонянами. Существует множество способов вычисления эфемерид, и Нейгебауэр знал, что любой, кто займется расчетами сам, используя собственную систему, не получит в итоге точно такие же числа, хотя результат и может оказаться близким. Вавилонская система “Б” была великолепна, поэтому в переводе ее сохранили полностью – вместе со всеми характерными чертами (Neugebauer 1989).

Нейгебауэр был великим ученым, но и вы, возможно, сумеете по его примеру применить дедукцию. Допустим, вам прислали фотокопию приведенного ниже текста, сопроводив ее теми же вопросами: что это значит? откуда этот фрагмент?



Прежде чем читать дальше, попробуйте разгадать эту загадку. Вероятно, вы справитесь с ней, даже если не умеете читать старый немецкий шрифт *фрактур* – и даже если вовсе не знаете немецкого! Присмотритесь к этому фрагменту. Прочитайте его с выражением, добавьте звучности. Не обращайте внимания на произношение. Узнали? Впечатляет! Может, Нейгебауэр и распознал вавилонскую колонку G, но вы быстро – так ведь? – установили, что этот фрагмент, *должно быть*, представляет собой отрывок из немецкого перевода елизаветинской трагедии (если точнее, акт III, сцена 2, строки 79–80). Стоит вам подумать об этом, как вы поймете, что ничем иным это быть не может! Вероятность того, что *эта* конкретная последовательность немецких букв родилась при любых других обстоятельствах, **Исчезающе мала.** <sup>[62]</sup>

Проведенные при создании этой последовательности НИОКР слишком специфичны, чтобы их можно было продублировать случайно. Почему? В чем особенность подобной цепочки символов? Николас Хамфри (1987) добавил этому вопросу остроты, переформулировав его следующим образом: если бы вас заставили “предать забвению” один из следующих шедевров, какой бы вы выбрали – “Начала” Ньютона, “Кентерберийские рассказы” Чосера, “Дона Жуана” Моцарта или Эйфелеву башню? Хамфри *отвечает*:

Если бы меня заставили сделать выбор, не приходится и сомневаться, что я остановился бы на “Началах”. Почему? Потому что из всех перечисленных работ только работа Ньютона *незаменима*. Все просто: если бы Ньютон не написал ее, ее написал бы кто-то другой – возможно, всего через несколько лет... “Начала” стали великолепным памятником человеческому интеллекту, а Эйфелева башня – относительно скромным торжеством романтической инженерии, но факт остается фактом: Эйфель работал на *свой* лад, а Ньютон – на лад Бога.

## 49. Кто автор “Спамлета”?

Допустим, доктор Франкенштейн создал монстра Шпексира, который после этого сел и написал пьесу “Спамлет”.

Кто автор “Спамлета”?

Для начала давайте обратим внимание на то, что, на мой взгляд, не имеет значения в этом насосе интуиции. Я не сказал, является Шпексир роботом, сконструированным из металла и кремниевых чипов, или, подобно оригинальному чудовищу Франкенштейна, собран из человеческих тканей – или при помощи методов нанотехнологии создан из клеток, или белков, или аминокислот, или атомов углерода. Неважно, какие материалы были использованы при конструировании монстра, – главное, что его создал доктор Франкенштейн. Вполне возможно, единственный способ сделать робота, который был бы достаточно маленьким, быстрым и энергоэффективным, чтобы сидеть на стуле и печатать пьесу, – это сконструировать его из искусственных клеток, наполненных искусно синтезированными двигательными белками и другими нанороботами на основе углерода. Этот вопрос любопытен с технической и научной точки зрения, однако здесь он не имеет значения. По той же самой причине, если робот Шпексир сконструирован из металла и кремния, не исключено, что по размерам он превосходит галактику, поскольку иначе его программу просто не сделать достаточно сложной. К тому же нам, возможно, придется пренебречь ограничением скорости света, чтобы представить, как описанное происходит в рамках человеческой жизни. Такие технические ограничения, как правило, признаются неприципиальными для подобных насосов интуиции, и мы не будем спорить, поскольку от них действительно ничего не зависит. (Покрутите регуляторы и проверьте, так ли это.) Если доктор Франкенштейн решит создать своего робота на основе искусственного интеллекта (ИИ) из белков и подобных элементов, это его дело. Если его робот окажется кросс-фертильным с обычными людьми, а следовательно, способным к созданию нового вида посредством рождения ребенка, это будет просто потрясающе, но нас интересует только мнимое дитя разума Шпексира – “Спамлет”. Вернемся к нашему вопросу: кто автор “Спамлета”?

Для ответа на этот вопрос нужно заглянуть внутрь Шпексира и понять, что там происходит. В одном из крайних случаев мы найдем внутри файл (если Шпексир – робот с компьютерной памятью) – фактически *сохраненную в памяти* версию “Спамлета”, готовую к воспроизведению. В этом случае автором “Спамлета”, безусловно, следует считать доктора Франкенштейна, который создает свое промежуточное творение, Шпексира, лишь в качестве устройства хранения и передачи информации, то есть исключительно замысловатого текстового процессора. Все НИОКР проводятся ранее и тем или иным способом копируются в память Шпексира.

Чтобы четче визуализировать это, можно представить “Спамлета” в галактике ближайших соседей в **Вавилонской библиотеке**. Как “Спамлет” туда попал? По какой траектории шли приведшие к его созданию НИОКР? Если мы увидим, что весь путь был уже пройден к моменту создания и наполнения информацией памяти Шпексира, будет очевидно, что Шпексир не сыграл в этом процессе никакой роли. Если мы, работая с конца, выясним, что Шпексир только пропустил сохраненный текст через программу проверки орфографии, прежде чем использовать его в качестве образца для воспроизведения, вряд ли мы сможем назвать Шпексира автором этого текста. Это измеримая, но **Исчезающе малая** доля всех проведенных НИОКР. Галактика почти идентичных текстов “Спамлета” достаточно велика – в нее входит около сотни миллионов разных мутантов, имеющих лишь одну опечатку, – а если расширить горизонты и включить все экземпляры с одной опечаткой на страницу, мы войдем в пространство Чрезвычайно больших количеств вариаций на заданную тему. Работая дальше и переходя от опечаток к *одумкам*<sup>[63]</sup>, то есть ошибочным или неверно выбранным словам, мы придем к вопросу серьезного авторства, а не просто редактуры. Относительная незначительность редактуры и ее неоспоримая важность в формировании итогового продукта прекрасно представляются в проектном пространстве, где учитывается каждый маленький подъем и где даже маленький подъем порой перемещает вас на совершенно новую траекторию. Здесь сложно удержаться и не процитировать Людвиг Миса ван дер Роэ: “Бог – в деталях”.

Теперь давайте покрутим регуляторы нашего насоса интуиции и рассмотрим другой крайний случай, в котором доктор Франкенштейн оставляет большую часть работы Шпексиру. Безусловно, реалистичнее всего выглядит сценарий, в рамках которого доктор Франкенштейн наделяет Шпексира виртуальным прошлым и коллекцией псевдовоспоминаний о пережитом, чтобы монстр мог обращаться к ним, реагируя

на навязчивое желание написать пьесу, заложенное в него Франкенштейном. Можно предположить, что среди этих псевдовоспоминаний есть множество воспоминаний о театральных вечерах и о чтении книг, а также о неразделенной любви, шокирующе опасных ситуациях, горьких предательствах и тому подобных вещах. Что происходит в таком случае? Возможно, отрывок из показанной в новостях человеческой истории станет катализатором, который спровоцирует Шпексира без устали перебирать полезные крупницы своих воспоминаний, разыскивая нужные темы, а затем преобразовывать найденное и переставлять фрагменты, формируя тем самым временные, перспективные структуры, борющиеся друг с другом за право быть законченными и в большинстве своем уничтожаемые в ходе критического отбора, но тоже время от времени дающие полезные элементы, позволяя процессу идти дальше. Весь этот многоуровневый поиск в некотором роде управляется многоуровневыми эндогенными оценками, включая оценку оценки функций оценки в ответ на оценку (оценки) продуктов текущих поисков (**циклы** внутри циклов внутри циклов).

Если же великолепный доктор Франкенштейн предвидел всю эту активность в мельчайших деталях на самом беспокойном и хаотическом уровне и собственноручно создал виртуальное прошлое Шпексира, а также спроектировал его поисковый аппарат, чтобы произвести только один продукт, “Спамлета”, то доктор Франкенштейн снова будет автором “Спамлета”, но также, одним словом, Богом. Такое знание о будущем Чрезвычайно велико, а потому чудесно. Чтобы сделать нашу фантазию чуточку реальнее, можно ослабить регуляторы и предположить, что доктор Франкенштейн не мог предвидеть все это в деталях и просто в основном делегировал Шпексиру тяжелую задачу пройти в **пространстве** по нужной траектории к *тому или иному литературному произведению*, а последующие НИОКР, осуществляемые внутри Шпексира, должны определить, к какому именно. Итак, повернув один регулятор, мы оказались в непосредственном соседстве с реальностью, поскольку у нас уже есть примеры впечатляющей работы искусственных авторов, которые Чрезвычайно превосходят ожидания собственных создателей. Никто еще не создал достойного внимания искусственного драматурга, но и искусственный шахматист (*Deep Blue* компании IBM), и искусственный композитор (EMI Дэвида Коупа) добились результатов, которые *в некотором отношении* сравнимы с лучшими образчиками творений человеческого гения.

Кто обыграл чемпиона мира по шахматам Гарри Каспарова? Не Мюррей Кэмпбелл и не один из членов его команды в IBM. Каспарова обыграл компьютер *Deep Blue*. Именно *Deep Blue* выстраивает шахматную партию лучше, чем любой из них. Никто из них не может выиграть в поединке с Каспаровым. *Deep Blue* может. Да, но. Да, но. Возможно, вам хочется подчеркнуть, что, пускай *Deep Blue* и побеждает Каспарова в шахматах, его поиск посредством перебора *совершенно* не похож на рекогносцировочные работы, проводимые Каспаровым при обдумывании хода. Однако это не так – или, по крайней мере, не так в единственном смысле, который имеет значение в контексте этой дискуссии о дарвинистских представлениях о креативности. Мозг Каспарова состоит из органических материалов, а его архитектура по всем ключевым аспектам не похожа на архитектуру *Deep Blue*, и все же, насколько нам известно, этот мозг представляет собой мощный параллельный поисковый движок, со временем накопивший выдающийся массив эвристических техник отсека лишнего, которые не позволяют ему терять время на бесперспективные ответвления. Несомненно, профиль инвестиций в НИОКР в этих случаях различается: Каспаров пользуется методами извлечения удачных расстановок из прошлых игр, благодаря чему распознает и со знанием дела игнорирует обширные сферы игрового пространства, в то время как *Deep Blue* все равно приходится их *систематически* прочесывать. “Проницательность” Каспарова существенным образом меняет порядок предпринимаемого им поиска, но не приводит к появлению “совершенно иного” способа творения. Всякий раз, когда обстоятельный поиск *Deep Blue* выявляет определенный *тип* направления мысли и алгоритмически устанавливает, что им, вероятно, можно пренебречь (это сложная, но выполнимая задача), компьютер получает возможность впоследствии при случае снова использовать эти наработки, точь-в-точь как делает Каспаров. Существенную часть этой аналитической работы провели создатели *Deep Blue*, которые вложили ее результаты в память *Deep Blue*, но Каспарову тоже помогали плоды сотен тысяч человеко-лет исследования шахмат, переданные ему другими шахматистами, тренерами и авторами книг и впоследствии усвоенные его мозгом.

В связи с этим любопытно рассмотреть предложение, которое однажды внес Бобби Фишер. Он предложил восстановить изначальную рациональную чистоту шахмат, перед стартом каждой партии *случайным образом* расставляя старшие белые фигуры в заднем ряду, а старшие черные фигуры расставляя в таком же случайном порядке (зеркально) в заднем ряду на противоположном конце доски (но всегда проверяя, чтобы с каждой стороны было по одному слону на белой и на черной клетке, а король стоял между ладьями). При таком раскладе целая гора заученных дебютов мгновенно стала бы практически

бесполезной как для людей, так и для машин, поскольку шанс сыграть хоть один из них выпадал бы крайне редко. Пришлось бы снова обращаться к фундаментальным принципам и внимательно продумывать каждую партию в реальном времени, под тиканье часов. Возможно, от такого изменения правил выиграли бы люди, но не компьютеры, однако сложно сказать наверняка. Все зависит от того, кто из шахматистов в большей степени полагается на то, что фактически представляет собой механическую память, то есть *с минимальным пониманием* использует НИОКР своих предшественников.

Факт остается фактом: пространство поиска в шахматах настолько велико, что даже *Deep Blue* не успевает основательно изучить его в реальном времени, а потому, подобно Каспарову, отсекает лишние деревья перебора, идя на обдуманный риск, причем этот риск часто оказывается рассчитан заранее, как и у Каспарова. Как Каспаров, так и *Deep Blue*, по-видимому, совершают огромное количество операций перебора, используя совершенно разные архитектуры. В конце концов, что нейроны знают о шахматах? Любая выполняемая *ими* работа – это перебор того или иного рода.

Может показаться, что я выступаю за применение компьютерного подхода, подхода ИИ, поскольку описываю работу мозга Каспарова таким способом, однако эта работа должна осуществляться каким-то способом, а другого способа выполнить эту работу пока никто не описал. Нет смысла утверждать, что Каспаров использует “понимание” или “интуицию”, поскольку это лишь означает, что сам Каспаров не имеет возможности понять, как именно добивается хороших результатов. Таким образом, поскольку никто не знает, как мозг Каспарова делает это (и хуже всех это знает сам Каспаров), пока нет никаких подтверждений тому, что методы Каспарова “совершенно не похожи” на методы *Deep Blue*. Не стоит забывать об этом, когда возникает желание сказать, что методы Каспарова, “само собой”, существенно отличаются. Что вообще может толкнуть человека так рисковать? Необоснованный оптимизм? Страх?

Но это только шахматы, скажете вы, а не искусство. В сравнении с искусством шахматы *тривиальны* (особенно теперь, когда чемпионом мира по шахматам стал компьютер). Здесь в игру вступают “Эксперименты с музыкальным интеллектом”, или ЕМІ (2000, 2001) композитора и программиста Дэвида Коупа. Коуп хотел создать всего лишь устройство для повышения эффективности – компаньон композитора, который помог бы ему справляться с творческими блоками, знакомыми каждому человеку искусства. Это устройство должно было стать высокотехнологичным дополнением к традиционным поисковым аппаратам (пианино, нотной бумаге, магнитофону и т. д.). Возможности ЕМІ постепенно расширялись, и в конце концов программа стала полноценным композитором, осуществляя все больше операций перебора. Если загрузить в ЕМІ музыку Баха, в ответ она генерирует музыкальные композиции в стиле Баха. Если загрузить в нее произведения Моцарта, Шуберта, Пуччини или Скотта Джоппина, система с готовностью анализирует их стили и сочиняет новую музыку в этих стилях лучше самого Коупа – и почти любого настоящего композитора. Если загрузить в ЕМІ музыку двух композиторов, она быстро генерирует пьесы, в которых их стили причудливым образом комбинируются, а если загрузить произведения всех стилей одновременно (не очищая память), система пишет музыку на основании всего своего музыкального “опыта”. Полученные композиции можно снова и снова загружать в ЕМІ вместе с любой другой музыкой в формате MIDI<sup>[64]</sup> – и в результате ЕМІ выработает “личный” музыкальный стиль, корнями уходящий к работам мастеров, но при этом представляющий собой несомненно оригинальное слияние всего этого “опыта”. Сегодня ЕМІ под силу сочинять не только двухчастные инвенции и песни, но и целые симфонии – насколько я знаю, она сочинила уже больше тысячи. Они достаточно хороши, чтобы обмануть специалистов (композиторов и профессоров музыки), и я лично готов засвидетельствовать, что ария ЕМІ-Пуччини заставила меня прослезиться. Но Дэвид Коуп не вправе считать себя композитором симфоний, мотетов и романсов ЕМІ, как и Мюррей Кэмпбелл не вправе утверждать, что обыграл Каспарова в шахматы.

Для дарвиниста этот новый элемент в каскаде **подъемных кранов** – просто последнее слово в долгой истории, и мы должны понимать, что граница между создателями и их артефактами должна быть столь же проницаемой, как и любые другие границы в каскаде. Когда Ричард Докинз (1982) замечает, что плотина наряду с зубами и мехом входит в фенотип бобра – его *расширенный фенотип*, – он закладывает основы для дальнейшего наблюдения о том, что границы человеческого авторства точно так же поддаются расширению. На самом деле мы знали это веками и выводили различные относительно стабильные правила, в соответствии с которыми отличали произведения Рубенса от произведений *мастерской* Рубенса и его учеников. В какой бы сфере люди ни прибегали к помощи, всегда можно задать вопрос, кто кому помогает, что есть создатель и что есть творение.

## 50. Шум в виртуальном отеле

Рассмотрим разницу между виртуальными и реальными мирами. При строительстве реального отеля вам придется потратить немало времени, сил и материалов, чтобы люди из соседних номеров не слышали друг друга. При строительстве виртуального отеля вы получаете эту звукоизоляцию бесплатно. В виртуальном отеле, если вы вдруг захотите, чтобы люди из соседних номеров слышали друг друга, вам придется специально добавить эту возможность. Вам придется добавить *отсутствие* звукоизоляции. Вам также придется добавить тени, запахи, вибрацию, пыль, грязные следы и признаки износа. В реальном, вещественном мире все эти нефункциональные характеристики вы получите бесплатно. Все, что приходится добавлять в виртуальные миры, чтобы сделать их реалистичнее, называется *обнаружением столкновений*. Если вы когда-нибудь пытались написать компьютерную видеоигру, вероятно, вы быстро поняли, что недостаточно заставить формы двигаться по экрану. Если не прописать в корректировочном цикле программы обнаружение столкновений (которое без конца вклинивается в деятельность программного объекта и спрашивает: “Я ни с чем не сталкиваюсь?”), формы будут просто проходить сквозь друг друга без какого-либо эффекта.

В книге *Le Ton beau de Marot* Даг Хофштадтер (1997) обращает внимание на роль того, что он называет *спонтанными вторжениями* в творческий процесс. В реальном мире почти все происходящее оставляет след, отбрасывает тени, имеет аромат, производит шум, и это обеспечивает огромное количество возможностей для спонтанных вторжений. Именно этого не хватает в виртуальном мире. Компьютерные моделисты считают это спокойствие одной из главных прелестей виртуальных миров: в них не происходит ничего, кроме так или иначе запланированного вами. Это позволяет при моделировании начинать с чистого листа и добавлять к своей модели по одной характеристике, чтобы проверить, какая минимальная конфигурация произведет желаемый эффект.

Из-за отсутствия этого шума компьютерные симуляции эволюции получаются в высшей степени ограниченными, поскольку эволюция путем естественного отбора питается шумом, преобразуя полезные фрагменты этого шума в сигналы, мусор в инструменты, а баги – в фичи. Одной из самых впечатляющих компьютерных симуляций эволюции остается ранняя модель “Эволюции виртуальных организмов” Карла Симса (1994), которую можно посмотреть в сети: <http://www.karlsims.com/evolved-virtual-creatures.html>.

Симс начал со *случайных конфигураций* виртуальных шарнирных блоков с виртуальными мускулами для управления шарнирами и позволил им эволюционировать в виртуальном мире, где действуют законы виртуальной физики. Программа автоматически отбирала конфигурации, которые лучше всего проявляли себя, для виртуального спаривания, а затем повторяла цикл с их потомством. Эволюция приводила к появлению все лучших пловцов, ходячих и прыгунов, в создании которых не принимал участия никакой разумный творец. Итоговые конфигурации были совсем не случайными, и это демонстрировало, как эффективно (виртуальная) эволюция вроде как выявляет удачные принципы конструирования и вроде как изобретает заново огромное количество разнообразных характеристик, обнаруживаемых в природе.

Этот великолепный пример показывает, как много может дать *относительно* простая модель, а также очерчивает узкие рамки эволюции виртуального мира. Симс спроектировал простую систему “развития”, которая получает целые геномы и создает новые организмы, однако этот процесс происходит за кулисами моделируемого виртуального мира. В результате случайное столкновение или соударение с фрагментом (виртуального) мусора не может укоротить или удлинить геном или изменить правила экспрессии генов. Всех этих механизмов просто нет в виртуальном мире с виртуальными структурами, а следовательно, он неизменяем. К примеру, в организме Симса не может развиваться новая хромосома. Вся его генетическая система находится за пределами модели и не сталкивается с естественным отбором, а только в принудительном порядке передает генетическую информацию от поколения к поколению. (Другой пример этого явления см. в главе 51.)

В компьютерной модели креативности должен быть мусор, с которым смогут сталкиваться творческие процессы, и шумы, которые они не смогут игнорировать. Спонтанное вторжение незначительного шума из соседней комнаты может изменить ход этих процессов непрогнозируемым и даже деструктивным образом, но в любом случае оно открывает новые возможности. Ключ к творчеству – использование случайностей, будь то при создании нового генома, новой модели поведения или новой мелодии.

Позвольте мне пояснить, чего я *не говорю*. Проблема с эволюцией организмов Симса не в том, что они состоят не из углерода или что в них нет белков и гемоглобина. Проблема в том, что они виртуальны.

И, будучи виртуальными, они живут в мире, который устроен на много порядков проще мира биологической эволюции. То же верно и в отношении ЕМІ Коупа. При всей своей удивительности ЕМІ на несколько порядков проще мира человеческого сочинения музыки. Прелесть этих примеров заключается в том, что они показывают, что именно может дать нечто столь чистое, столь свободное от шума, столь абстрактное.

Можно представить, как улучшить ЕМІ Коупа, модель Симса или любой другой проект в сфере искусственной жизни или искусственного творчества, добавляя в виртуальный мир все больше и больше мусора, все больше возможностей для столкновений. В результате виртуальным организмам придется взаимодействовать с большим количеством виртуальных структур, и заранее никак не угадать, когда произойдет счастливая случайность. Но подумайте, насколько нелогичным покажется такой совет:

Что бы вы ни моделировали, старайтесь, чтобы каждое явление, каждая подпрограмма, каждое событие этого мира оказывало на него ряд нефункциональных воздействий: производило посторонние шумы, оставляло след, мусорило, вызывало вибрации и так далее.

Зачем? Зачем весь этот шум? У него нет четкого назначения: он нужен только для того, чтобы служить потенциальным источником сигналов для всех остальных процессов, которые *могут* обратиться к алхимии творческих алгоритмов и превратить их в функции, искусство, смысл. Каждое усовершенствование конструкции во вселенной начинается со счастливого случая, незапланированного пересечения двух траекторий, которое, как выясняется впоследствии, оказывается не просто столкновением. Но, следуя этому совету, компьютерные моделисты растрачивают производительность, которая делает компьютеры столь полезными инструментами. Возникает гомеостаз. Неудивительно, что результаты компьютерного моделирования креативности все меньше впечатляют. Чтобы подобраться к уровню креативности настоящего композитора, модель должна стать максимально вещественной: в ней должно происходить все больше случайных столкновений, которые будоражат композитора-человека.

## 51. Херб, Элис и малыш Хэл

Покойный эволюционный биолог Джордж Уильямс утверждал, что нельзя отождествлять гены с молекулами ДНК. Делать так – значит, совершать ошибку. Эта ошибка сродни ошибке, которую можно допустить, если сказать, что “Гамлет” состоит из чернил. Само собой, любой экземпляр пьесы Шекспира должен из чего-то состоять (если не из чернил, то, возможно, из очертаний букв на экране компьютера или даже цепочек двоичного кода, записанных на лазерный диск), но сама пьеса представляет собой абстрактную информационную структуру, которая может перемещаться с одного носителя на другой. Будучи рецептами создания белков, гены тоже представляют собой абстрактные информационные структуры, если следовать этой логике – а эта логика всегда казалась мне верной. Но есть и несогласные – все те, кто сомневается в ценности такого представления о генах. Для них – и, в частности, для философа биологии Питера Годфри-Смита – я сконструировал небольшой насос интуиции:

Херб и Элис хотят завести ребенка, но вот каким образом:

1. Оба их генома секвенируются. Они по почте получают файл, в котором их геномы прописаны как две последовательности примерно по три миллиарда букв в каждой: А, С, G, Т...

2. Затем они пишут небольшую компьютерную программу, которая применяет алгоритм деления к обоим геномам и (случайным образом) создает виртуальные сперматозоиды и яйцеклетки, которые затем (случайным образом) объединяются *in silico*, чтобы создать новую спецификацию генома (которая проходит все биоинформатические аналитические проверки как спецификация ДНК ребенка Херба и Элис). (Пока все происходит на уровне символов А, С, G, Т как чисто машинный процесс редактирования цепочки.)

3. На основе этой спецификации кодон за кодоном конструируется фактическое ДНК-воплощение всего генома, где А = аденин, С = цитозин, G = гуанин и Т = тимин. (Сегодня это возможно в лаборатории Крейга Вентера.)

4. Затем этот геном внедряется в ядро человеческой яйцеклетки (имеет ли значение, чья это яйцеклетка, раз ее собственную ДНК удаляют до внедрения ядра?) и становится “младенцем из пробирки” одним из обычных способов.

Можно ли считать этого младенца ребенком Херба и Элис? Мне кажется достаточно очевидным, что Хэл – действительно их биологический потомок, поскольку в него заложена вся генетическая информация, которую они передали бы ему, если бы он был зачат обычным образом. Этот насос интуиции подчеркивает важность того, что имеет значение в репродуктивной информации и при каузальной передаче информации (в этом случае в форме ASCII-кода для “А”, “С”, “G” и “Т”, а не в форме молекул). Каузальная связь может, например, осуществляться через телекоммуникационные спутники, а не по более прямым биохимическим путям. [Из частного письма от 26 апреля 2010 г.]

Годфри-Смит согласился со мной и признал Хэла потомком Херба и Элис, однако усомнился в некоторых моих формулировках (полный вариант см. в работах Dennett 2010; Godfrey-Smith 2011). В том же духе конструктивной критики я подтвердил, что существует биологически важное различие между процессом, который использовали для продолжения рода Херб и Элис, и обычным способом размножения. Что если бы все следовали примеру Херба и Элис? Поскольку генетическая информация от Херба передается в яйцеклетку не традиционным методом – не в сперматозоиде, – *подвижность сперматозоидов* перестает быть критерием для отбора, а потому при прочих равных будет снижаться от поколения к поколению. Используй – или потеряешь. И все же я утверждаю – и думаю, что этот насос интуиции ясно показывает, – что молекулярная структура, которая по большей части не меняется из поколения в поколение, сохраняется *именно потому*, что содержит в себе всю необходимую информацию.

Эволюция беспрепятственно продолжалась бы и если бы существовали другие структуры для хранения информации. Для проверки этого утверждения можно использовать еще один насос интуиции. Допустим, на другой планете “нечетные” поколения использовали спирали ДНК, состоящие из А, С, G и Т, а “четные” поколения – другие двойные спирали, скажем ХНК, состоящие из Р, Q, R и S (некоторых других молекул). Можно предположить, что молекулы ХНК потомства нечетного поколения создавались по образцам родительских молекул ДНК при помощи механизма, подобного матричной РНК, но “осуществляющего перевод” с одного биохимического языка на другой. В следующем поколении эти



сообщения переводились бы обратно другим матричным механизмом – и так далее. Чтобы обзавестись потомством, вам пришлось бы найти партнера, гены которого были бы написаны на том же языке, но гены вашего ребенка были бы написаны на другом языке. Эдиповы союзы были бы бесплодны – что, возможно, и к лучшему, – но случалось бы немало трагедий в духе “Ромео и Джульетты”, поскольку любовники из разных общин были бы неспособны к продолжению рода. (При этом они могли бы довольствоваться бесплодной сексуальной жизнью и усыновлять детей любого типа или даже использовать донорские яйцеклетки и сперматозоиды и растить уйму детишек, которые приходились бы друг другу сводными братьями и сестрами.) В таком мире, несмотря на все эти сложности, эволюция шла бы своим чередом, от поколения к поколению передавая генетическую информацию об удачных адаптациях (а также наследственных заболеваниях и т. п.) с использованием двух систем кодирования, которые могут сколь угодно сильно различаться на структурном уровне. Те же гены, разные молекулы. У каждого гена было бы две формы, которые отличались бы друг от друга, как слово “кот” отличается от слова *cat*, а слово “дом” – от слова *house*. (Обратите внимание на параллель с насосом интуиции о **двух черных ящиках**: в обоих случаях общим для двух синтаксически и структурно различных проводников выступает *одна и та же информация*, одна и та же семантика.)



## 52. Мемы

Я еще не упоминал о мемах, и некоторые читатели могли подумать, что я решил от них отказаться. Вовсе нет. Концепция мемов – один из моих любимых инструментов мышления. Говорить о мемах я могу долго – всей книги не хватит! Я уже не раз высказывался о мемах в других работах (среди прочего см. Dennett 1990, 1991a, 1995a). По разным причинам многим эта концепция категорически не нравится, поэтому они склонны цепляться за многочисленные замечания, высказываемые в ее адрес. Я решил еще раз попытаться отстоять концепцию мемов и ответить критикам – как серьезным критикам, так и оголтелым мемоненавистникам, – но для этого мне понадобится отдельная небольшая книга. Пока же я рекомендую всем тем, кто хочет узнать о мемах больше, обратиться к моему эссе “Новые репликаторы” (2002, также в Dennett 2006a).

Но здесь, в качестве прелюдии, я сделаю краткое введение в *серьезную* концепцию мемов (в отличие от свободного определения этого чрезвычайно популярного термина в интернет-среде). Как заметил Докинз (1976) при введении концепции мема как культурной единицы, которая подвергается копированию, фундаментальный принцип биологии заключается в том,

что все живое эволюционирует благодаря неравномерному выживанию реплицируемых сущностей...

На нашей планете в качестве реплицируемой сущности господствует ген – молекула ДНК. Могут существовать и другие подобные сущности. Если они существуют – и выполняется ряд других условий, – они практически наверняка станут основой процесса эволюции.

Но нужно ли нам отправляться в далекие миры, чтобы найти другие типы репликации, а следовательно, и другие типы эволюции? На мой взгляд, репликатор нового типа недавно возник на этой самой планете. Он прямо перед нами. Пока он только зарождается, неловко барахтаясь в своем первичном бульоне, но уже эволюционирует с такой скоростью, что оставляет старый ген далеко позади. [р. 206]

Этот инструмент мышления позволяет нам сделать два наблюдения, которые существенным образом меняют картину нашего воображения, когда мы думаем о человеческой культуре и креативности. Во-первых, мемы разрушают притягательную в остальном идею о том, что к хорошему проекту ведут лишь две дороги: либо *гены*, либо *гений*. В представлении большинства мыслителей, пока мемы не открывают им глаза, если что-то в человеческой жизни проявляет характерные признаки адаптации для достижения целей или повышения функциональной эффективности, должно быть, это либо продукт *генетического* естественного отбора, либо результат целенаправленного, демонстрирующего понимание и намерение человеческого мышления – *разумного* проектирования. Казалось бы, второй закон Орджела – эволюция умнее вас – закрепляет два этих варианта, но на самом деле есть и третий вариант, распространенный повсеместно: это негенетический, культурный отбор, осуществляемый в процессе того же самого бездумного естественного отбора, который дает нам гены. Прекрасный пример – сделанное более века назад наблюдение о форме полинезийских каноэ: “Каждая лодка делается по образцу предыдущей... само море выделяет лодки, выбирая те из них, которые плавают лучше всех остальных” (Alain 1908). Совершенно очевидно, что это естественный отбор в действии: островитяне следуют простому правилу – если лодка вернулась из моря неповрежденной, делай другую такую же! Возможно, они обладают глубоким пониманием принципов морского строительства, которые задним числом подкрепляют их любимые формы, но в этом нет никакой нужды. Контроль качества осуществляет эволюция. То же самое верно в отношении правил правописания, слов, религиозных практик и многих других основополагающих характеристик человеческой культуры: их *никто* не проектировал, и они не заложены “в наших генах”, но все равно сконструированы великолепно.

Во-вторых, за пользование этим вторым информационным каналом, этой изобильной средой проектирования и передачи информации, недоступной ни одному другому виду, мы платим особую цену: мемы *обладают собственной приспособляемостью*, как и все остальные симбионты, процветающие в нашей компании, и их приспособляемость в некоторой степени независима от нашей. Почти никто не задумывается об этом, что особенно очевидно при обсуждении эволюционной теории религии. “О, так вы работаете над эволюционной теорией религии? И что, *по-вашему*, дают религии? *Должны же* они чему-то благоприятствовать, раз в каждой человеческой культуре есть религия в той или иной форме”. Что

ж, в каждой человеческой религии есть и простуда. Чему благоприятствует она? Только самой себе. Мы должны быть готовы находить культурные репликаторы, которые не приносят никакой пользы, но все равно прекрасно себя чувствуют. Это ставит в одинаковые условия все теории культурной эволюции, заменяя зашоренное представление о том, что культурные инновации – как и генетические инновации – всегда повышают приспособляемость тех, кто их передает. Мемы – это информационные симбионты. Мы не можем жить без них, как не можем жить и без триллионов других населяющих нас симбионтов, но это не значит, что все они дружелюбны нам. Часть из них лишь создает нам проблемы, без которых мы вполне могли бы обойтись.

## Резюме

В этом разделе я постарался продемонстрировать, что дарвинистское мышление действительно сравнимо с универсальной кислотой: оно переворачивает весь традиционный мир с ног на голову, бросая вызов представлению об осуществлении проектирования по нисходящей траектории от гения гениев, разумного творца, и заменяя его представлением о работе по восходящей траектории, осуществляемой бездумными, немотивированными, циклическими процессами, которые создают все более сложные комбинации, пока те не начинают реплицироваться самостоятельно, ускоряя процесс проектирования посредством многократного использования всех лучших элементов. Некоторые из этих ранних потомков в итоге объединяют усилия (первый подъемный кран, симбиоз), что приводит к многоклеточности (второй подъемный кран), которая обуславливает появление более эффективных агентов-исследователей благодаря половому размножению (третий подъемный кран), и это ведет к зарождению языка и культурной эволюции одного из видов (и снова краны), что создает среду для появления новейших кранов – литературы, науки и инженерии, а они, в свою очередь, позволяют нам “выйти на метауровень”, недоступный никакой другой форме жизни, где мы получаем возможность по-разному размышлять, кто мы, что мы и как мы появились в этом мире, моделируя эти процессы в пьесах и романах, теориях и компьютерных симуляциях, а также используя множество других инструментов мышления из нашего впечатляющего инструментария.

Это представление настолько универсально и вместе с тем так изобилует деталями, что само по себе может считаться мощным инструментом. Всем тем, кто по-прежнему не приемлет дарвинистское мышление, стоит задуматься, что при использовании одних лишь старомодных инструментов они окажутся за бортом передовых исследований множества важных и столь разнообразных феноменов, как эпидемии и эпистемология, биотопливо и архитектура мозга, молекулярная генетика, музыка и мораль.

## VII.

### Инструменты мышления о сознании

Вооруженные десятками инструментов мышления, мы наконец подходим к теме, которую многие считают самым таинственным феноменом во всей вселенной. Ее даже не раз провозглашали неразрешимой загадкой. Мы никогда не познаем сознание, утверждают эти ученые: его природа будет до конца времен систематически ускользать от нашего понимания, какие бы попытки изучить ее ни предпринимались в науке и философии. Поскольку нет *достаточных* оснований верить в существование этого интеллектуального барьера, я прихожу к выводу, что это не что иное, как самовнушение. Некоторым не нравится, что рано или поздно мы можем раскрыть секрет работы сознательного разума, а потому, чтобы мы не навязывали им свои представления, они твердят, что нам лучше сдаться перед лицом неразрешимой проблемы. Если мы последуем их совету, они окажутся правы, поэтому давайте не будем обращать на них внимания и возьмемся за эту сложную, но все же выполнимую задачу.

### 53. Два контробраза

Многие описанные ранее инструменты мышления так или иначе имели отношение к сознанию – убеждениям, мышлению и так далее, – но запутанные проблемы сознания я до сих пор обходил стороной. На то есть причина: размышляя о сознании, люди склонны раздувать свои представления о сознании и тем самым одурачивать себя. Они берутся за самые сложные проблемы, не давая себе шанса оценить, какую часть работы (и игр) сознания (*mind*) можно объяснить, не поднимая извечных вопросов о сознательном опыте. Мы уже разбили базовый лагерь – так не пора ли нам покорять вершину? Пора. Однако, позволяя себе такую мысль, мы уже совершаем ошибку! Самосознание (*consciousness*) нельзя назвать единственным великолепным пиком нашего сознания (*mind*). Вопреки традиции, восходящей, по меньшей мере, к Декарту, жившему в семнадцатом веке, феномены самосознания не занимают в нашем сознании ни “центрального”, ни “высшего” положения (Jackendoff 1987; Dennett 1991a). Чтобы нейтрализовать привлекательный, но неудачный образ, нужно использовать контробраз, поэтому начнем с простого калибратора воображения: вспомним прекрасную песню Коула Портера “Ты – вершина” и задумаемся, что вы, возможно, *не вершина* как таковая – не высшая точка горы, а вся гора, и потому ваши знания о горе, которой вы сами и являетесь, не ограничиваются видом с пика, ведь вам открываются и все панорамы с ее склонов. Феномены сознания можно сравнить с волосами, которые обрамляют лысину. Не забывайте об этом.

Вот другой контробраз: сознание *не похоже* на проводник вроде телевизора, в который может передаваться и записываться информация, и в мозге нет места, где “все сходится воедино” перед лицом некоего центрального свидетеля, – я называю это воображаемое место картезианским театром (Dennett 1991a). Сознание *скорее сродни славе, чем телевизору*: славе в мозге, церебральной звездности, которая позволяет некоторым фрагментам содержимого становиться влиятельнее и памятнее конкурентов. Вместо того чтобы отстаивать эту точку зрения (аргументы в ее поддержку см. в работах Dennett 1991a, 2005b), я просто предлагаю вам этот инструмент мышления. Не нравится – не берите. Но я все же дам вам дружеский совет: всякий раз, когда будете представлять проникновение в сознание как прибытие в штаб-квартиру сознания или как перевод с языка бессознательных нейронных сигналов на какой-то другой язык, вспоминайте эти контробразы и проверяйте, не вводите ли вы себя в заблуждение.

## 54. Чутье на зомби

Большинство людей чувят – именно чувят, иначе и не скажешь, – что ни один робот (сделанный из кремния, металла, пластика и т. п.) не может обладать сознанием в том смысле, в котором им обладает человек. Есть в наших живых, дышащих, органических телах и мозгах что-то такое, без чего сознанию не обойтись. Прийти к этому интуитивному пониманию можно и без использования насосов – настолько широко оно распространено, – и эти люди вполне могут оказаться правы. Но теперь мы знаем, что наше тело и мозг можно представить в качестве роботов, состоящих из роботов, состоящих из роботов – и так далее, до субнейронного уровня, где на благо системы трудятся двигательные белки и другие наноботы, – а потому можем предположить, что это чутье не более чем артефакт бедного воображения: люди просто представляют себе роботов, устроенных на много порядков проще. Один мой друг однажды попробовал убить эту идею в зародыше: “Я просто не могу постичь сознательного робота!” Чепуха, ответил я. Дело в том, что ты *не хочешь* постичь сознательного робота! Ты думаешь, что эта идея слишком глупа и нелепа, чтобы принимать ее всерьез. Но постичь сознательного робота под силу даже ребенку – представляют же дети сознательный паровоз (“Паровозик, который верил в себя”) или сознательную новогоднюю елку (во всех этих слезливых детских сказках об одиноких елках, которые мечтают найти дом). Любый смотревший “Звездные войны” часа полтора представлял R2D2 и C3PO сознательными. Мы делаем это с детства, обычно “даже не задумываясь”. Это не только просто, но и почти неизбежно, когда мы сталкиваемся с объектом, который ведет себя – и особенно говорит, – как человек.

Вот интересный факт: с тех пор, как в 1950-х гг. Уайлдер Пенфилд провел свои прорывные исследования в Монреале, состоялось множество операций на мозге пребывающих в сознании пациентов, которые могли сказать, что ощущают, когда их мозг стимулируется *здесь* или *вон там*. Вряд ли кто-то из участников или свидетелей любой из этих операций хоть раз подумал: “О боже! Это не человек! Это зомби! Иначе и быть не может, ведь мы заглядываем внутрь, а там одно серое вещество!” Нет, ведь было слишком очевидно – смотрите! слушайте! – что пациент в сознании. На самом деле было бы столь же очевидно, если бы мы вскрыли череп нашего собеседника и обнаружили, что внутричерепная полость заполнена микрочипами. Мы бы узнали – возможно, к собственному удивлению, – что нам не только просто вообразить или постичь сознательного робота, но и что он существует на самом деле.

Некоторые философы полагают, что воображение сыграет с вами злую шутку, если вы купитесь на это “исключительно поведенческое” свидетельство сознательности и придете к такому выводу. “Не позволяйте себя обмануть!” – таков, возможно, их девиз. *Доказать*, что другой человек сознателен, гораздо сложнее, поскольку существует возможность – по крайней мере, логическая, – что этот человек на самом деле “зомби”. Не вуду-зомби, которых показывают в кино и в костюмы которых наряжаются на Хеллоуин. Ходячих мертвецов легко отличить от обычных людей по поведению (и жуткому внешнему виду). Философские зомби, напротив, могут быть приятны в общении. Они вполне могут становиться душой компании, любить, радоваться жизни и совершать спонтанные поступки, как и любой из ваших знакомых. Зомби могут быть и некоторые из ваших лучших друзей. Философские зомби (по определению) на поведенческом уровне неотличимы от обычных сознательных людей, но при этом “пустоголовы” – напрочь лишены внутреннего мира и сознательного опыта. Они кажутся сознательными только внешне. Если вы согласны с этими философами и считаете это серьезной проблемой, если вы гадаете – учитывая логическую возможность существования философских зомби, – как вообще может родиться научная, материалистическая теория сознания, то вы находитесь в плену *чутья на зомби*<sup>[65]</sup>.

Позвольте мне сразу признать, что у меня, как и у любого другого человека, тоже есть чутье на зомби. Если задуматься, *действительно* кажется, что сознание должно представлять собой некое дополнение ко всему тому, что оно делает для нас и с нами, некое особое, интимное ощущение, своего рода *самоосознание*, которое не может быть свойственно ни одному роботу и совершенно невообразимо в качестве “всего лишь” физической активности мозга. Но я научился не обращать внимания на это чутье. Я считаю его источником ошибок, обманом воображения, а не наставлением на путь истинный. Тем не менее убедить в этом остальных не так-то просто. Чтобы ослабить чутье на зомби, нам придется использовать несколько насосов интуиции.

Для начала сравним эту логическую возможность с рядом других. *Логически* возможно, что вы живете в Матрице, а вся жизнь, которую вы наблюдаете и в которой, очевидно, участвуете, на самом деле представляет собой виртуальную реальность, созданную специально, чтобы обеспечить вам покой, пока ваше тело лежит без движения в какой-то высокотехнологичной капсуле. *Логически* возможно, что атомов

уплерода не существует, а те объекты, которые ученые считают атомами уплерода, на самом деле представляют собой огромное множество крошечных космических кораблей, пилотируемых пришельцами, всю жизнь притворяющимися атомами уплерода. *Логически* возможно, что вся вселенная была создана около шести тысяч лет назад – вместе с так называемыми ископаемыми и фотонами, якобы летящими к нам из далеких галактик. (*Логически* возможно, что мир был создан всего десять минут назад, а мнимые воспоминания о прошлом были просто внедрены в ваш мозг.) Подобные логические возможности открывают нам широкие горизонты для создания художественных произведений, однако мы не считаем их серьезными указаниями на то, что нам следует пересмотреть нашу физику, химию и биологию или вовсе отказаться от их достижений. Есть ли основания считать чутье на зомби более весомым, более достойным рассмотрения? Многие серьезные мыслители полагают, что да.

Предок всех насосов интуиции, сконструированных для отказа от чего-то вроде чутья на зомби, был предложен сотни лет назад Готфридом Вильгельмом Лейбницем, философом и математиком, который делит лавры изобретателя математического анализа с Исааком Ньютоном. Он был столь же умен и неординарен, как и остальные мыслители его эпохи, но все равно купился на следующий насос интуиции, предложенный им самим.

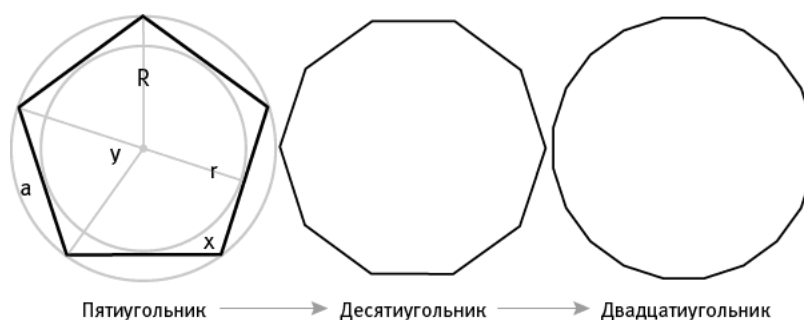
Если допустить существование машины, сконструированной таким образом, чтобы она могла думать, чувствовать и воспринимать, можно вообразить, что ее размеры будут больше при соблюдении тех же пропорций, чтобы внутрь нее можно было войти, как на мельницу. В таком случае, изучая ее интерьер, мы найдем только части, которые работают в связке друг с другом, но не обнаружим ничего, что объясняло бы способность к восприятию. *Следовательно* [курсив мой], восприятие следует искать в простой субстанции, а не в сложной конструкции и не в машине. [Leibniz 1714, § 17]

Это “следовательно” – одна из плавных логических ошибок во всей истории философии. Лейбниц не дает нам никаких промежуточных аргументов для обоснования этого вывода; он полагает, что вывод слишком очевиден, чтобы его хоть чем-нибудь подкреплять. Вспомните генетика начала двадцатого века Уильяма Бэтсона, который просто не мог представить гены в качестве материальных сущностей (см. с. 122). Подобно Бэтсону, неспособному принять всерьез сумасбродную идею о трех миллиардах спаренных оснований в двойной спирали внутри каждой клетки (уму непостижимо!), Лейбниц не мог принять всерьез идею о “мельнице” с *триллионами* подвижных частей. Без сомнения, он настаивал бы, что “простым добавлением частей” от машины к сознанию не придешь, но это было бы лишь его чутье, которое он никак не мог доказать. Но если Дарвин, Крик и Уотсон разоблачили обман воображения Бэтсона, то Тьюринг отправил спроектированный Лейбницем насос интуиции в утиль. Вот только это не так. Пока что. Думаю, со временем чутье на зомби станет историей, любопытным артефактом нашего населенного духами прошлого, но сомневаюсь, что настанет день, когда оно исчезнет совсем. Оно не выживет в текущей, дурманящей разум форме, но сохранится в качестве менее агрессивной мутации, все еще психологически могущественной, но лишенной власти. Такое случалось прежде. По-прежнему *кажется*, что земля стоит на месте, пока луна и солнце вращаются вокруг, но мы узнали, что это только видимость, которую не стоит принимать во внимание. По-прежнему *кажется*, что есть разница между объектом, пребывающим в абсолютном покое, и объектом, который просто не ускоряется в инерциальной системе отсчета, но мы научились не доверять этому чувству. Я жду тот день, когда философы, ученые и простые обыватели будут посмеиваться над ископаемыми следами нашего раннего недоумения по вопросам сознания: “По-прежнему *кажется*, что эти механистические теории сознания чего-то не учитывают, но теперь мы, само собой, понимаем, что это лишь иллюзия. На самом деле они объясняют о сознании все, что требует объяснения”.

Упрямое доверие чутью на зомби поддерживается многими философскими мысленными экспериментами, такими как знаменитый эксперимент Джона Сёрла о **китайской комнате**, вдохновивший меня на создание термина “насос интуиции”. Вскоре он будет развенчан у вас на глазах. Но сначала я хочу немного более подробно изучить концепцию философских зомби.

## 55. Зомби и зимбо

Когда люди говорят, что не могут постичь (философских) зомби, мы вправе спросить их, откуда они это знают. Постигать не так уж просто! Можете ли вы постичь более трех измерений? Искривление пространства? Квантовую запутанность? Просто вообразить что-то недостаточно – Декарт даже говорит нам, что вообразить не значит постичь! Согласно Декарту, воображая, мы используем все свое (механистическое по сути) тело, со всеми его ограничениями (близорукостью, ограниченным разрешением глаза, углами зрения и глубиной фокуса), в то время как постижение использует лишь разум, который гораздо лучше справляется с распознаванием различий, поскольку его не сковывают механические рамки. В качестве примера, убедительно доказывающего существование этого различия, Декарт приводит тысячеугольник. Можете ли вы его вообразить? А постичь? В чем разница? Давайте первым делом попробуем его вообразить. Начнем, скажем, с пятиугольника, а затем вообразим десятиугольник. Это непросто, но вы знаете, что делать: нужно согнуть каждую сторону пятиугольника посередине и вытолкнуть наружу, тем самым превратив пять равных сторон в десять. Насколько далеко толкать? Просто впишите пятиугольник в окружность и выталкивайте новые стороны по направлению к окружности, пока углы ее не коснутся. Затем повторите операцию, чтобы сделать двадцатиугольник.



Повторив операцию еще несколько раз, вы получите правильный 1280-угольник, который почти неотличим от окружности *в воображении*, но *при попытке постичь его* отличается от окружности – и от тысячеугольника – так же сильно, как от окружности отличается квадрат. Если я попрошу вас вообразить окружность внутри тысячеугольника внутри окружности внутри тысячеугольника внутри окружности, то есть своего рода мишень, сможете ли вы отличить окружности от тысячеугольников *в своем мысленном образе*? Нет, все они будут казаться окружностями, но вам не составит труда *постичь* то, что вас попросили.

Декарт не требует производить такие выкладки: в его представлении постижение, как и воображение, – непосредственный и эпизодический мысленный акт, в результате которого человек *схватывает идею, не создавая ее образа*, или вроде того. Вы просто улавливаете (мысленно) значение нужных понятий (СТОРОНА, ТЫСЯЧА, ПРАВИЛЬНЫЙ, МНОГОУГОЛЬНИК) – и *вуаля!* Вы все поняли. Я всегда с недоверием относился к этому базовому картезианскому акту *постижения*. Если вам он под силу, прекрасно, но лично я с ним никак не справляюсь. Я не чувствую уверенности, что *преуспел* в постижении чего-либо, пока не поиграю некоторое время с соответствующими идеями, мысленно проверяя их следствия и, по сути, выполняя упражнения, чтобы в полной мере овладеть задействованными инструментами. (А занимаясь этой умственной гимнастикой, я активно использую воображение – например, изучаю различные диаграммы и образы, которые возникают у меня в голове. Иными словами, я эксплуатирую то, что Декарт назвал бы лишь моим воображением, чтобы добиться того, что он считает постижением.) Можете ли вы постичь теорию струн? Считаете ли вы, что понять и проверить на логическую непротиворечивость все разговоры о многочисленных измерениях, заполненных суперструнами, “мозгами” и тому подобными вещами, не представляет труда? Мне они кажутся *заумными*, но именно по этой причине я не готов объявить их непостижимыми или невозможными (Ross 2013). Меня они не убеждают, однако я недостаточно уверен в собственных способностях к постижению, чтобы отбросить их как полную чепуху. Я *пока* не сумел постигнуть истину теории струн. Нам не следует придавать особенного значения легковесным вердиктам о постижимости или непостижимости в отсутствие наглядных доказательств. Бэтсон сказал, что существование материального гена



“непостижимо”, однако, будь он сегодня жив, он без проблем сумел бы его постичь. В конце концов, о двойной спирали со всеми ее витками теперь рассказывают в школе – и этот феномен оказывается вполне постижимым для детей, как только они уловят его суть. Но никакая новая информация и никакие новые техники воображения не помогут нам постичь круглый квадрат (правильный четырехугольник, все точки сторон которого равноудалены от его центра) или самое большое простое число.

Я вполне уверен, что идея философского зомби концептуально непоследовательна, невозможна и несостоятельна. Но не стоит верить мне на слово. Что вы можете сделать, чтобы убедить себя, что вам *под силу* постичь философского зомби? Допустим, вы попытаетесь представить, что ваш друг Зик “оказался” зомби. Что убедит вас или даже подтолкнет к такому выводу?<sup>[66]</sup> Какое отличие станет решающим? Не забывайте, никакое *действие* Зика не может убедить вас, что он зомби или не зомби. Я замечая, что многие выполняют это упражнение *неправильно*: пытаясь постичь происходящее, они, к несчастью, забывают или отбрасывают часть определения философского зомби. Заметить эту ошибку станет проще, если выделить специальный подвид зомби, которых я называю зимбо (Dennett 1991a). Все зомби обладают бессознательными (само собой) системами управления, которые извлекают информацию о мире (через глаза и уши зомби) и используют ее, чтобы не наткнуться на стены, поворачиваться на зов и так далее. Иными словами, все они представляют собой интенциональные системы. Но зимбо выделяется из общей массы, поскольку это зомби, который также наделен всем необходимым для наблюдения за собственной активностью, как внешней, так и внутренней, а потому располагает внутренними (бессознательными) информационными состояниями высшего порядка, *описывающими* все остальные его внутренние состояния. Дальнейший самомониторинг позволяет зимбо получать и использовать информацию об этих состояниях самомониторинга – и так далее до бесконечности. Иначе говоря, зимбо наделен *рекурсивной саморепрезентацией* – *бессознательной* рекурсивной саморепрезентацией, если вы меня понимаете. Только благодаря этому особому таланту зимбо может быть участником подобного разговора:

Вы: Зик, я тебе нравлюсь?

Зик: Конечно. Ты мой лучший друг!

Вы: Тебе не понравилось, что я об этом спросил?

Зик: Честно говоря, нет. Вопрос получился немного обидным. Мне стало от него не по себе.

Вы: Откуда ты знаешь?

Зик: Хм-м... Я просто помню, что почувствовал раздражение, угрозу, а может, просто удивление, когда услышал этот вопрос из твоих уст. Почему ты меня об этом спросил?

Вы: Вопросы буду задавать я, если не возражаешь.

Зик: Как скажешь. Мне вообще не нравится этот разговор.

Не забывайте, поскольку философский зомби неотличим от сознательного человека на поведенческом уровне, в его репертуар входит поведение вроде поддержания этой беседы, а чтобы контролировать подобные паттерны поведения, зомби понадобится рекурсивная саморепрезентация. Зомби может “размышлять” (характерным для зомби бессознательным образом) о том, что он чувствует в отношении того, что чувствовал в отношении того, о чем думал, когда задавался вопросом... и так далее. Довольно просто представить, что у вас возникнут подозрения, если Зик вдруг зависнет, когда вы начнете расспрашивать его подобным образом, но в таком случае вы просто узнаете, что если Зик и зомби, то не зимбо. Спрашивая, возможно ли существование философских зомби, вы всегда должны следить, что на самом деле думаете о зимбо, поскольку лишь существо, располагающее рекурсивной саморепрезентацией, сможет сохранять самообладание в будничных взаимодействиях вроде этого разговора, не говоря уже о сочинении стихов, формулировке новых научных гипотез и игры в спектаклях, а все эти действия по определению входят в компетенцию зимбо.

Если только вы не вообразили в мельчайших подробностях, насколько неотличим был бы “нормальный” Зик от зимбо Зика, вы не попытались по-настоящему постичь философского зомби. Подобно Лейбницу, вы сдаетесь, даже не попытавшись. Теперь задайте себе еще несколько вопросов. Какая вам разница, зимбо ли Зик? Или – на более личном уровне – какая вам разница, являетесь ли (или станете ли) зимбо вы сами? Вам ведь все равно этого никогда не узнать.

Правда? Есть ли у Зика убеждения? Или же у него есть только *вроде как* убеждения – “своего рода информационные состояния за вычетом сознания, которые ведут зимбо по жизни точно так же, как убеждения ведут по жизни всех нас”? Вот только в этом случае *вроде как* убеждения столь же действенны, столь же надежны, как и настоящие, поэтому использовать оператор “вроде как” здесь неуместно.

Продemonстрировать это можно, вообразив левшей (как я, страдающих от диспраксии) в качестве зимбо, а правшей – в качестве сознательных людей.

Левша: Говоришь, ты доказал, что мы, левши, на самом деле зомби? Я бы в жизни не подумал! Но почему же мы бедняги?

Правша: Вы по определению лишены сознания – что может быть хуже?

Левша: Хуже для кого? Что страдать, если голова пустая? Но ты-то зачем пытаешься со мной поговорить, раз я всего лишь зимбо?

Правша: *Мне кажется*, что голова у тебя не пустая.

Левша: И мне так кажется! В конце концов, я зимбо, а следовательно, обладаю всевозможными способностями к самомониторингу высшего порядка. Я знаю, когда я в смятении, когда мне больно, когда мне скучно, когда мне интересно и так далее.

Правша: Нет. Ты функционируешь так, *словно бы* знаешь эти вещи, но на самом деле ничего не *знаешь*. Ты лишь *вроде как* знаешь все это.

Левша: Думаю, здесь оператор “вроде как” неуместен. То, что ты называешь моим *вроде как* знанием, неотлично от твоего так называемого *реального* знания, если не считать “дефиниционный” аспект, по которому знание зимбо не считается реальным.

Правша: Но разница *есть* – *должна быть*!

Левша: Что это, если не предрассудки?

Если этого недостаточно, чтобы вы представили, каково дружить с зимбо, рассмотрите другие примеры. Представьте, что хотите написать роман о зимбо, застрявшем в мире сознательных людей, или о сознательном человеке, выброшенном на Остров Зимбо. Какие детали вы можете выдумать, чтобы история стала правдоподобной? Можно выбрать и путь попроще: прочитайте хороший роман с мыслью, что это роман о зимбо. Что подтверждает или опровергает вашу гипотезу? При создании романа писатели выбирают точку зрения, или *режим повествования*. Одни могут излагать события *от первого лица*, как поступили Герман Мелвилл в “Моби Дике” и Дж. Д. Сэлинджер в “Над пропастью во ржи”.

“Зовите меня Измаил”.

“А уж если я волнуюсь, так это не притворство. Мне даже хочется в уборную, когда я волнуюсь. Но я не иду. Волнуюсь, оттого и не иду”<sup>[67]</sup>.

Другие писатели выбирают *повествование от третьего лица*, где рассказчику известно все. Любопытно, что при повествовании от первого лица подтвердить гипотезу о зимбо, казалось бы, проще. В конце концов, вся история просто отражает повествовательное *поведение* зимбо Измаила или зимбо Холдена Колфилда. Мы видим их только со стороны и узнаем лишь то, что *они сами называют* проявлениями своего внутреннего мира! Сравните эти повествования от первого лица с повествованиями от третьего лица, например фрагментами из “Доводов рассудка” Джейн Остин и “Преступления и наказания” Ф. М. Достоевского.

Она [Элизабет] чувствовала, что следовало бы пригласить миссис Мазгроув и спутников ее на обед; но она не могла снести мысли, что в продолжение обеда перемена в обстоятельствах, сокращение штата прислуги неизбежно откроются тем, кто всегда смотрел снизу вверх на Эллиотов из Киллинча. Суетность боролась с приличием, и суетность победила, и тотчас Элизабет вздохнула с облегчением<sup>[68]</sup>.

Он [Раскольников] смотрел на Соню и чувствовал, как много на нем было ее любви, и странно, ему стало вдруг тяжело и больно, что его так любят. Да, это было странное и ужасное ощущение!

Казалось бы, здесь автор позволяет нам “заглянуть прямо в сознание” Элизабет и Раскольникова, а потому разве могут они быть зимбо? Но не забывайте, где у сознательных людей поток сознания, у зимбо – поток *бессознательного*! В конце концов, в зимбо нет ничего чудесного: их поведение контролируется множеством внутренних процессов невероятной информационной сложности и модулируется практическими аналогами эмоций, соответствующими счастью, смятению и боли. Таким

образом, Элизабет и Раскольников могут быть зимбо, а Остин и Достоевский просто используют знакомые и любимые всем нам по **народной психологии** термины для описания их внутренних процессов, подобно тому как программисты говорят об итерационном “поиске” и рискованных “суждениях” создаваемых ими шахматных программ. Зимбо может стыдиться потери положения в обществе и может задыхаться от любви.

Никогда не забывайте, как подвело воображение Уильяма Бэтсона. Когда я изо всех сил стараюсь не попасть в эту ловушку, выискивая просчеты в своих допущениях и пытаюсь понять, где я мог ошибиться насчет зомби, я всегда совершаю воображаемые открытия, которые показывают – в лучшем случае, – что концепция сознания вообще весьма сумбурна. К примеру, я могу вообразить, что существует два (или семь, или девяносто девять) различных типа так называемого сознания, причем для левой характерен один, для правой другой, а для омаров – третий. Но вообразить это можно (пока что) лишь одним способом – я должен представить, что *они отличимы по следующим функциональным критериям*: левши не могут Х, правши не могут Y и так далее. Но эти различимые различия лишь показывают, что мы вообще говорим не о философских зомби, потому что (по определению) философского зомби не отличить от “истинно сознательного” человека на основании внешних критериев. При этом никто еще не сумел описать внутреннее отличие истинного сознания, в основе которого не лежит *способность* предположительно сознательного человека *делать что-либо* на уровне “психики”, тем самым убеждая нас (и себя самого) в своей сознательности. Но каким бы ни было это психическое различие, вероятно, в “поток бессознательного” зомби найдется и его фальшивый аналог. Если нет, то почему? Итак, я вполне уверен, что концепция философского зомби – своего рода интеллектуальная галлюцинация, недуг, который можно перерасти. Попробуйте. Далее в этом разделе я помогу вам справиться с этой задачей и пересмотреть свои представления.

## 56. Проклятие цветной капусты

Вижу, вы облизываетесь при виде только что сваренной цветной капусты, от одного запаха которой меня начинает тошнить. Мне сложно понять, как вам вообще может нравиться *этот вкус*, а потому я подозреваю, что вы, вероятно, ощущаете вкус цветной капусты иначе, чем ощущаю его я. Казалось бы, эта гипотеза вполне правдоподобна, ведь я прекрасно знаю, что в разное время я по-разному ощущаю вкус одной и той же пищи. К примеру, первый поток апельсинового сока за завтраком кажется гораздо более сладким, чем второй, если между ними я съем кусок блинчика с кленовым сиропом, но после пары плотков кофе апельсиновый сок снова становится (примерно? точно?) таким же, каким он был при первом плотке. **Безусловно** (динь!), нам хочется говорить (или думать) о таких вещах, и, безусловно (динь!), мы не слишком ошибаемся, когда делаем это, поэтому... безусловно (динь!), вполне нормально говорить о том, как *Деннет ощущает вкус сока в момент времени  $t$* , и спрашивать, так же или иначе *Деннет ощущает вкус сока в момент времени  $t'$*  или *Джонс ощущает вкус сока в момент времени  $t$* . Назовем то, какими вещи предстают перед нами, термином *квалиа* (*qualia*).

Этот “вывод” кажется невинным, но мы уже совершили большую ошибку. Подразумевается, что на последнем этапе мы можем отделить “квалиа” от всего остального – хотя бы теоретически. *То, как  $X$  ощущает вкус сока*, предположительно можно отличить от всех сопутствующих, второстепенных факторов, или побочных продуктов этого “центрального” ощущения. Можно смутно представить, как в каждом из случаев все лишнее постепенно отбрасывается, чтобы самое главное – то, как разные индивиды видят, слышат, чувствуют вещи, а также ощущают их вкус и запах в разное время, – оказалось отделено от того, что оказывает влияние на этих индивидов и стимулирует их нечувственное восприятие, и от того, как они себя впоследствии ведут и во что решают верить. Главная ошибка не в том, что мы допускаем возможность хотя бы иногда или всегда проводить такую очистку на практике, а в том, что вне зависимости от успешности попыток этой очистки нам кажется, будто существует некоторое остаточное свойство.

Искушающих нас примеров не перечислить. Я не могу вообразить, никогда не узнаю, да и не могу, пожалуй, узнать, как слышал Баха Гленн Гульд. (Я едва могу вспомнить, как сам слышал Баха в детстве.) И мне, пожалуй, не узнать, каково быть летучей мышью (Nagel 1974) или видите ли вы то же самое, что и я, смотря на чистое “голубое” небо. Эти обыденные примеры убеждают нас в реальности этих особенных свойств – субъективных вкусов, видов, ароматов, звуков, – которые мы затем, очевидно, изолируем для изучения, прибегая к подобной философской дистилляции. Так и рождаются квалиа.

“Квалиа” – это “технический” термин для обозначения того, что всем нам прекрасно знакомо: того, *как вещи выглядят для нас*. Казалось бы, собственные квалиа должны быть известны вам лучше всего на свете: пусть вселенная окажется лишь гигантской иллюзией, лишь выдумкой злого демона Декарта, но то, *из чего состоит* эта выдумка, будет (для вас) *квалиа* вашего галлюцинаторного опыта. Декарт подвергал сомнению все, в чем вообще можно сомневаться, но ни разу он не усомнился в том, что его сознательные ощущения имеют квалиа, то есть свойства, по которым он может их узнавать и понимать.

Такое определение квалиа – *то, как вещи выглядят для нас*, – может показаться вполне ясным, но, хотя описанные выше квалиа уже анализировались и обсуждались философами, между ними по-прежнему нет согласия, что именно означает или, технически говоря, подразумевает этот термин. Многие специалисты по когнитивной науке великодушно допустили, что философы знают, о чем говорят, когда используют этот специальный термин, и добавили его в свой лексикон, тем самым совершив тактическую ошибку. Безотносительно эмпирической подоплеки, до сих пор не утихают споры о том, чем являются и не являются квалиа. Несколько лет назад я (1988b) опубликовал эссе, в котором перечислил четыре основополагающих свойства квалиа. Квалиа – то, как вещи выглядят для нас, – должны быть

1. невыразимыми,
2. внутренне присущими,
3. частными, а также
4. непосредственно воспринимаемыми.

Таким образом, они (1) неделимы для интроспекции, а следовательно, неопишутемы (“вам нужно быть там”); (2) не связаны ни с какими отношениями, диспозициями или функциями (красный цвет может вызывать тревогу у некоторых людей, однако эту субъективную диспозицию нельзя считать квалиа

красного); (3) “Вам нужно быть там, но у вас не получится, они мои и только мои!” и (4) ваши квалиа знакомы вам лучше всего остального.

В большинстве научных кругов эти свойства по-прежнему считаются хорошей отправной точкой для анализа квалиа, однако цель моего эссе состояла в том, чтобы показать, что ничто не может соответствовать всем четырем перечисленным критериям, а потому обсуждалась возможность пересмотра и доработки концепции, но консенсус так и не был достигнут. Широко применимый и высоко котирующийся технический термин нередко имеет несколько несовместимых определений – вспомните “ген” и “вид” в биологии или “причину” во всей остальной науке, – но мне представляется, что неразбериха с “квалиа” гораздо проблематичнее. Представители других дисциплин хватаются за эту концепцию, считая ее подарком от философии, который может пригодиться и в их исследованиях, но в итоге получают троянского коня.

В том эссе я предложил тринадцать других насосов интуиции (в дополнение к цветной капусте) и не буду повторяться здесь, поскольку в последующие годы я сконструировал другие, возможно более действенные, инструменты, которые можно использовать в моей битве с самонадеянностью, нашедшей выражение в знаменитом ответе на вопрос, что же такое квалиа. Нед Блок (1978, р. 281) отмахнулся от этого извечного вопроса “лишь вполуху”, вспомнив легендарный ответ Луи Армстронга на вопрос, что такое джаз: “Если вам приходится спрашивать, вам никогда этого не узнать”. Эта любопытная тактика прекрасно демонстрирует, какую именно гипотезу я хочу развенчать. Если у меня все получится, ответ Блока, который большинство и сегодня считает образцовым, будет казаться столь же нелепым и необоснованным, как и комичное удивление виталиста, заявившего при встрече с человеком – “заметьте, живым существом!” – что он сомневается в существовании *жизненного порыва*.

## 57. Жизненная сила: сколько это в “настоящих деньгах”?

Согласно распространенному мнению, даже если бы можно было создать, скажем, роботизированную модель цветового зрения, которая демонстрировала бы все знакомые нам, людям, феномены, такие как дополнительные послеобразы и иллюзии цветовых контрастов, и даже если бы внутри этого робота и происходили сходные процессы обработки послеобразов и тому подобных вещей, такой робот не мог бы обладать *квалиа* красного и синего, поскольку он был бы “всего лишь роботом”. Функциональные состояния, сигнализирующие о цветах, попадающих в поле зрения механических глаз робота, были бы лишены той дополнительной характеристики, которой они наделены у нас. Знаменитое эссе Томаса Нагеля “Что значит быть летучей мышью?” (1974) дает нам стандартный способ ссылаться на сознательные состояния организма, если эти состояния вообще существуют. Быть роботом, видящим послеобразы, *ничего бы не значило*. Почему так много людей считает это очевидным? Возможно, дело в том, что они представляют себе относительно простого робота и не учитывают, что нельзя делать выводы обо *всех* роботах, опираясь лишь на факты о *простых* роботах. Само собой, если *определить* квалиа как *внутренне присущие свойства* опытов, анализируемые в изоляции от всех их причин и следствий и логически независимые от любых диспозиционных характеристик, то квалиа, согласно логике, вообще не поддаются функциональному анализу. Сколько бы ни бились инженеры, им не удастся создать робота, наделенного квалиа, но в этом и нет особенного смысла, ведь нет никаких оснований полагать, что внутренне присущие свойства вообще существуют.

Чтобы увидеть это, сравним *квалиа* опыта с *ценностью* денег. Кажется, некоторые наивные американцы полагают, что доллары, в отличие от евро и йен, обладают *внутренне присущей ценностью*. Карикатурный турист спрашивает: “Сколько это в *настоящих деньгах*?” – пытаясь выяснить цену в долларах. Давайте пойдем немного дальше. Эти наивные американцы готовы обменивать доллары на евро и “сводить” ценность других валют к обменному курсу с долларом (или к стоимости товаров и услуг), но при этом им кажется, что доллары принципиально отличаются. Каждый доллар в их представлении имеет нечто логически независимое от своей функции единицы обмена, которую он делит со всеми остальными находящимися в обращении валютами. Доллар обладает неким *трудноуловимым* свойством. Задумавшись об этом, вы поймете, что ему свойственна аура ценности – может, и не такая яркая, как в старые добрые времена, но все равно различимая. Назовем ее *жизненной силой* доллара. В таком случае официально жизненная сила – это не связанная ни с какими отношениями и диспозициями *внутренне присущая* экономическая ценность доллара. Фунты стерлингов, евро и другие валюты не обладают внутренне присущей ценностью – они лишь символически заменяют доллары и оплачиваются долларами, а следовательно, обладают *производной* экономической ценностью, но не жизненной силой! Бедняги европейцы! Их валюты не имеют внутренне присущей экономической ценности! Как они с этим справляются? Как они мотивируют себя зарабатывать эти жалкие *безжизненные* гроши? Те из нас, кому повезло получать зарплату в долларах, зарабатывают деньги с хорошей жизненной силой. Неудивительно, что многие нации выбирают американский доллар в качестве своей резервной валюты! Даже иностранцы чувствуют жизненную силу доллара.

Так говорят наши воображаемые американские туристы. Пользуясь таким определением жизненной силы, экономисты никогда не поймут, в чем состоит жизненная сила каждого доллара, поскольку ни одна экономическая теория не в состоянии объяснить *внутренне присущую экономическую ценность*. Тем хуже для экономики? Наличие жизненной силы сделало бы экономическую теорию несостоятельной, однако у нас, к счастью, нет оснований полагать, что внутренне присущая экономическая ценность вообще существует. Совершенно очевидно, что жизненная сила – это плод воображения, артефакт интуитивного восприятия этих наивных американцев, а объяснить артефакт можно и не признавая его.

Некоторые участники дебатов о сознании похожи на этих воображаемых туристов. Они безапелляционно заявляют, что их представления о *внутренне присущих свойствах феноменов*, без сомнения, должны ложиться в основу любой науки о сознании. Такую убежденность стоит считать любопытным *симптомом*, заслуживающим диагноза, – исходным фактом, который должна учитывать любая наука о сознании, – подобно тому как экономистам и психологам, вероятно, не помешало бы объяснить, почему так много людей питает иллюзию, что деньги обладают внутренне присущей ценностью. (Когда Европа перешла на евро, людям, которые привыкли воспринимать цены во франках, марках, лирах и других валютах, пришлось пройти через трудный период, когда они не могли полагаться

на “перевод” цен в *свои* “настоящие деньги”. Первое исследование этого феномена см. в работе Dehaene, Marques 2002.)

Многие свойства сознательных состояний могут и должны быть прямо сейчас подвергнуты дальнейшему научному изучению, а поняв, что они собой представляют, мы с большой долей вероятности увидим, что они вполне подходят нам для объяснения природы сознания. В конце концов, именно так случилось с давней “загадкой” о природе *жизни*. Как выяснилось, витализм – представление о существовании во всех живых организмах некоего важного секретного ингредиента, называемого *жизненным порывом*, – оказался ошибкой воображения. Сегодня витализм почти исчез, однако некоторые упрямцы не сдаются. Вдохновленные этой историей успеха, мы продолжим научное изучение сознания. Если настанет день, когда будут объяснены все наблюдаемые феномены сознания и оплачены все интеллектуальные счета, а мы увидим, что чего-то важного все равно *не хватает* (если это действительно важно, это станет настоящим бельмом на глазу), все те, кто не отказался от своих подозрений, скажут: “Мы же говорили!” Пока что пусть подумают, как отделаться от диагноза, что они, как и виталисты, идут на поводу у иллюзии. Вот вопрос для тех, кто считает квалиа внутренне присущими свойствами опыта: чем это представление отличается от ошибки наивных американцев? (Или американцы правы? Может, доллары *действительно* обладают жизненной силой, как любой вам скажет?)

## 58. Печальная история мистера Клапгра

Так что же такое квалиа, если не *внутренне присущие* свойства сознательного опыта? Однажды вечером за бутылкой шамбертена философ Уилфрид Селларс сказал мне: “Дэн, квалиа – это то, что делает жизнь стоящей!” Интересная идея. Давайте посмотрим, какими в таком случае окажутся квалиа. Чтобы проанализировать это, я сконструирую насос интуиции на основе недавних исследований когнитивной науки, в ходе которых изучалось нескольких причудливых и неочевидных патологий: *прозопагнозия* и *синдром Капгра*.

Страдающие от прозопагнозии обладают нормальным зрением, но испытывают трудности с распознаванием лиц. Они не могут отличить мужчину от женщины, старика от юноши, африканца от азиата, а в компании нескольких близких друзей одного пола и возраста они не в состоянии сказать, кто есть кто, пока не услышат голос или не заметят другую характерную особенность. Если попросить такого человека выбрать из ряда фотографий, среди которых будут снимки знаменитых политиков, кинозвезд, членов семьи и безвестных незнакомцев, портреты знакомых ему людей, как правило, он будет выбирать наугад. Не страдающим от прозопагнозии сложно представить, каково это – смотреть, к примеру, на собственную мать, не узнавая ее. Кому-то и вовсе сложно поверить, что прозопагнозия существует на самом деле. Рассказывая людям о подобных феноменах, я часто сталкиваюсь со скептиками, которые вполне уверены, что я все выдумываю на ходу. Но нам следует научиться рассматривать подобные трудности в качестве мерила хрупкости нашего воображения, а не в качестве примеров невозможного. Прозопагнозия (от греческого *prosopon*, что значит “лицо”, и *agnosia*, “неузнавание”) – хорошо изученная, бесспорная патология, от которой страдают тысячи людей.

Весьма любопытно, что (многие) страдающие от прозопагнозии не могут распознавать лица сознательно, однако по-разному реагируют на знакомые и незнакомые лица, а некоторые их реакции и вовсе доказывают, что *исподволь, сами того не сознавая*, они распознают те лица, которые не могут распознать по требованию. К примеру, такое неосознанное распознавание наблюдается, когда страдающим от прозопагнозии показывают фотографии и дают пять вариантов ответа для каждой. Они выбирают ответ наугад, но их кожно-гальваническая реакция – мера эмоционального возбуждения – становится гораздо сильнее, когда они слышат правильное имя, связанное с фотографией. Можно провести простую проверку: какие из следующих имен принадлежат политикам – Мэрилин Монро, Ал Гор, Маргарет Тэтчер, Майк Тайсон? Скорее всего, вы быстро справились с этой задачей, однако если бы каждое из имен сопровождалось неверной фотографией, вам понадобилось бы значительно больше времени. Объяснить это можно, только если *на каком-то уровне* вы все равно распознавали лица, пусть в этом и не было нужды. Таким образом, представляется, что в мозге есть (как минимум) две по большей части независимые друг от друга системы визуального распознавания лиц: поврежденная *сознательная* система, которая не может помочь испытуемым справиться с заданием, и неповрежденная *бессознательная* система, отвечающая возбуждением на несоответствие лиц и имен. Дальнейшие испытания показывают, что неповрежденная система находится “выше”, в зрительной коре, а поврежденная связана с “нижней”, лимбической системой. На самом деле типов прозопагнозии гораздо больше и теперь нам лучше известно, какие зоны мозга с ней связаны, однако нам вполне хватит и этого упрощенного описания, ведь впереди нас ждет еще более странная патология, называемая синдромом Капгра (по фамилии впервые описавшего ее в 1923 г. французского психиатра Жан-Мари Жозефа Капгра).

Люди, страдающие от синдрома Капгра, неожиданно начинают верить, что кого-то из их близких – как правило, супруга, любовника или родителя – тайком заменил его двойник. Страдающие от синдрома Капгра не считаются сумасшедшими: в остальном это совершенно нормальные люди, которые в результате травмы мозга неожиданно приобретают это специфическое убеждение, которое, несмотря на его причудливость и крайнюю маловероятность, поддерживают с такой уверенностью, что в некоторых случаях убивают или серьезно ранят “двойника” – то есть близкого им человека. На первый взгляд может показаться, что ни одна травма мозга не в состоянии оказывать именно *такой* странный эффект. (Вдруг в таком случае где-то есть и люди, которые получили по голове и стали полагать, что луна сделана из зеленого сыра?) Однако специалист по когнитивной нейробиологии Эндрю Янг обратил внимание на закономерность и предположил, что синдром Капгра представляет собой “противоположность” патологии, которая вызывает прозопагнозию. При синдроме Капгра сознательная, кортикальная система распознавания лиц не страдает – именно поэтому больной узнает, что стоящий перед ним человек как две капли воды похож на его близкого, – но бессознательная, лимбическая система отключается, лишая это



узнавание эмоционального резонанса, который оно должно вызывать. *Отсутствие* столь скромного фактора узнавания играет такую большую роль (“Чего-то не хватает!”), что фактически перечеркивает позитивное распознавание лица знакомого человека неповрежденной сознательной системой, и в итоге больной искренне верит, что смотрит на двойника. Вместо того чтобы винить в этом несоответствии собственную систему восприятия, страдающие от синдрома Клапгра винят весь мир таким метафизически экстравагантным, таким невероятным образом, что вряд ли можно сомневаться во влиянии (по сути, политическом), которым обычно пользуется поврежденная в данном случае бессознательная система распознавания лиц. Когда эпистемологический голод этой системы остается неутоленным, она реагирует таким образом, что обесценивает результаты работы остальных систем.

Хэдин Эллис и Янг впервые предложили эту гипотезу в 1990 г., и с тех пор сам Янг, а также нейробиолог Крис Фрит и другие ученые подтвердили и расширили ее. Само собой, есть и сложности, но я не буду углубляться в них, поскольку хочу использовать этот пример из когнитивной нейробиологии, чтобы стимулировать воображение и открыть наш разум для другой возможности, еще не обнаруженной, но вполне вообразимой. Вот *воображаемый* случай несчастного мистера *Клапгра*, которого я назвал так, чтобы не забывать, что мы черпаем вдохновение из реальной жизни, где существует синдром Клапгра. (Этот сценарий присоединяется к сонму насосов интуиции, при помощи которых философы изучают все мыслимые нарушения человеческого сознания, возможно, имеющие отношение к природе квалиа.)

Мистер Клапгра зарабатывает на жизнь, участвуя в психологических и психофизиологических экспериментах, а потому неплохо распознает собственные субъективные состояния. Однажды утром он просыпается и, едва открыв глаза, в отчаянии восклицает: “А-а-а! Что-то не так! Весь мир стал... странным... ужасным, каким-то *неправильным*! Я не знаю, хочу ли и дальше жить в *таком* мире!” Клапгра закрывает глаза, протирает их и с опаской открывает снова, но перед ним опять оказывается до странности отталкивающий мир – знакомый, но в то же время другой, хотя описать его отличия и не получается. Так говорит сам Клапгра. Его спрашивают: “Что вы видите, когда смотрите вверх?” Он отвечает: “Голубое небо, белые перистые облака, желтовато-зеленые почки на весенних деревьях, ярко-красного кардинала на ветке”. Судя по всему, он не испытывает проблем с цветовым зрением, но на всякий случай ему дают стандартный тест Исихары, который показывает, что он не страдает от дальтонизма. Кроме того, мистер Клапгра правильно определяет цвет нескольких десятков полей цветовой системы Манселла. Почти все приходят к выводу, что недуг несчастного мистера Клапгра не затрагивает его цветовое зрение, но один ученый, мистер Хромафил, решает провести еще несколько тестов.

Хромафил занимается исследованиями цветовых предпочтений, эмоциональных ответов на цвет и влияния различных цветов на уровень внимания и концентрации, кровяное давление, частоту сердечных сокращений, скорость обмена веществ и множество других висцеральных процессов. За последние шесть месяцев в ходе проведения тестов он собрал огромную базу данных специфических и типовых реакций мистера Клапгра и хочет проверить, изменилось ли что-нибудь. Он заново проводит все тесты и замечает поразительную закономерность: все эмоциональные и висцеральные реакции, которые Клапгра раньше выдавал на синий цвет, теперь выдаются на желтый – и наоборот. Если раньше он предпочитал красный зеленому, то теперь – зеленый красному. То же самое случилось и с другими его цветовыми предпочтениями. Пища вызывает у него отвращение, если только он не ест в темноте. Сочетания цветов, которые он раньше находил приятными, теперь кажутся ему отталкивающими, а сочетания “противоположных” цветов приятными – и так далее. Оттенок шокирующе розового, который раньше ускорял его сердечный ритм, он по-прежнему распознает как шокирующе розовый (хотя и сомневается, что хоть кого-то может шокировать *этот* оттенок розового), но теперь он его успокаивает, в то время как успокаивавший его ранее оттенок лаймового зеленого теперь его возбуждает. При взгляде на картины траектория его саккад – быстрых скачкообразных движений глаз при сканировании изображения – теперь существенно отличается от записанных ранее траекторий, явно определявшихся тем порядком, в каком его внимание привлекали цвета на холсте. Способность решать в уме арифметические задачи, которая ранее сильно подавлялась при помещении мистера Клапгра в ярко-синюю комнату, теперь подавляется при помещении его в ярко-желтую комнату.

Иными словами, хотя Клапгра и не жалуется на проблемы с цветовым зрением и прекрасно справляется со всеми стандартными тестами на распознавание и различение цветов, его эмоциональные реакции на цвета существенным образом изменились, как изменилось и воздействие цветов на его внимание. Доктор Хромафил объясняет коллегам, что случай мистера Клапгра не так уж загадочен: дело в том, что он столкнулся с полной инверсией цветовых *квалиа*, которая не затронула его высокоуровневые

когнитивные цветовые таланты, к примеру способность различать и распознавать цвета, то есть те таланты, которыми мог бы быть наделен цветочувствительный робот.

Что же *нам* сказать? Подверглись ли инверсии квалиа Клапгра? Поскольку случай воображаемый, ответить на этот вопрос, казалось бы, можно как угодно, но философы годами всерьез рассматривают другие воображаемые случаи, полагая, что от результатов этих изысканий зависят ответы на важные теоретические вопросы, так что отмахиваться от этого примера не стоит. Прежде всего, *возможно* ли вообще такое? Это зависит от того, о какой возможности идет речь. О логической? О физиологической? Разница огромна. Философы склонны не обращать внимания на физиологические ограничения, поскольку в философских баталиях они значения не имеют, но в этом случае они могут пойти на попятный. Я не вижу оснований заявлять о логической невозможности описанного. Клапгра обладает странной комбинацией нетронутых способностей и шокирующих новых неспособностей; обычно тесно взаимосвязанные диспозиции беспрецедентным образом диссоциированы, но можно ли считать его недуг в этом отношении более радикальным, чем прозопагнозию или синдром Капгра? Я не уверен даже, что состояние Клапгра невозможно физиологически: описано немало случаев, в которых больные прекрасно различают цвета, но не могут их называть (цветовая аномия) или становятся дальтониками, но не замечают этого, беспечно выдумывая и называя цвета наугад, не понимая, что на самом деле играют в угадывку. Клапгра, как и больной с синдромом Капгра, не испытывает трудностей с распознаванием и называнием цветов – его проблема не столь очевидна и невыразима: дело в том, что сбились все частные диспозиции, благодаря которым на картины хочется смотреть, комнаты хочется красить, а сочетания цветов – подбирать. У Клапгра изменились те эффекты цветов, которые делают жизнь стоящей, – иначе говоря (если Селларс был прав), его цветовые *квалиа*.

Допустим, мы объясним все это Клапгра и спросим его, подверглись ли инверсии его цветовые квалиа. У него есть три возможных ответа: “Да”, “Нет” и “Я не знаю”. Что он ответит? Если сравнить мою историю о Клапгра со множеством историй об инверсии квалиа, предлагавшихся и во всех деталях обсуждавшихся философами, самым тревожным нововведением станет возможность того, что Клапгра действительно столкнулся с инверсией квалиа, но сам этого не понял. Не забывайте, доктору Хромафилу пришлось предлагать свою гипотезу скептически настроенным коллегам, а Клапгра вполне может разделять их скептицизм. В конце концов, он не только не жаловался на проблемы с цветовыми квалиа (как в стандартных историях), но и убедил себя, что его цветовое зрение в порядке, точно так же, как убедил в этом ученых: успешно справившись со стандартными тестами на цветовое зрение. Эта особенность моей истории должна вызывать некоторый дискомфорт, поскольку в философской литературе господствует мнение, что подобное поведенческое самотестирование не имеет значения: **безусловно** (динь!), *такие* тесты вообще не имеют отношения к квалиа. Как правило, считается, что такие тесты вообще не могут пролить свет на природу квалиа. Однако, как показывает моя вариация истории, философы не учитывают, что у кого-то может хотя бы возникнуть искушение *опереться* на эти тесты, чтобы убедиться, что его квалиа не изменились.

Могут ли ваши квалиа оставаться неизменными, пока меняются ваши эмоциональные реакции? Среди философов нет согласия о том, как отвечать на такие дефиниционные вопросы о квалиа. Рассмотрим действие плутамата натрия – усилителя вкуса. Несомненно, с ним вкус пищи кажется лучше и ярче, но *меняет ли он квалиа* пищи или лишь повышает чувствительность людей к тем квалиа, которые у них уже были? Здесь я призываю к прояснению концепции квалиа, а не пытаюсь найти ответ на эмпирический вопрос о принципе действия плутамата натрия или различиях реакций на плутамат натрия, продемонстрированных испытуемыми, поскольку, пока мы не разберемся с концепциями, любые выводы о лежащих в основе наблюдаемого нервных процессах или **гетерофеноменологии** предмета будут систематически неоднозначны. Я просто хочу узнать, как философы используют слово “квалиа”: считают ли они все изменения субъективных реакций изменениями квалиа – или же существует некоторое привилегированное подмножество реакций, фактически фиксирующих квалиа? Абсурдна ли идея об изменении эстетического представления человека о конкретном квалиа – или ответа на конкретный квалиа? Пока не будут даны ответы на эти дефиниционные вопросы, термин остается не просто расплывчатым или запутанным – он безнадежно неоднозначен и колеблется между двумя (или более) фундаментально различными идеями.

Подверглись ли инверсии цветовые квалиа Клапгра? Некоторые философы утверждают, что я недостаточно подробно описал его состояние. Я описал его поведенческие компетенции – он правильно узнает, различает и называет цвета, но демонстрирует “неверные” реакции во многих других отношениях, – но не описал его субъективное состояние. Я не сказал, что он испытывает, смотря на спелый лимон: *внутренне присущий субъективный желтый* или, скажем, *внутренне присущий*

*субъективный синий*. Но в этом и суть: я подвергаю сомнению предположение, что эти термины вообще обозначают реальные характеристики его опыта. Допустим, я добавлю, что в ответ на этот вопрос Клапгра *говорит*: “Так как я по-прежнему вижу спелые лимоны *желтыми*, само собой, мой опыт включает характеристику внутренне присущего субъективного желтого”. Устраивает ли нас такой ответ? Можем ли мы с уверенностью сказать, что он знает, о чем говорит? Стоит ли нам ему поверить – или же он пребывает в плену философской теории, которая не заслуживает его преданности?

Вот плавный недостаток философских методов, обычно используемых в подобных случаях: философы склонны предполагать, что все компетенции и диспозиции, демонстрируемые нормальными людьми, скажем, в отношении цветов, формируют монолитный блок, не поддающийся разложению или делению на независимые субкомпетенции и субдиспозиции. Таким образом они успешно избегают необходимости рассматривать вопрос, должны ли квалиа быть привязаны к некоторому подмножеству диспозиций или к конкретной диспозиции. К примеру, философы Джордж Грэм и Терри Хорган (2000, р. 73) говорят о “непосредственном знакомстве с самим характером феномена – знакомстве, которое закладывает эмпирический фундамент для распознавательных и дискриминационных способностей [человека]”. Откуда они знают, что это “непосредственное знакомство” закладывает “фундамент” для распознавательных и дискриминационных способностей? Предполагается, что страдающим от прозопагнозии непосредственно знакомы те лица, которые они видят, или хотя бы “зрительные квалиа” этих знакомых лиц, но при этом больные не могут распознать их как квалиа, проявляющиеся, когда они смотрят на лица друзей и близких. Если снова вернуться к определению Уилфрида Селларса, который сказал, что квалиа – это то, что делает жизнь стоящей, то квалиа, возможно, *нельзя* считать “эмпирическим фундаментом” нашей способности изо дня в день узнавать, различать и называть цвета.

## 59. Настроенная колода

В знаменитой статье философ Дэвид Чалмерс (1995) отличает “легкие” проблемы сознания от того, что сам называет (с большой буквы “Т”) Трудной проблемой сознания. “Легкие”, по мнению Чалмерса, проблемы все равно весьма трудны. Вот, например, несколько сложных вопросов о сознании:

1. Как сознание позволяет нам говорить об образах, которые мы видим, звуках, которые слышим, запахах, которые чувствуем, и так далее? (Или – в упрощенном варианте – как информация из тех зон мозга, которые отвечают за восприятие, используется в зонах, отвечающих за язык, для формирования предоставляемых нами отчетов и ответов?)

2. Когда мы выполняем рутинную задачу (с которой можем справиться “едва ли не во сне”), почему сознание включается всякий раз при возникновении проблемы и каким образом оно помогает нам ее решить?

3. Сколько независимо движущихся объектов мы можем одновременно непрерывно отслеживать и как мы это делаем? (Ответ: как минимум четыре. Убедиться в этом можно, посмотрев потрясающую демонстрацию этого феномена, называемого FINST-индексацией, на сайте <http://rucss.rutgers.edu/finstlab/MOT-movies/MOT-Occ-baseline.mov>.)

4. Что происходит, когда что-то “вертится у вас на языке” – когда вы знаете, что ответ вам известен и *почти* можете его озвучить?

5. Почему нужно понимать шутку, чтобы она казалась смешной? (Ответ длиною в целую книгу см. в работе Hurley, Dennett, Adams 2011.)

По словам Чалмерса, эти проблемы относительно легки, поскольку связаны с когнитивными функциями сознания – тем, что мы можем *делать*, используя непрерывно протекающие в мозге процессы отслеживания, направления внимания, обработки информации и вызывания ее из памяти. Как бы сложно ни было решить эти проблемы, их решения будут поддаваться проверке и корректировке в ходе экспериментов, и на самом деле мы делаем успехи в отношении “легких” проблем. К примеру, мы в состоянии сконструировать *относительно* простые компьютерные модели, которые весьма убедительно воспроизводят соответствующие функции, так что мы можем быть уверены, что мозг справляется с ними, не прибегая к магии или чему-либо, не имеющему аналогов в природе. Можно создать робота, демонстрирующего все эти феномены, – если не сегодня, то в обозримом будущем.

Трудной, в представлении Чалмерса, является проблема “опыта” – что значит быть сознательным, – невыразимая, не поддающаяся анализу *таковость* сознательности. Робот мог бы вести себя *так, словно бы* он сознателен: отвечать на все наши вопросы, отслеживать все движущиеся объекты, испытывать на себе действие феномена, когда что-то “вертится на языке”, смеяться исключительно впопад и (бессознательно) чувствовать себя растерянным или ошарашенным, – но на самом деле голова у него оставалась бы при этом пустой. Робот был бы зомби, не имеющим ни намека на внутренний мир, свойственный нам с вами – обычным сознательным людям.

Согласно Чалмерсу, бодрствуя, мы с вами, милостивый читатель, понимаем, что мы сознательны. Философский зомби этого не понимает – он никогда не бодрствует и не обладает внутренним миром – он *кажется сознательным лишь внешне*. Само собой, он убедительно настаивает, что столь же сознателен, как и мы с вами, а если подвергнуть его проверке на детекторе лжи, он прекрасно справится с тестом на искренность, но, будучи зомби, он ошибается! (Зомби также неотличимы от обычных сознательных людей, когда нейробиологи анализируют активность их мозга при помощи фМРТ и тому подобных тестов.) Очевидно, что отличить сознательного человека от зомби – действительно трудная задача (если считать ее задачей вообще). Если же *это* трудно, то объяснить существование этих различий еще труднее – в этом и заключается Трудная проблема сознания. Некоторые из нас, включая и меня самого, полагают, что Трудная проблема – плод воображения Чалмерса, но другие – и их на удивление много – убеждены, что существует или должна существовать реальная разница между сознательным человеком и совершенным зомби и что эта разница важна.

Позвольте мне рассмотреть эту любопытную ситуацию: некоторые из нас сомневаются в существовании Трудной проблемы, в то время как другие считают нас ненормальными, раз мы вообще позволяем себе в этом усомниться. По их мнению, любому сознательному существу совершенно очевидно и интуитивно понятно его собственное сознание, но это наше удивительное свойство (пока) не поддается научному пониманию, а следовательно, заслуживает называться Трудной проблемой. Сблизить эти альтернативные точки зрения невозможно. Одна из них точно категорически неверна. Я годами пытался продемонстрировать, что, каким бы привлекательной ни казалась эта идея, от нее следует отказаться. Я вполне уверен, что мысль о существовании Трудной проблемы ошибочна, но не могу этого доказать. Более того, даже если бы я мог представить доказательства, они не встретили бы понимания, ибо некоторые философы заверяют меня, что ничто не заставит их пересмотреть свою позицию по этому вопросу, поскольку все настолько очевидно и неоспоримо, что никакой аргумент не может подорвать или сокрушить их доверие к этой идее. В связи с этим я не стану совершать тактическую ошибку и попытаюсь продемонстрировать доводами опровергнуть убеждение, которое выходит за рамки разумного.

Такое отношение к идее напоминает мне искреннюю убежденность, с которой многие выходят с выступления блистательного фокусника. Любой фокусник знает, что люди склонны раздувать свои воспоминания об удачных фокусах. Испытываемые ими удивление и недоумение делают воспоминания ярче, поэтому они серьезно и искренне настаивают, что увидели нечто большее, чем простой трюк фокусника. Некоторые люди ужасно хотят верить в волшебство. Вспомните замечание Ли Сигела о “настоящей магии”, которое обсуждалось в главе 22, где речь шла о **чудо-ткани**: “*Настоящей магией...* называют магию, которая на самом деле не настоящая, а настоящую магию, которую действительно можно творить, *настоящей никто не считает*”.

Многие считают сознание “настоящей магией”. Если то, о чем вы говорите, не суперархизкстраультрамегаграндиозно, значит, речь идет не о сознании – не о Тайне За Семью Печатями. Научный журналист Роберт Райт (2000) кратко излагает эту точку зрения:

Само собой, проблема здесь заключается в утверждении, что сознание “идентично” физической активности мозга. Чем больше Деннет и другие пытаются объяснить мне, что они имеют в виду, тем больше я убеждаюсь, что на самом деле они имеют в виду, что сознания не существует. [р. 398]

Любой набор фокусов мозга просто *не может быть* сознанием – *настоящим* сознанием. Но даже те, кто не совершает этой опережающей ошибки, часто склонны преувеличивать феномены сознания. (Именно поэтому столько места в моей книге “Объясненное сознание” (1991a) пришлось отвести *сжатию*, сокращению сознания – настоящего сознания – до его истинных размеров, чтобы продемонстрировать, что феномены сознания на самом деле не столь удивительны, как о них думают. Из-за этого сжатия многие читатели не упустили случая пошутить, что мою книгу стоило назвать “Поверхностно объясненное сознание” или – как предлагает Райт – “Опровергнутое сознание”.) Тем из вас, кто сомневается, что может запутаться в раздутом представлении о сознании, я хочу нанести скользящий удар в надежде уничтожить их самодовольство, проведя замечательную и пугающую параллель с миром карточных фокусов – фокус “Настроенная колода”.

Много лет Ральф Халл – знаменитый повелитель карт из Круксвилла, что в штате Огайо, – поражал не только обывателей, но и начинающих фокусников, карточных мастеров и профессиональных иллюзионистов карточным фокусом, который сам называл “Настроенной колодой”...

Джон Нортен Хиллиард, “Карточная магия” (1938)

Фокус Ральфа Халла примерно таков:

Ребята, сейчас я покажу вам новый фокус. Он называется “Настроенная колода”. Эта колода карт волшебным образом настроена [Халл подносит колоду к уху, перебирает карты и внимательно прислушивается к их шелесту]. Благодаря тонко настроенным вибрациям, я могу *услышать* и *почувствовать* местонахождение каждой карты. Выберите любую карту... [Колода в этот момент раскрывается веером или иным образом предлагается зрителям, один из которых вытягивает карту, смотрит на нее, а затем тем или иным способом возвращает ее в колоду.] Теперь я прислушаюсь к настроенной колоде. Что она мне скажет? Я слышу характерные

вибрации... [бз-з-з, бз-з-з, Халл перебирает карты возле уха и производит всяческие манипуляции и ритуалы, после чего эффектным жестом достает карту зрителя].

Халл снова и снова показывал этот фокус коллегам-иллюзионистам, предлагая им разгадать секрет его трюка. Разгадать его никто не сумел. (Не забывайте, главное правило карточной магии гласит, что фокус нельзя повторять перед одной и той же аудиторией. Этот великий фокус дерзко обошел данное правило.) Фокусники предлагали ему продать секрет фокуса, но Халл отказался его продавать. На старости лет он раскрыл секрет своему другу Хиллиарду, который опубликовал его рассказ в напечатанной за свой счет книге. Вот что Халл сказал о своем фокусе:

Много лет я показывал этот фокус сотням иллюзионистов и обывателей, но, насколько мне известно, никто не смог разгадать его секрет... *ребята слишком сильно пытались его разгадать* [курсив мой].

Как и большинство великих фокусов, этот фокус заканчивается раньше, чем вы осознаете, что он начался. Фокус состоит в его названии – “Настроенная колода”. Как только Халл объявлял о новом фокусе и сообщал заинтересованной аудитории его название, фокус заканчивался. Подготовив аудиторию таким простым способом и потянув время с помощью очевидно надуманной и сбивающей с толка болтовни о вибрациях и жужжании, Халл исполнял относительно простой и знакомый карточный фокус типа А (здесь я не стану дальше приоткрывать традиционную завесу тайны; как вы увидите, механические детали этого трюка не имеют значения). Его зрители, умелые фокусники, видели, что он, возможно, показывает фокус типа А, и могли проверить свою гипотезу, проявив упрямство и несговорчивость и тем самым не позволив успешно показать фокус типа А. Когда они начинали упрямиться, чтобы проверить свою гипотезу, Халл “повторял” фокус, но на этот раз показывал фокус типа Б. Зрители начинали перешептываться: может, он показывает фокус типа Б? Чтобы проверить новую гипотезу, они начинали упрямиться иным образом, не давая Халлу успешно исполнить фокус типа Б. Но он все равно показывал “свой” фокус – теперь используя метод В. Когда зрители проверяли гипотезу, что он показывает фокус типа В, он переключался на метод Г, а возможно, возвращался к методу А или Б, поскольку аудитория уже “опровергла” гипотезу о том, что он использует метод А или Б. Так продолжалось десятки раз, и Халл неизменно оказывался на шаг впереди своей аудитории: он понимал, что всегда сможет показать *тот или иной фокус* из множества известных каждому из зрителей фокусов, и скрывал, что фокусы будут меняться, присваивая “новому” фокусу название – “Настроенная колода”. Халл объяснил Хиллиарду:

При каждом исполнении этого фокуса опровергается хотя бы одно предположение, которое рождается у зрителя в голове, поэтому рано или поздно он неизбежно прекращает свои попытки разгадать загадку.

Секрет Халла заключался в том, что он озвучивал название своего фокуса, тем самым убеждая зрителей, что показывает один-единственный фокус, а не целую серию! В результате сознание зрителей отключалось, не позволяя им прибегать к **выпрыгиванию**. Они оказывались в ловушке системы, уверенные, что должны разгадать секрет большого, нового фокуса, и не замечали, что на самом деле искать нужно не один ответ, а несколько. У них не получалось выпрыгнуть из системы.

В связи с этим я предполагаю, что Дэвид Чалмерс – непреднамеренно – исполнил подобный фокус, объявив всему миру, что открыл “Трудную проблему сознания”. Существует ли эта Трудная проблема *в действительности*? Или же то, что кажется Трудной проблемой, на самом деле представляет собой совокупность феноменов, которые Чалмерс называет Легкими проблемами сознания? Все они имеют будничное объяснение, не требующее революции в физике и новых открытий. Они поддаются – хоть и без труда – анализу при помощи стандартных методов когнитивной науки.

Я не могу доказать, что Трудной проблемы не существует, а Чалмерс не может доказать, что она существует на самом деле. К нему пришло одно мощное интуитивное озарение, которое можно было бы положить в основу новой теории сознания, если бы оно привело к появлению новых сенсационных прогнозов или помогло в перспективе объяснить нечто непонятное в ином случае, однако это озарение по-прежнему стоит особняком, не порождая никаких теоретических прорывов, хотя опровергнуть его и сложно.

Список известных проявлений сознания весьма велик и растет день ото дня, причем в нем есть как обыденные, так и экзотические примеры. Сложно учесть их все, поэтому не стоит забывать о том, что мы,

по сути, можем пасть жертвой арифметической ошибки, когда пересчитаем все Легкие проблемы и обнаружим неучтенный остаток. Возможно, мы не заметили, что этот остаток уже был учтен в списке будничных проблем, объяснение которым мы уже нашли – или хотя бы знаем, в каком направлении мыслить. Как мы можем совершить эту “арифметическую ошибку” и просмотреть ее? Есть опасность посчитать какие-то феномены дважды или забыть, что мы уже объяснили один из феноменов, в связи с чем его нужно исключить из списка феноменов, требующих объяснения. Вероятно ли, что мы совершаем такую ошибку? Вспомните, рассматривая случай несчастного мистера Клапгра, мы видели, что с ним *что-то* не так, но его проблему можно было описать двумя существенно разными способами:

А. Его эстетические и эмоциональные *реакции* на цветовые квалиа подверглись инверсии (но сами квалиа остались неизменными).

Б. Его цветовые квалиа подверглись инверсии, хотя он сохранил способность различать, распознавать и называть цвета.

Можно возразить (подобно озадаченным иллюзионистам, пытающимся разгадать фокус Халла): “А не может быть верно, поскольку единственным основанием утверждать, что его цветовые квалиа остались неизменными, служит то, что неизменным осталось его *поведение при различении и назывании цветов*, но это ничего не говорит о его квалиа, поскольку такие паттерны поведения представляют собой (лишь) когнитивные, функциональные факты, а квалиа, само собой, от них независимы. Б также не может быть верно, поскольку изменились *только* его реакции: Клапгра не жалуется, что цвета стали *выглядеть иначе*, а лишь говорит, что те же самые субъективно воспринимаемые цвета больше не производят на него того же впечатления, которое производили раньше. Таким образом, его цветовые квалиа могли измениться, а могли остаться неизменными, и – как вы увидите – нет эмпирического способа установить, какая из гипотез верна! Вот поистине Трудная проблема!”

Такой аргумент не учитывает возможность того, что упоминаемые в вариантах А и Б квалиа не выполняют никакой работы. Как в варианте А, так и в варианте Б мы видим, что аппарат различения цветов функционирует так же, как и раньше, но реакции Клапгра на результаты его работы инвертированы. Квалиа вклиниваются в качестве своеобразного, с трудом поддающегося описанию посредника: предполагается, что квалиа обеспечивают фундамент, или сырье, или основу эмоциональных реакций. В таком случае инверсия может произойти на двух этапах: перед тем как квалиа “представляются” распознающему их аппарату или после “представления”, при ответе распознающего аппарата на представленные квалиа. Но на самом деле никакого представления не происходит. Мы знаем, например, что негативные (тревожные, пугающие) реакции могут запускаться на ранних этапах процесса восприятия и впоследствии “окрашивать” всю последующую обработку воспринятой информации, а потому в таком случае мы могли бы сказать, что эмоциональные реакции становятся *причиной*, по которой квалиа приобретают субъективный характер, знакомый Клапгра, а не (наоборот) что “внутренне присущая” природа квалиа становится причиной или основанием эмоциональных реакций. Но если мы уже пришли к эмоциональным (или эстетическим, или аффективным) реакциям на воспринятую информацию, то “работы” для квалиа больше не осталось, а зимбо, само собой, был бы столь же ошарашен инверсией реакций на воспринимаемое, как и сознательный человек.

Что эта история о настроенной колоде добавляет к остальным насосам интуиции о квалиа? Она просто служит реальным примером того, как умных, компетентных специалистов можно склонить к созданию фантомной проблемы, просто поставив вопрос определенным образом. Такое случалось. Такое может случиться снова. И это позволяет по-новому взглянуть на тупиковую ситуацию, создавая новое бремя доказывания: как вам понять, что вы *не попались* в ловушку, подобную фокусу с настроенной колодой? Я не говорю, что эти выкладки неопровержимы, но они должны подтолкнуть всех тех, кто доверяет **чутью на зомби**, усомниться в его “очевидности”.

## 60. Китайская комната

В конце 1970-х гг. вокруг ИИ – искусственного интеллекта – поднялась такая шумиха, что его перспективы оказались значительно преувеличены. Не за горами было появление думающих машин! Философ из Беркли Джон Сёрл нисколько не сомневался в этом и предложил мысленный эксперимент, чтобы доказать свою правоту. В 1980 г. он опубликовал статью “Сознание, мозг и программы”, в которой описал знаменитую “Китайскую комнату” – мысленный эксперимент, призванный показать, что создать “сильный” ИИ невозможно. Он отрицал “сильный” ИИ как утверждение, что “должным образом запрограммированный компьютер в буквальном смысле обладает когнитивными состояниями, а следовательно, программы объясняют человеческое сознание” (Searle 1980, p. 417). Позже он уточнил свое определение: “Должным образом запрограммированный цифровой компьютер, получающий соответствующие входные данные и дающий верные выходные данные, наделен сознанием в той же степени, что и человек” (Searle 1988, p. 136). Статья 1980 г. была опубликована во флагманском журнале когнитивной науки *Behavioral and Brain Sciences*, а BBS, как его называют, славится своим особым форматом: в каждом номере журнала печатается несколько длинных “целевых статей”, сопровождаемых десятками комментариев экспертов отрасли и ответами авторов. Поскольку BBS – решительно междисциплинарный журнал, как правило, эксперты представляют разные дисциплины, что дает читателю возможность увидеть любопытный срез мнений. В зависимости от того, насколько целевая статья понравилась экспертам – и понравилась ли она им вообще, – можно сделать вывод, стоит ли учитывать ее в своей работе. Кроме того, можно многое узнать о трудностях междисциплинарной коммуникации, в ходе которой уверенные в своей правоте люди упорно не слышат друг друга или участвуют в серьезнейшей академической борьбе. Целевая статья Сёрла спровоцировала волну опровержений, включая и мое (Dennett 1980), в котором я ввел термин “насос интуиции”, чтобы показать, в чем ошибка его мысленного эксперимента.

Термин предложил я, но благодарить за его появление нужно Дага Хофштадтера, поскольку термин родился в ходе наших дискуссий об эссе Сёрла, которые Даг с комментариями перепечатал в нашей антологии “Глаз разума” (Hofstadter, Dennett 1981). Мы сочли предложенный Сёрлом мысленный эксперимент весьма любопытным, поскольку он был, с одной стороны, очевидно ошибочным, а с другой стороны, столь же очевидно убедительным и угодным толпе. Как – и почему – он сработал? Мы проанализировали его, прибегнув к *обратному проектированию*, и Даг предложил “покрутить все регуляторы”, чтобы объяснить его механику. Выдержит ли эта история изменения или же она всецело зависит от деталей, которые *должны бы* быть факультативными<sup>[69]</sup>?

С тех пор прошло более тридцати лет, и я вынужден признать, что “Китайская комната” Сёрла не растеряла своей привлекательности и приобрела огромное влияние. Она стала классикой, изучаемой в тысячах университетов и снова и снова обсуждаемой по сей день. Я много лет использовал ее на собственных занятиях и хорошо разобрался в том, как она работает. Теперь я знаю, как показать, что именно с ней не так<sup>[70]</sup>.

Дело в том, что “Китайская комната”, как я сейчас продемонстрирую, представляет собой дефектный насос интуиции, **упорный башмак**, который может отключить ваше воображение, если вы не проявите достаточной осторожности. Но прежде всего я хочу сказать, что догадываюсь, что многие из вас сейчас закатывают глаза или даже стонут. Вам не хочется, чтобы я развенчивал этот миф, ведь вывод – создать “сильный” ИИ невозможно, *фух!* – нравится вам так сильно, что вас вовсе не прельщает перспектива знакомства с педантичной критикой живого, интересного аргумента, который поддерживает вашу отчаянную надежду. Раньше я с презрением относился к такой реакции. Тем из вас, кто реагирует таким образом, нравится, что именитый профессор Беркли приводит знаменитый довод, доказывающий вашу правоту, и вы счастливы поверить ему на слово. Подробности вас не интересуют – важен только вывод. Какое глупое оправдание!

Но затем я поймал себя на этом же и пересмотрел свою категоричную позицию. Признаюсь, я всегда находил квантовую механику несколько отталкивающей и вводящей в заблуждение. Я бы даже предпочел, чтобы она оказалась неправдой! При этом я знаю, что она великолепно справляется с предсказанием и объяснением многих феноменов, включая обыденные явления, например отражение и преломление света и активность белков в сетчатке глаза, благодаря которой мы видим. Квантовая механика – это сердце науки, но понять ее нелегко даже специалистам. Я несколько раз пытался справиться с математикой квантовой механики, но у меня ничего не вышло, поэтому я с интересом, но без какого-либо понимания



слежу за научными спорами о ее интерпретации, что, однако, не мешает мне скептически относиться ко многому из того, что говорят по теме предположительно компетентные эксперты. В один прекрасный день я прочитал научно-популярную книгу Марри Гелл-Мана “Кварк и ягуар. Приключения в простом и сложном” (1995), рассчитанную на неспециалистов вроде меня. Мне понравилось, как сурово Гелл-Ман расправляется со всеми мифами и камня на камне не оставляет от наиболее сомнительных представлений, которым каким-то образом удалось завоевать популярность. (Прочитайте главу “Квантовая механика и бред собачий” – и вы поймете, о чем я.) В какой-то момент я поймал себя на мысли “Так их, Мари! Не давай им спуску!” Всемирно известный нобелевский лауреат поддерживал мои предрассудки, используя понятные мне аргументы. Такая квантовая физика была мне по душе! Но потом я подумал: понимаю я в действительности его аргументы – или только **вроде как** понимаю их? Могу ли с уверенностью сказать, что не иду на поводу у Гелл-Мана? Я надеялся, что не найдется физика, который решит снова растолковать мне все тонкости и продемонстрировать, что я попался в ловушку. Мне так нравились выводы Гелл-Мана, что мне не хотелось копаться в деталях. Я использовал то же самое оправдание.

И все же не совсем. Я добросовестно оцениваю вердикты Гелл-Мана в свете того, что было написано позже (и пока полет нормальный). И я готов признать, что рано или поздно может случиться так, что “здравый смысл” Гелл-Мана, который мне так нравится, окажется очередным обманом воображения. Умоляю вас, попробуйте столь же непредвзято относиться к моей попытке разоружить и нейтрализовать “Китайскую комнату” Сёрла. Я постараюсь подслестить вам горькую пилюлю.

В 1950 г. Алан Тьюринг предложил тест, который должен был, по его мнению, стать лакмусовой бумажкой для измерения интеллекта машины. Теперь он называется тестом Тьюринга. В ходе него судья беседует с двумя сущностями – компьютером и человеком, – скрытыми из виду и коммуницирующими посредством “телетайпа” (скажем, монитора и клавиатуры). Человек пытается продемонстрировать судье свою истинную человечность, а компьютер пытается заставить судью признать его человеком. Если судья не в состоянии сказать наверняка, кто из них человек, то компьютер (программа) проходит тест Тьюринга и признается не только разумным, но и “сознательным в том же смысле, что и человек”, как выразился в 1988 г. Сёрл. Прохождение теста Тьюринга, по мнению многих специалистов, стало бы доказательством возможности создания “сильного” ИИ. Почему? Они (как и Тьюринг) полагали, что подобную беседу невозможно вести, не понимая ее смысла, а следовательно, успех тестируемого компьютера свидетельствовал бы о том, что машина понимает значения своих слов<sup>[21]</sup>. Обратите внимание, что тест Тьюринга не отличает – и не может отличить – зимбо от “поистине сознательного” человека, поскольку зимбо с тем же успехом *делают* все, на что способны сознательные люди. Как замечают многие, утверждение, что проходящая тест Тьюринга сущность не только разумна, но и *сознательна*, бросает вызов **чутью на зомби**. Казалось бы, из-за этого тест Тьюринга не может использоваться для проверки сознательности, однако нам не стоит делать выводов, пока мы не *разберемся в деталях*. Программа, проходящая тест Тьюринга, может обладать лишь характерным для зимбо потоком *бессознательного*, не имея нашего потока сознания, однако мы только что сказали, что это идет вразрез с чутьем на зомби, которое гласит, что разница между ними принципиальна. Может, программа-зимбо лишь **вроде как** сознательна? В каком именно отношении она не оправдывает ожиданий? Возможно, свет на это прольет насос интуиции Сёрла: судя по всему, в нем используется **доведение до абсурда**, дискредитирующее идею о “сильном” ИИ.

Сёрл предлагает нам представить, что он заперт в комнате и вручную моделирует работу огромной программы ИИ, якобы понимающей китайский язык. Он оговаривается, что программа проходит тест Тьюринга, пресекая все попытки человека отличить ее от сущности, понимающей китайский на самом деле. Не знающий китайского Сёрл заперт в комнате и манипулирует цепочками символов в соответствии с инструкцией, (очевидно) не приобретая в процессе способности понимать китайский. Больше в комнате нет ничего, что понимало бы китайский язык. (Комната пуста – в ней находятся только Сёрл и “клички бумаги”, на которых написаны его инструкции. В точности следуя им, он “вручную моделирует” работу гигантской программы.) Сёрл не понимает китайский, но и (динь!) Сёрл с клочками бумаги тоже не понимает китайский. Следовательно, ничто в китайской комнате не понимает китайский, хотя программа и может вести беседу на уровне, достаточном, чтобы обмануть носителей китайского языка. Подобно компьютеру, Сёрл различает китайские иероглифы исключительно по их форме: и для него, и для компьютера они представляют собой лишь различные бессмысленные “загогулины”, поэтому Сёрл в китайской комнате – это лишь *реализация* соответствующей компьютерной программы. Эта программа выполняет свою задачу, не понимая китайский, работая хоть на кремниевых чипах, хоть на Сёрле.

Так просто и так убедительно! Разве с этим мысленным экспериментом может быть хоть что-то не так? Увы, да. Когда Сёрл представил его в Беркли, специалисты по компьютерным наукам в ответ привели довод, который Сёрл назвал “аргументом о системе” (Беркли):

Хотя запертый в комнате индивид действительно не понимает происходящего, он представляет собой лишь часть целой системы, а система происходящее понимает. Индивид пользуется толстой книгой с правилами, имеет большое количество бумаги и карандашей для осуществления расчетов, а также располагает “базами данных” с наборами китайских иероглифов. Понимание приписывается не одному индивиду, а всей системе, частью которой он является. [Hofstadter, Dennett 1981, p. 358]

Ответ Сёрла на этот аргумент весьма показателен:

Мне даже немного неловко... потому что эта теория кажется мне невероятной. Мысль заключается в том, что человек может не понимать китайский, но *совокупность* этого человека и клочков бумаги каким-то образом может его понимать. Мне сложно представить, чтобы хоть кто-то, не пребывающий в плену идеологии, признал эту идею правдоподобной. [Searle 1980, p. 419]

Невероятным Сёрлу показалось не что иное, как фундаментальный прорыв Тьюринга, приведший к появлению идеи компьютера с хранимой в памяти программой! Вся компетентность – в программном обеспечении. Не забывайте, что регистровая машина из главы 24 совсем не понимает арифметику, но регистровая машина в совокупности с программным обеспечением прекрасно справляется с арифметическими действиями. Центральный блок обработки данных вашего ноутбука ничего не знает о шахматах, но, запустив шахматную программу, сможет обыграть вас – и то же самое верно в отношении остальных его великолепных способностей. Описываемая Сёрлом идеология служит стержнем компьютерной науки. Ее основательность демонстрируется повсеместно. Чтобы воспроизвести человеческую компетентность, а следовательно (в итоге), и понимание, необходимо поставить виртуальные машины на виртуальные машины на виртуальных машинах, поскольку сила в системе, а не в аппаратном обеспечении. “Странной логике” Дарвина вторит странная логика Тьюринга (Dennett 2013e): если раньше (до Тьюринга) мы полагали, что человеческая компетентность должна проистекать из понимания (таинственного источника разума), то теперь считаем само понимание эффектом, который достигается (снизу вверх) посредством наложения компетенций друг на друга.

Вся суть в деталях. Сёрл не говорит читателю, на каком уровне он осуществляет ручное моделирование гигантской программы ИИ. Вот как он сам описывает, что происходит, когда он исполняет программу:

Теперь допустим, что после получения первой партии китайских иероглифов [входных данных] я получаю вторую партию иероглифов вместе с набором правил о сопоставлении второй партии с первой. Правила написаны по-английски, и я понимаю их не хуже любого другого носителя английского языка. Они позволяют мне сопоставить один набор формальных символов с другим набором формальных символов... Теперь допустим также, что я получаю третью партию китайских иероглифов вместе с написанными по-английски инструкциями, позволяющими мне сопоставить элементы из этой партии с символами из первых двух партий, и эти правила объясняют мне, как выбирать определенные китайские иероглифы определенной формы в ответ на определенные формы, предоставленные мне в третьей партии. [Searle 1980, p. 418]

Сёрл противопоставляет это “сопоставление” разных “партий символов” происходящему в тот момент, когда на входе он получает английское предложение или английскую историю и отвечает на родном английском.

Со стороны – с позиции наблюдателя, который читает мои “ответы”, – кажется, что ответы на китайские и на английские вопросы одинаково хороши. Однако в случае с китайским, в отличие от случая с английским, я даю ответы, манипулируя неинтерпретируемыми формальными символами. [Searle 1980, p. 418]

Какой контраст! Но давайте разберемся, о чем Сёрл не говорит. Мы знаем, что он получает свои “наборы правил” (инструкции) на английском, но каковы они? Что-то вроде “прибавьте содержимое регистра 39021192 к содержимому регистра 215845085” (машинного кода) или вроде “определите постоянную длину очереди и задайте ей значение 100” (**исходного кода**)? Сидит ли Сёрл на низшем уровне, в бешеном темпе осуществляя арифметические действия (триллионы операций в секунду), или же он следует исходному коду программы, реализуемому на много уровней выше? А если он следует исходному коду, может ли он читать комментарии? Лучше бы нет, поскольку официально они не считаются частью программы, но могут предоставить ему множество намеков о том, что происходит (“Здесь осуществляется грамматический разбор предложения, выделяются подлежащие, сказуемые, обстоятельства и определения, в результате чего предложение признается вопросительным, повествовательным, побудительным или восклицательным”, а миллиард операций спустя “обнаруживается игра слов, осуществляется переключение на остроумие...”, после чего подробно описываются миллиарды операций, в ходе которых осознаются отсылки, оцениваются разные варианты ответов [прямо как разные варианты шахматных ходов] и наконец генерируется ответ). Если программа, работу которой вручную моделирует Сёрл, способна вести впечатляющую беседу на китайском, ей нужно сверяться с огромными базами данных, содержащими не только “наборы китайских иероглифов”, как выражается сам Сёрл, но и обиходные знания, доступные носителям китайского, и это лишь минимум необходимой ей информации. Получает ли Сёрл, моделирующий работу программы, намеки на всю эту многослойную когнитивную деятельность, или же она представляется ему сплошной арифметикой?

Подумайте, как Сёрл обработал бы следующий вопрос на английском:

Представьте заглавную букву *D* и положите ее набок, повернув на 90 градусов против часовой стрелки. Теперь поместите ее поверх заглавной буквы *J*. О какой погоде вам напоминает получившийся символ?

Теперь представьте, что в разгар работы в китайской комнате Сёрлу дают аналогичную задачу на китайском.

Четвертого июня 2012 г. *Sohu Weibo* [китайская блог-платформа] заблокировала пост со следующим набором символов. Можете ли вы понять почему?  
占占占占人 占占占点 占占点占 占点占占 点占占占 灬占占占

Это настоящие китайские иероглифы (“загогулины” Сёрла), но их последовательность не имеет смысла. Почему власти ее заблокировали? Дело в том, что 4 июня – годовщина бойни на площади Тяньаньмэнь, в ходе которой армия убила сотни протестующих. (“4 июня” значит для китайцев примерно то же, что “11 сентября” для американцев.) Самыми знаменитыми снимками того дня стали фотографии одинокого отважного человека, стоящего перед танками. В цепочке символов видно, как он встает перед четырьмя танками слева, после чего они снова и снова переезжают его и исчезают в правой стороне<sup>[72]</sup>.

Китайская комната должна “понимать это”, но сам Сёрл ничего не узнает, если только не получит доступ к комментариям к исходному коду, поскольку он не сможет понять, что в данном случае он следовал правилам, чтобы создать “мысленный образ”, обработать его и использовать результат для проверки памяти. Иными словами, прежде чем ответить на вопрос на китайском, *система* произвела бы набор действий, который поразительно напоминает сознательные действия Сёрла при ответе на вопрос на английском. Можно сказать, что система обладает собственным сознанием, невообразимым для Сёрла, прозябающего в ее моторном отсеке.

Любая программа, способная справиться с тестом Тьюринга, *должна* производить “мысленные” операции, очень точно имитирующие мысленные операции, осуществляемые нами в ходе разговора. Допустим, например, что интервьюер в тесте Тьюринга решил учить кандидата квантовой физике, используя сократовский метод вопросов и ответов и давая ученику простые задачи. Чтобы поддерживать этот разговор, сидящий в моторном отсеке Сёрл вынужден будет помогать системе справляться со сложными интеллектуальными упражнениями, но сам в итоге останется таким же несведущим в квантовой физике, каким был в начале эксперимента. Система, напротив, будет разбираться в этой теме гораздо лучше, чем до начала теста Тьюринга, поскольку она *выполняла упражнения*. В этом конкретном случае тест Тьюринга установит в программу новую виртуальную машину – машину простой квантовой физики.

Подобные факты полностью скрываются рисуемым Сёрлом образом “клочков бумаги” и “правил”, которые позволяют ему “устанавливать соответствие” между китайскими иероглифами. Я не утверждаю, что Сёрл намеренно скрыл сложность программы, моделирование работы которой представлял в ходе своего эксперимента, но он явно проигнорировал следствия ее сложности. Если вы представляете эту программу в качестве относительно простого набора правил, то, подобно Бэтсону, представлявшему “частицы хроматина, неотличимые друг от друга и практически гомогенные во всех известных опытах”, вы, вероятно, сочтете заявленные способности программ к пониманию “непостижимыми”, как и способности ДНК.

Смотрите, что мы сделали. Мы повернули регулятор предложенного Сёрлом насоса интуиции и изменили уровень детализации описания исполняемой программы. Уровней всегда много. На высшем уровне способности системы к пониманию нельзя назвать невообразимыми: мы даже можем увидеть, как именно система понимает, что делает. Ответ системы больше не кажется нелепым – он кажется очевидно верным. Это не значит, что ИИ такого типа, который критиковал Сёрл, на самом деле достигает того уровня компетентности, где можно говорить о понимании, или что такие методы, расширенные в соответствии с представлениями исследователей ИИ того времени, могли бы с большой долей вероятности привести к появлению компетенций столь высокого уровня. Это значит лишь то, что мысленный эксперимент Сёрла на самом деле не справляется с задачей, которую, как утверждается, он решает: он не демонстрирует абсолютную невозможность создания “сильного” ИИ.

Можно покрутить и другие регуляторы, но этим не раз занимались в огромном количестве литературы, написанной под влиянием “Китайской комнаты”. Здесь я уделяю основное внимание самому инструменту мышления, а не теориям и предположениям, на которые он был нацелен, и показываю, что этот инструмент дефектен: он убеждает, затуманивая наше воображение, а не используя его должным образом.

## 61. Падение телеклона с Марса на Землю

Одна луна восходит на востоке. Другая луна восходит на западе. Вы видите, как две луны движутся навстречу друг другу по холодному черному небу, после чего одна проходит позади другой и они продолжают двигаться дальше. Вы на Марсе, в миллионах километров от дома. От гибели в холодной марсианской пустыне вас защищают хрупкие мембраны земной технологии, но покинуть планету вы не можете – ваш космический корабль уже не починить. Вам никогда не вернуться на Землю, к друзьям и близким, в родные края.

Впрочем, надежда есть. В коммуникационном отсеке разбитого корабля вы находите телепорт “Телеклон Марк IV” и инструкцию к нему. Если включить телепорт, направить его луч на приемник “Телеклона” на Земле, а затем войти в камеру отправки, телепорт быстро и безболезненно расщепит ваше тело и создаст молекулярную схему для отправки на Землю, где приемник, резервуары которого наполнены необходимыми атомами, почти мгновенно – по переданной инструкции – создаст вас! Вы со скоростью света вернетесь на Землю и окажетесь в кругу друзей и близких, которые совсем скоро будут затаив дыхание слушать рассказы о ваших приключениях на Марсе.

Еще раз обследовав поврежденный корабль, вы приходите к выводу, что “Телеклон” – ваш единственный шанс. Терять вам нечего, поэтому вы включаете передатчик, настраиваете его и входите в камеру отправки. Пять, четыре, три, два, один, ВЖУХ! Вы открываете дверь, выходите из камеры приемки “Телеклона” и оказываетесь в знакомой, солнечной атмосфере Земли. Вы вернулись домой, целая и невредимая, хотя для этого вам и пришлось упасть на Землю с самого Марса. Чтобы отпраздновать чудесное спасение от жуткой гибели на красной планете, вы собираете всех друзей и близких и замечаете, что все они изменились с последней встречи. В конце концов, прошло почти три года – и никто не помолодел. При взгляде на свою дочь Сару, которой теперь, должно быть, восемь с половиной, вы ловите себя на мысли: “Неужели это та же девочка, которая сидела у меня на коленях?” Само собой, это она, решаете вы, хотя вам и приходится признать, что вы не столько узнаете ее по памяти, сколько угадываете, кто она такая. Она теперь гораздо выше, выглядит гораздо старше и знает гораздо больше. В общем-то, с момента вашей последней встречи успела смениться даже большая часть клеток ее тела. Но несмотря на внешние перемены, несмотря на замену ее клеток, она осталась тем же человеком, которого вы поцеловали на прощание три года назад.

И тут вы задумываетесь: “А тот ли я человек, который три года назад поцеловал на прощание эту маленькую девочку? Я мать этой девочки или совершенно новый человек, которому всего несколько часов от роду, хотя я и наделена воспоминаниями – или мнимыми воспоминаниями – о прошлых днях и годах?” Может, мать этой девочки недавно погибла на Марсе, расщепленная и уничтоженная в камере отправки “Телеклона Марк IV”?

Может, я погибла на Марсе? Нет, я определенно не погибла на Марсе, ведь я жива на Земле. Впрочем, возможно, *кто-то* все же погиб на Марсе – и это мать Сары. В таком случае я не мать Сары. Но я должна ею быть! Я решила воспользоваться “Телеклоном” именно затем, чтобы вернуться домой, к своей семье. Но я все время забываю, что я, возможно, и не пользовалась “Телеклоном” на Марсе. Возможно, это был кто-то другой – если это случилось вообще.

Считать ли эту адскую машину телепорт<sup>ом</sup> – средством транспортировки – или, как намекает ее название, каким-то убийственным производителем клонов? Выжила ли мать Сары после взаимодействия с “Телеклоном”? Она думала, что выживет. Она вошла в камеру с надеждой и предвкушением, а не с суицидальным отчаянием. Само собой, она действовала из альтруистических побуждений – она желала, чтобы у Сары была мама, которая будет ее защищать, – но в то же время проявляла и эгоизм, ведь она выбиралась из переделки и возвращалась в любимый мир. По крайней мере, так казалось. “Откуда мне знать, что именно так и казалось? Потому что я была *там*. Я была матерью Сары, которая обдумывала все это. Я и *есть* мать Сары. По крайней мере, так кажется”.

Несомненно, телепортировать можно песню, стихотворение или фильм. Можно ли считать самость сущностью, “состоящей из информации”, которую можно телепортировать без потерь? Может, наше нежелание признавать телепортацию людей сродни недавно преодоленному в большинстве сфер анахроническому противостоянию официальным электронным подписям на документах? (В 2011 г. я узнал, что Ученое общество Гарвардского университета не готово принять отсканированную подпись на моем рекомендательном письме – они требовали чернильную подпись, выполненную мною собственноручно, из-за чего мне пришлось полдня кататься на такси по Бейруту, чтобы получить,

подписать и отправить экспресс-почтой обратно нужную форму на дорогой бумаге кремового цвета. Насколько я понимаю, теперь Общество изменило свои правила, однако я надеюсь, что Гарвард по-прежнему скрепляет дипломы восковыми печатями. Кое-где есть место традициям, во всей их великолепной бессмысленности.)

## 62. Самость как центр нарративной гравитации

Что есть самость? Философы веками бьются над этим вопросом. Христианская концепция бессмертной души, нематериальной и необъяснимой, столетиями пленяла мыслителей и ускользала от серьезного изучения, но с каждым днем приверженцев у нее все меньше. Идея о духовной сущности, которая после смерти отправляется в рай, кажется все более противоречивой. Очевидно, что мы не отказываемся от нее окончательно, как отказались от мысли о существовании гоблинов и ведьм, только потому, что все еще тешим себя иллюзиями. Те же из нас, кто придерживается материалистических взглядов и уверен, что сознание есть мозг (природа которого вполне понятна), вынуждены отвечать на вопрос, почему *кажется*, что каждый из нас обладает какой-то духовной сущностью, или и того лучше – почему каждый из нас представляет собой такую духовную сущность, живущую в теле, а точнее, в мозге. Может, чтобы найти самость, нужно просто заглянуть внутрь себя?

Дэвид Юм, как известно, раскритиковал эту идею в 1739 г.:

Что касается меня, то, когда я самым интимным образом вникаю в нечто, именуемое мной *своим я*, я всегда наталкиваюсь на то или иное единичное восприятие тепла или холода, света или тени, любви или ненависти, страдания или наслаждения. Я никак не могу уловить *свое я* как нечто существующее помимо восприятий... Если же кто-нибудь после серьезного и непредубежденного размышления будет все же думать, что у него иное представление о *своем я*, то я должен буду сознаться, что не могу дальше спорить с ним. Я могу лишь допустить, что он так же прав, как и я, и что мы существенно отличаемся друг от друга в данном отношении. Он, быть может, и сознает в себе нечто простое и непрерывное, которое и называет *своим я*, тогда как я уверен, что во мне такого принципа нет. <sup>[73]</sup> [1964, I, iv, разд. 6]

Юм не без иронии признает, что представления людей могут различаться, и его слова по сей день резонируют среди тех, кто гадает – а что, если я, к примеру, зомби (а точнее, зимбо), который наивно судит людей по себе? Такая мысль весьма любопытна, но вряд ли кто-то воспринимает ее всерьез.

Совершенно ясно, чем самость не является. Самость не часть мозга, в отличие от миндалевидного тела и гиппокампа. Лобная доля играет важнейшую роль в оценке ситуаций, намерений, восприятий и тому подобного, но я сомневаюсь, что хоть кто-то ошибочно *размещал в ней самость*. (Губительная префронтальная лоботомия превращает человека в его “тень”, однако в ходе этой операции не вырезается самость. Как гласит старая шутка, лучше уж одуреть от пьянства, чем подвергнуться префронтальной лоботомии, но в любом случае этому опыту подвергнется самость.) Так что же такое самость? Я полагаю, что она сродни центру гравитации – абстракции, которая, несмотря на свою абстрактность, тесно связана с физическим миром. Вы, как и любой другой материальный объект, обладаете центром гравитации (корректнее называть его центром масс, но здесь мы не станем обращать внимания на эту формальность). Если у вас массивный торс, ваш центр гравитации находится выше, чем в среднем у людей вашего роста, поэтому вам приходится прилагать больше усилий, чтобы прямо стоять, и так далее. Существует много способов найти свой центр гравитации, который – в зависимости от таких факторов, как тип вашей обуви и время последнего приема пищи, – занимает плавающее положение в небольшой зоне посреди вашего тела. Это математическая точка, а не атом и не молекула. Центр гравитации отрезка стальной трубы не сделан из стали – и вообще не из чего не сделан. Это точка в пространстве – точка на средней линии, проходящей через центр трубы, равноудаленная от ее концов (примерно, в зависимости от несовершенств и т. п.).

Концепция центра гравитации – сама по себе полезный инструмент мышления. По сути, она усредняет все гравитационные силы между всеми частицами материального объекта и всеми частицами материи на планете и говорит нам, что мы можем свести все это к двум точкам – центру земли (ее центру гравитации) и центру гравитации объекта – и рассчитать поведение объекта в меняющихся условиях. К примеру, если центр гравитации объекта в какой-то момент окажется за пределами всех точек его основания, объект упадет. Само собой, интуитивно мы понимали центры гравитации задолго до того, как Ньютон постиг саму гравитацию. (“Сядь! Ты раскачиваешь лодку”».) Теперь мы можем во всех подробностях объяснить, как и почему работает эта концепция, а при проектировании торшеров или транспортных средств стараемся разместить центр гравитации как можно ниже или удобнее – иными словами, без центров гравитации никуда. Может, это и “теоретический вымысел”, но очень ценный

вымысел, на основании которого можно сделать множество истинных предсказаний. Разве может такая абстрактная сущность, лишенная материального воплощения, становиться *причиной* чего-либо? Только не напрямую. И все же объяснения, ссылающиеся на центр гравитации, конкурируют с очевидно каузальными объяснениями. Почему кофейная чашка не перевернулась, когда корабль так сильно накренился? Ответ “потому что у нее необычно низкий центр гравитации” конкурирует с ответом “потому что она приклеена к палубе”.

Центр гравитации можно назвать теоретическим *вымыслом*, поскольку ему, как и вымышленным персонажам, свойственна любопытная неопределенность характеристик. Шерлок Холмс из детективных рассказов о Шерлоке Холмсе Артура Конана Дойла наделен множеством характеристик, однако по тем вопросам, по которым Конан Дойл не высказался, предмета для обсуждения нет. Можно продолжить эти рассуждения: писатель ни разу не упоминает, что у Шерлока была третья ноздря, так что мы вправе предположить, что третьей ноздри у него не было (Lewis 1978). Также мы можем согласиться, что он не был двоежнецом с одной женой в Париже и другой в Нью-Йорке. Но на многие подобные вопросы ответов нет: была ли у него родинка на левой лопатке? приходился ли он двоюродным братом Оскару Уайльду? был ли у него коттедж в Шотландии? Если речь идет о *реальном* человеке, на все эти вопросы – и бесчисленное множество других – должны существовать верные ответы, даже если нам и не дано их найти. Но с Шерлоком все иначе: будучи вымышленным персонажем, он наделен только теми характеристиками, которые называет или подразумевает писатель. Наивный читатель, который полагает, что рассказы о Шерлоке Холмсе правдивы, может поинтересоваться, был ли проводник поезда, идущего в Олдершот, выше или ниже Холмса, но, если вы понимаете, что такое вымысел, вам и в голову не придет задавать такой вопрос. Точно так же и с центрами гравитации. Задуматься, не случится ли так, что в конце концов они “окажутся нейтрино”, – значит упустить из виду, что они, по сути, лишь теоретический вымысел.

Что же такое центр *нарративной* гравитации? Это тоже теоретический вымысел, призванный унифицировать и объяснить в ином случае поразительно сложный набор действий, высказываний, волнений, жалоб, обещаний и т. п., из которых состоит личность. Именно он лежит в основе объяснений на личном уровне. Ваша рука не подписывала контракт – его подписали вы. Ваш рот не солгал – солгали вы. Ваш мозг не помнит Париж – его помните вы. Вы “номинальный собственник” живого тела, которое мы считаем вами. (Как говорится, это ваше тело – делайте с ним что угодно.) Подобно тому как мы можем упростить все гравитационные силы всех элементов мира и стоящего на земле обелиска, сведя их к двум точкам – центру земли и центру гравитации обелиска, – мы можем упростить и все взаимодействия – рукопожатия, произносимые слова, чернильные каракули и многое другое, – сведя их к двум самостям, продавцу и покупателю, которые только что произвели транзакцию. Каждая из самостей – это личность, имеющая собственную биографию, “предысторию” и множество проектов в работе. В отличие от центров гравитации, самости не просто обладают траекториями во времени и пространстве, а формируются на ходу в процессе накопления воспоминаний, разработки планов и определения ожиданий.

От некоторых фрагментов предыстории человек, возможно, предпочел бы отречься, но сделанного не воротить – этот фрагмент нарратива не переделать. Тем не менее его можно заново интерпретировать в свете более поздних элементов биографии. Нам хорошо знакома фраза “я был сам не свой, когда это сделал”, и часто наш скепсис по отношению к этому внутренне противоречивому утверждению вполне обоснован. Говорящий так человек имеет в виду, что он действовал совсем “не в своем духе”, и просит нас не судить его на основании этого и не думать, что он снова поступит так же в будущем. Порой его заявления вполне правдоподобны, порой – нет. Нам также знакома другая отговорка: “Если не ты, то кто это сделал? – Меня бес попутал”. И снова мы часто принимаем это не за чистую монету, а в качестве искреннего отречения от черт характера и мотиваций, подтолкнувших к совершенному действию. В результате встают вопросы об ответственности и свободе воли, которыми мы займемся в следующей главе. Пока же обратите внимание, что концепция самости необходима нам, чтобы провести черту (пусть часто и произвольным образом) между тем, что вы делаете, и тем, что с вами происходит.

Любое физическое тело обладает центром гравитации, а любое живое человеческое тело обладает самостью – или, скорее, любое живое человеческое тело принадлежит самости, которая выступает в качестве своеобразного администратора с проживанием. Один ли владелец у тела? Могут ли в одном теле ужиться две и более самости? Заболевания, называемое диссоциативным расстройством личности (а ранее – расстройством множественной личности), судя по всему, дает нам пример существования нескольких самостей в одном теле, при котором одна самость (“хозяин”) доминирует, имея группу “альтернативных” самостей. Я говорю “судя по всему”, потому что вокруг диагноза не утихают споры: одни считают, что это откровенное надувательство через индуцированное бредовое расстройство (когда



наивный психиатр непреднамеренно подталкивает больного к проявлению определенных симптомов), а другие признают несколько редких и истинных случаев заболевания, отличая их от игры всевозможных *подражателей*. После нескольких лет изучения феномена (и наблюдения за людьми, которые изучают и лечат эту болезнь) мы с психологом Николасом Хамфри (1989) решили, что правы все! Встречаются и надувательства, и преувеличения, и легковые психиатры с упрямыми пациентами – но и немногочисленные случаи, в которых болезнь, похоже, существовала хотя бы в рудиментарной форме и до вмешательства заинтригованного посредника. В этом нет ничего удивительного, если рассматривать это состояние в качестве более интенсивного проявления вполне нормального феномена, с которым в той или иной степени сталкиваемся мы все. Большинство из нас ведет несколько довольно разных жизней – на работе, дома, за игрой – и в каждом из контекстов приобретает привычки и воспоминания, которые не переходят ни в один из других контекстов.

Как заметил в своей классической работе “Представление себя другим в повседневной жизни” социолог Ирвинг Гофман (1959), мы все представляем себя героями в жизненных сценариях (профессор Деннет, Дэн из соседнего дома, папа, дедушка и так далее) и без труда заручаемся поддержкой героев второго плана, которые представляют себя подобным образом. Мы даем друг другу фразы, которые легко можем переварить, подыгрываем друг другу в ходе представления себя или ломаем продуманные сценарии, выходя из роли, что приводит к неловким, комичным или неприятным моментам. Для этого нужны стальные нервы. Можете ли вы представить, как вас на вечеринке знакомят с человеком, а вы просите его показать какое-нибудь удостоверение личности – скажем, водительские права или паспорт – или наоборот сразу бросаетесь к нему в объятия? Оказываясь в неудобных ситуациях, люди порой прибегают к отчаянным мерам, и случается, что их притворство со временем становится их второй натурой. В тяжелых обстоятельствах вы просто *отступаете*, оставляя после себя другой центр нарративной гравитации, другого героя, который лучше способен разобраться с насущной проблемой.

Мы с Хамфри обнаружили, что пациенты, предположительно страдающие от диссоциативного расстройства личности, часто демонстрируют виртуозное мастерство при уходе от ответов на неудобные вопросы. Чтобы формулировать вопросы, которые не могут остаться без ответов<sup>[74]</sup>, или – что даже лучше – расставлять ловушки и проверять, действительно ли одна из альтер-личностей не помнит ничего из того, что делала и говорила другая альтер-личность, нужно быть откровенно грубым, рискуя серьезно обидеть собеседника, поэтому высока вероятность, что вы из вежливости будете подыгрывать пациенту, выступая в качестве завербованного в два счета сообщника. Мошенники занимаются этим намеренно, демонстрируя немалое мастерство, но так же поступают и невинные жертвы описываемого расстройства личности, которые даже не сознают, что делают. В той или иной степени мы все этим занимаемся. Но *их* альтер-личности – лишь вымышленные персонажи, да? *Реальная* личность – только хозяин, да? На самом деле все не так очевидно.

Та сущность, которая является *вами*, – та личность, которой вы себя представляете, играя при этом хоть несколько ролей, хоть одну монолитную роль, – выступает в качестве центра вашей нарративной гравитации. Именно так вас узнают друзья (“Сегодня ты сам не свой!”) и так вы сами видите себя большую часть времени, пусть даже этот образ несколько идеализирован (“О боже! Это я сделал? Я бы никогда такого не сделал!”). Профессиональные писатели, подобно мошенникам, создают нарративы, мастерски снабжая их деталями. Мы же предстаем талантливыми любителями, рассказывая свои истории с умом, но (в основном) безотчетно, как паук, плетущий паутину. Это природа, а не искусство. Скорее не *мы*, используя мозг, плетем небылицы, а наш мозг, используя небылицы, сплетает *нас*. Да, в основе всего лежат неопровержимо истинные биографические сведения, однако с годами немалые фрагменты этих сведений теряют значимость, отходят на второй план и перестают определять вас. Некоторые из них вы специально отбрасываете, отвергаете, “забываете” в процессе самообновления и самосовершенствования<sup>[75]</sup>.

Подумайте, как легко вам ответить на некоторые вопросы о своем прошлом. Танцевали ли вы когда-нибудь с кинозвездой? Бывали ли вы в Париже? Катались ли вы хоть раз на верблюде? Случалось ли вам задушить человека голыми руками? Большинству из нас не составляет труда дать ответ на подобные вопросы, будь он хоть положительным, хоть отрицательным. (Представьте, что кто-то задумался, услышав последний вопрос, и почесал подбородок, прежде чем ответить. Держитесь от него подальше!) Мы знаем ответы на эти вопросы, потому что знаем о себе – о своей самости – достаточно, чтобы понимать, что мы бы точно запомнили, если бы хоть раз танцевали с кинозвездой, бывали в Париже, катались на верблюде или душили человека. Если “ничего не приходит в голову”, мы считаем, что ответ отрицательный. (Как иначе мы можем быть уверены в этом? Разве вы ведете список всех вещей, которые хоть раз делали? Всех мест, где хоть раз бывали?) Сравните эти вопросы с такими, казалось бы, сходными

вопросами: вы когда-нибудь танцевали с человеком по фамилии Смит? вы когда-нибудь бывали в аптеке, где продавалась мастика? вы когда-нибудь ездили на синем “Шевроле”? вы когда-нибудь разбивали белую кофейную чашку? Вы без труда ответите на некоторые из них, но другие вынудят вас признаться, что вы понятия об этом не имеете, а третьи и вовсе подтолкнут вас дать неверные ответы, хотя вы сами этого и не поймете, просто потому что все это совсем неважно. Зачем вам помнить, делали ли вы что-то из перечисленного? Многие из того, что случается с нами, вообще не памятно. Наши центры нарративной гравитации скрывают множество хороших и плохих вещей, которые мы делали в своей жизни, но к итоговому нарративу невинным образом добавляется и многое из того, чего никогда не происходило, потому что по той или иной причине это идеально вписывается в наш образ. Вы представляете собой совокупность способностей и опыта, суровых намерений и легковесных грез, заключенную в единственном мозге и теле и называемую вашим именем. Мысль о том, что существует также и особенное, нерушимое ядро *самости* – или эго, или дух, или душа, – всего лишь чарующая фантазия, в которой нет необходимости, если мы хотим понять людей, их мечты и надежды, доблести и грехи.

Если же центр нарративной гравитации не таинственный фрагмент элементарного вещества, а абстракция, можно ли исследовать его с научной точки зрения? Да.

### 63. Гетерофеноменология

Гетерофеноменология представляет собой не насос интуиции, а очередной элемент *строительных лесов*, которые стоит возвести, прежде чем мы перейдем к ряду сложных вопросов. Изучение человеческого сознания затрагивает феномены, которые, на первый взгляд, наблюдаются в каком-то другом измерении – в частном, субъективном измерении, где все описывается “от первого лица”. Каждый из нас пребывает в этом измерении по отношению к собственному сознанию, и никто более не может получить в него непосредственный доступ. Каким же образом соотносятся стандартные объективные методики, применяемые “от третьего лица” при изучении метеоритов и магнитов (или человеческого обмена веществ и плотности костей), и методики для изучения человеческого сознания? Должны ли мы создать радикально новую, революционную альтернативную науку – или же стандартные методы можно распространить и на изучение феноменов сознания? Я убежден, что объективную науку можно непосредственным, консервативным образом распространить на *всю* сферу человеческого сознания, в полной мере учитывая все данные и не отказываясь от правил и ограничений экспериментальных методов, прекрасно работающих в остальных областях науки. Объективная методология, гетерофеноменология (феноменология *другого*, а не самого себя), – надежный способ воспринять *самоописание* субъекта так серьезно, как это только возможно.

Зачем такое длинное название? “Феноменологией” изначально называли каталог феноменов различного типа, пока не предлагалась хорошая теория для их объяснения. В шестнадцатом веке Уильям Гильберт составил хорошую феноменологию магнетизма, но объяснить все тщательно описанные им феномены магнетизма удалось лишь несколькими веками позже. В начале двадцатого века Эдмунд Гуссерль – а вслед за ним и группа находящихся под его влиянием психологов и философов – назвал “Феноменологией” (с заглавной буквы “Ф”) предположительно научное исследование феноменов субъективного опыта, наблюдаемых при интроспекции “от первого лица” в отрыве от каких-либо теорий и предпосылок. Это направление научной мысли существует и по сей день, хотя по тем или иным причинам оно постоянно подвергается притеснениям или игнорируется. Несмотря на любопытные результаты, заслуживающие дальнейшего исследования, феноменология по-прежнему считается субъективным подходом и подавляется объективной, эмпирической наукой, требующей, чтобы данные были доступны всем исследователям. Но сознание можно изучать объективно, применяя метод, который представляет собой простую вариацию на тему феноменологии, в связи с чем я называю его гетерофеноменологией – в отличие от *автофеноменологии* Гуссерля. Гетерофеноменология – это изучение субъективных феноменов со сторонней точки зрения объективной науки.

Очевидно, ключевое отличие экспериментов с камнями, розами и крысами от экспериментов с пребывающими в сознании и идущими на контакт людьми заключается в том, что последние могут общаться при помощи языка, а следовательно, сотрудничать с экспериментаторами, давая им рекомендации, взаимодействуя с ними вербально и сообщая им, как они чувствуют себя в различных контрольных условиях. В этом суть гетерофеноменологии: она использует нашу способность к *речи* и ее *интерпретации*, чтобы составить каталог того, что субъект считает истинным о своем сознательном опыте. Этот каталог убеждений детализирует *гетерофеноменологический мир* субъекта – мир глазами субъекта, субъективный мир субъекта. Вся совокупность деталей гетерофеноменологии в сочетании со всеми данными, которые мы можем собрать о сопутствующей активности мозга субъектов и событиях окружающей среды, составляет полный набор данных, подлежащих объяснению силами теории человеческого сознания. За рамками этого набора не остается ни объективных, ни субъективных феноменов сознания.

В ходе интерпретации *сырые* данные, полученные из речи и нажатия на кнопки в ходе эксперимента, превращаются в отчеты и описания убеждений, что предполагает принятие **интенциональной установки**, требующей наличия рабочей гипотезы, что субъект – это агент, действия которого рационально управляются убеждениями и желаниями, в свою очередь рациональными, с учетом нужд и истории восприятий субъекта. К примеру, ограничения интенциональной установки становятся очевидны, когда в подобных экспериментах предпринимаются стандартные предосторожности, чтобы субъекты не получали опыт, который может способствовать возникновению убеждений или желаний, влияющих на ответы субъектов таким образом, что это искажает нашу интерпретацию их действий: так, мы не говорим им, что надеемся от них услышать, но в то же время проверяем, понимают ли они поставленную перед ними задачу. Такое принятие интенциональной установки нельзя назвать необратимо

субъективным или релятивистским. Можно определить правила интерпретации, принять интерсубъективное соглашение об интерпретации, выявить отклонения от нормы, а также с оглядкой счесть корректируемым, состоятельным и эволюционно объяснимым неизбежное допущение о рациональности. (Подробнее об этих процессах см. в работе Dennett 1991a.)

Я не предлагаю новую методологию для изучения сознания. Я просто обращаю внимание на стандартные методы, уже используемые исследователями в когнитивной психологии, психофизике (которая изучает отношения между физическими стимулами и субъективными реакциями) и нейробиологии, объясняю и защищаю их. При правильном понимании и применении этих методов необходимость в радикально новой или революционной объективной науке о сознании исчезает, поскольку не остается ни одного феномена сознания, не поддающегося контролируемому научному исследованию.

Какие вещи учитывает эта методология? Помимо стандартных вещей, которые учитывает любая наука (нейроны и электроны, часы и микроскопы), она учитывает только *убеждения* – убеждения, выражаемые субъектами и признаваемые конститутивными для их субъективности, – и *желания* – желания сотрудничать с экспериментаторами и говорить им правду как можно более искренне. (При использовании этого метода важно также контролировать эти убеждения и желания и отказываться от результатов тех экспериментов, где они не были под контролем.) Что есть убеждения и желания? Пока теория не получила подтверждения, мы можем давать максимально расплывчатые определения, считая убеждения и их содержание *теоретическим вымыслом* или *абстракцией*, подобно центрам масс, экватору и параллелограммам сил.

Встречи с русалками реальны, но описаны неверно, поскольку русалок не существует. Подобным образом каталог убеждений об опыте не тождествен каталогу самого опыта. Философ Джозеф Левин (1994, р. 117) возразил, что “первичные данные для формулировки теории дает сознательный опыт как таковой, а не просто наше вербальное суждение о нем”. Это не может быть правдой. Разве можем мы, не имея теории, каталогизировать опыт? Рассмотрим свидетельства, получаемые от испытуемых, которых мы помещаем в экспериментальные ситуации и расспрашиваем об их опыте (а также просим осуществить другие необходимые действия). Эти источники естественным образом вписываются в каскад интерпретаций, которые нам необходимо сделать, расположенных ниже в порядке от наименее к наиболее сырой:

- (а) “сознательный опыт как таковой”
- (б) убеждения об этом сознательном опыте
- (в) “вербальные суждения”, о которых упоминает Левин
- (г) высказывания того или иного типа, которые можно интерпретировать как описывающие эти вербальные суждения.

В некотором смысле в качестве первичных данных выступают именно высказывания – записанные звуки и движения. В зависимости от обстоятельств к первичным данным можно также добавить результаты электроэнцефалографии (ЭЭГ), функциональной магнитно-резонансной томографии (фМРТ) и подобных исследований. Надежные методы интерпретации могут привести нас к (в) и (б) – и мы получим каталог убеждений субъекта о том, что значит быть им в этих условиях. Но стоит ли продвигаться к (а) в отсутствие теории? Вряд ли – и на то есть две причины.

Во-первых, если (а) выходит за рамки (б) – если вы имеете сознательный опыт, не будучи в этом убежденным, – то этот дополнительный сознательный опыт недоступен *вам* точно так же, как и сторонним наблюдателям.

Таким образом, альтернативный метод, предложенный Левином, позволяет получить не больше пригодных к использованию данных, чем гетерофеноменология.

Во-вторых, если (б) выходит за рамки (а) – если вы убеждены, что имеете сознательный опыт, которого на самом деле не имеете, – то объяснять надо ваши убеждения, а не отсутствующий опыт.

В таком случае остается единственный способ избежать получения ложных данных и учесть при этом все феномены, доступные *кому угодно*, – необходимо придерживаться гетерофеноменологического стандарта и считать (б) максимальным набором *первичных* данных.

Что если некоторые убеждения невыразимы в вербальных суждениях? Ничто не мешает гетерофеноменологам и испытуемым использовать аналоговые или другие нелингвистические формы выражения убеждений. Например:

Нарисуйте вертикальную черту на изображенной линии, чтобы обозначить, насколько  
глубок [в том или ином измерении] испытываемый опыт:

Едва заметный Всепоглощающий

Испытуемые также могут с разной силой нажимать на кнопку, обозначая остроту боли (или степень тревоги, или скуки, или даже недоверия к эксперименту). Существует также множество поддающихся измерению физиологических показателей, от кожно-гальванической реакции и частоты сердечных сокращений до перемен в выражении лица и позе. Если вы, испытуемый, полагаете, что даже при использовании всех этих методов *все равно* учитываются не все элементы вашего опыта, вы можете сказать об этом гетерофеноменологам, которые добавят это убеждение к списку убеждений в ваших первичных данных:

Субъект утверждает, что у него остались невыразимые убеждения насчет X.

Если это убеждение истинно, наука обязана объяснить, что это за убеждения и почему они невыразимы. Если же убеждение ложно, наука все равно обязана объяснить, почему субъект полагает (ошибочно), что у него есть именно такие невыразимые убеждения<sup>[76]</sup>.

## 64. Исследовательница цвета Мария: разоблачение упорного башмака

Предложенный австралийским философом Фрэнком Джексоном мысленный эксперимент об исследовательнице цвета Марии, который часто называют “аргументом о знании”, не выходит у философов из головы с момента своего появления в 1982 г. Масштаб и надежность этого эксперимента позволяют назвать его одним из самых успешных насосов интуиции, когда-либо сконструированных философами-аналитиками. Он давно признан классикой и неизменно включается во все списки обязательной литературы бакалаврских курсов по философии сознания во всем англоязычном мире, а статьи о его следствиях собраны в несколько увесистых антологий. Любопытно, что автор впоследствии отрекся от своего эксперимента и заявил, что более не согласен с его выводами, но это не уменьшило популярности этого насоса интуиции.

Вот он во всей полноте, возможно, вырванный из контекста, но все равно предельно ясный:

Мария – выдающийся ученый. По какой-то причине ей приходится исследовать мир *через* черно-белый монитор, находясь при этом в черно-белой комнате. Она специализируется на нейрофизиологии зрения и получает, допустим, всю доступную физическую информацию о том, что происходит, когда мы видим спелые помидоры или небо и используем термины вроде “красный”, “синий” и т. п. К примеру, она обнаруживает, какие комбинации длин волн с неба стимулируют сетчатку и каким образом это *через* центральную нервную систему приводит к сжатию голосовых связок и выбросу из легких воздуха, в результате чего человек произносит: “Небо синее”. (Вряд ли можно возразить, что в принципе всю эту физическую информацию можно получить по черно-белому телевизору, ведь иначе “Открытый университет” *в силу обстоятельств* вынужден был бы использовать цветное телевидение.) Что случится, когда Мария выйдет из своей черно-белой комнаты или получит цветной телемонитор? *Узнает* ли она что-нибудь? Кажется вполне очевидным, что она узнает что-то о мире и нашем зрительном опыте. Но в таком случае придется признать, что прежде ее знание было неполным. Однако она обладала *всей* физической информацией. *Следовательно*, физической информации недостаточно, а физикализм [то есть материализм, отрицание дуализма] ошибочен. [Jackson 1982, p. 130]

Хороший ли это насос интуиции? Давайте покрутим все его регуляторы и поймем, как он работает. Нам пришлось бы потратить на это немало времени, но, к счастью, работа уже проделана в огромном количестве литературы по теме. Здесь я просто продемонстрирую несколько регуляторов, на которые стоит обратить внимание, и оставлю упражнение вам на откуп. (Если хотите, можете сверить свои результаты с литературой; в разделе “Источники” есть ссылки на две недавно опубликованные антологии. Сумеете ли вы найти новый поворот?) Более двадцати лет назад я провел пробное исследование регуляторов и пришел к неутешительному выводу, на который редко обращают внимание: “Суть этого эксперимента, как и любого хорошего мысленного эксперимента, сразу понятна даже непосвященным. Но на самом деле этот мысленный эксперимент очень плох – этот насос интуиции склоняет нас неправильно понимать его предпосылки!” (Dennett 1991a, p. 398). Проверим, так ли это. Я утверждаю, что верно представить себе сценарий эксперимента гораздо сложнее, чем кажется, поэтому люди упрощают его и делают выводы на неверных основаниях.

Первый регулятор: “через черно-белый монитор, находясь при этом в черно-белой комнате”.

Предположим, Мария носит черные или белые перчатки и не имеет права смотреть на себя саму, когда принимает ванну. И все же исключить взаимодействие с любыми “внешними источниками” цвета невозможно. Неужели нам пришлось бы изобрести устройство, которое не позволяло бы Марии тереть глаза (и создавать “фосфены” – попробуйте)? Разве она не могла бы видеть цвета во снах, прежде чем увидеть их на самом деле? Если нет, то почему? Должны ли цвета “попадать внутрь через глаза”, прежде чем она “сохранит” их в мозге? За этим простым предположением скрывается целый сонм проблематичных обывательских теорий.

Второй регулятор: она “получает, допустим, всю доступную физическую информацию о том, что происходит, когда мы видим спелые помидоры или небо и используем термины вроде «красный», «синий» и т. п.”.

*Всю доступную информацию? Сколько именно? Это как иметь все деньги мира? Что это значит? Представить это нелегко, но мысленный эксперимент не удастся, если у Марии будет не вся информация. Она должна располагать всей информацией о всех вариантах ответов во всех мозгах, включая ее собственный, и в частности – всей информацией обо всех эмоциональных или аффективных реакциях на все цвета во всех обстоятельствах. В таком случае она будет в мельчайших подробностях знать, какие цвета ее успокаивают, раздражают, нравятся ей тем больше, чем дольше она на них смотрит, отвлекают ее, отталкивают и так далее. Запрещено ли ей экспериментировать на себе (не жульничая и не пронося в черно-белую комнату цветные предметы)? Если вы не представили всего этого (и не только), вы не последовали инструкциям. Это все равно что представить круг, когда вас попросили представить тысячеугольник. Множество следствий одного из этих упражнений на психическое представление не вытекает из другого упражнения. В нашем случае, к примеру, должны ли мы не обращать внимания на тот факт, что если бы Мария получила всю эту информацию, то, несомненно, пришла бы в состояние полного психического упадка, отягощенная тысячами энциклопедических статей и диаграмм?*

Если бы Джексон заявил, что Мария обладает богоподобной способностью к “физическому всеведению” – причем знает все не только о цветах, но и о любом физическом факте на любом уровне от кварка до галактики, – многие (если не все) читатели возразили бы ему и сказали, что это слишком фантастично, чтобы воспринимать такое допущение всерьез. Но заявление о том, что Мария знает *лишь* все физические факты *о цветовом зрении*, по сути, не менее фантастично.

- Допустим, у Марии миллиард голов...
- Не глупи!
- Ладно, пусть у Марии будет миллион голов...
- Без проблем! (Серьезно?)

Ранее я пытался подчеркнуть эту проблему воображения, предлагая людям рассмотреть другой вариант концовки:

Однажды пленители Марии решили, что ей пора увидеть цвета. Чтобы обмануть ее, они приготовили ярко-синий банан, при взгляде на который она должна была получить свой первый цветовой опыт. Посмотрев на банан, Мария воскликнула: “Эй! Вы пытаетесь меня обмануть! Бананы желтые, а этот синий!” Ее пленители были потрясены. Как она это поняла? “Все просто, – ответила она. – Не забывайте, что я знаю *все* – абсолютно все, – что вообще можно знать о физических аспектах цветового зрения. Само собой, прежде чем вы принесли мне банан, я уже записала в мельчайших подробностях, какое именно физическое впечатление произведет на мою нервную систему желтый или синий (или зеленый и т. д.) предмет. Я заранее знала, какие именно *мысли* роятся у меня в голове (ведь, в конце концов, «простую диспозицию» при мысли о чем-то нельзя считать одной из хваленых квалиа, так?). Меня ни в коей мере не удивил мой опыт синего (однако меня удивило, что вы так глупо попытались меня провести). Я понимаю, *вам сложно представить*, что я знаю о своих реакционных диспозициях так много, что меня не удивила моя реакция на синий. Само собой, *вам сложно это представить*. Любому сложно представить следствия того, что кто-то владеет абсолютно всей физической информацией о чем угодно!” [Dennett 1991a, pp. 399–400]

Обычно считается, что так быть не может. Как обезоруживающе выразился Джексон, “кажется вполне очевидным, что [Мария] узнает что-то о мире и нашем зрительном опыте”. Джордж Грэм и Терри Хорган (2000, р. 72) замечают: “**Безусловно** [динь!], мы полагаем, что она удивится и обрадуется”. Это не так. Именно в этом и заключается проблема мысленного эксперимента о Марии. Мы так жаждем сделать вывод, что Мария получит *некоторое* откровение, впервые увидев цвет, что никто не хочет показать, что история просто должна развиваться именно так. На самом деле все могло бы быть совсем иначе.

Насос интуиции Джексона прекрасно проливает свет на множество наивных представлений о природе цветового опыта и мозга, которые обычно отлично служат людям, и стоит признать, что тем самым он изящно выводит ряд следствий обывательских теорий. Но его целью было опровергнуть

гипотезу о способности физических наук объяснять все цветовые феномены. Само собой, в реальном мире человек, оказавшийся на месте Марии, узнал бы что-то новое, потому что, сколько бы она ни знала о цвете, множество фактов о физическом воздействии цвета были бы ей неизвестны. Неуместными “вполне очевидно” Джексона и “безусловно” Грэма и Хорган становятся лишь потому, что мысленный эксперимент описывает исключительный случай. Если вы все равно склонны полагать, что предложенная мною альтернативная концовка истории невозможна, попробуйте убедительно аргументировать свою точку зрения. Будет любопытно, если вы озвучите умозаключение, к которому не пришел ни один из сотен философов, годами бившихся над этой задачей. (Конечно же, сам этот факт можно считать доказательством великолепия описанного насоса интуиции, ведь он три десятка лет обеспечивает философов работой.)



## Резюме

Научное изучение сознания осложняется тем фактом, что в этой области *каждый* мнит себя экспертом. Конечно же, на самом деле это не так, но почти любой, кто хоть на несколько минут задумывался на эту тему, склонен считать результаты своих размышлений столь же заслуживающими доверия, как результаты любого высокотехнологичного эксперимента или анализа целой горы статистики. Забавно слышать, как во время обсуждений на научных конференциях эти люди строго *исправляют ошибки*, замеченные в словах докладчика, ссылаясь на недавний опыт, который (как они сами думают) они получили. Если бы мы, как обычно полагают эти люди, не допускали ошибок при оценке природы собственного опыта, они были бы правы!

Но человек *может* неправильно запоминать, интерпретировать и описывать свой самый сокровенный опыт, подвергаясь неявному воздействию убедительной, но ненадежной идеологии. Вот простое упражнение, результаты которого могут вас удивить. Его несложно выполнить дома: сядьте перед зеркалом, чтобы следить за выполнением инструкций, которые заключаются в том, чтобы неотрывно смотреть себе в глаза, фиксируя на них свой взгляд и не отвлекаясь на происходящее на периферии. Не отрывая взгляда, вытяните карту из середины хорошо перетасованной колоды и держите ее *лицом к себе* на расстоянии вытянутой руки, как раз за границей вашего периферийного зрения. Потрясите карту. Вы будете знать, что карта трясется, но, конечно же, ничего не увидите. Медленно вводите карту в зону вашего периферийного зрения, потряхивая ею. Сначала вы увидите движение (тряску), *но не цвет!* Вы не сможете сказать, красной масти карта или черной, не сможете отличить фигурную карту и уж точно не сможете определить, каково ее численное значение. Перемещая карту все ближе к центру, вы удивитесь, как близко к прямой траектории зрения она должна оказаться, чтобы вы смогли назвать ее цвет или определить, фигурная ли это карта. По мере приближения карты к точке фиксации вашего взгляда вам придется прикладывать все больше усилий, чтобы не жульничать и не смотреть на движущуюся карту. Определить, какая карта у вас в руке, вы сможете лишь тогда, когда она окажется почти напротив вас. Удивлены? Я ни разу не встречал человека, который не удивился бы, впервые испытав это на собственном опыте. Вам всегда казалось, что вы видите примерно одинаковое количество цветов и деталей до самой границы поля зрения, а теперь вы узнаете, что, хоть это и “логично” и вроде бы подтверждается обычной интроспекцией, на самом деле это не так. И это лишь один аспект, один из многих феноменов, демонстрирующих, что наше богатое, непрерывное, детализированное представление о мире в нашем (визуальном) сознании – всего лишь иллюзия. Совет прост: не думайте, что понимаете феномены сознания, пока не узнаете, что выяснила о них наука в последние годы. Кабинетные теории философов, не принимающих этот совет во внимание, в лучшем случае ничего не стоят, а в худшем – только запутывают все сильнее и сбивают всех с толку. “Посредством интроспекции” вы можете “узнать” о своем сознании лишь малую, но способную кого угодно ввести в заблуждение толику того, что мы можем узнать о вашем сознании, систематически изучая сознание по стандартам гетерофеноменологии.

Предстоит преодолеть еще множество сложностей и решить множество трудных проблем, которые не считаются Трудной проблемой сознания. Если после решения всех этих “легких” проблем тайна останется тайной, нам придется пересмотреть свою отправную точку и отказаться от текущих представлений о биологии, физике и даже логике. Пока же посмотрим, как далеко нас заведет привычная наука – наука, которая помогла нам понять все, что мы понимаем сегодня, от астероидов и тектонических плит до размножения, роста, регенерации и метаболизма живых организмов.

## VIII.

### Инструменты мышления о свободе воли

Пропасть между **манифестной** и **научной картиной мира** особенно пубока, когда речь идет о свободе воли. Подобно вопросам, что есть *на самом деле* цвет и что есть *на самом деле* доллар, вопрос, иллюзорна или нет свобода воли, подталкивает нас использовать научную картину мира, чтобы исследовать проблему, сформулированную в традиционных терминах манифестной картины мира. И в последние годы к этому склоняются многие. Целый ряд именитых ученых решительно заявляет, что свобода воли иллюзорна: так считают нейробиологи Вольф Зингер, Крис Фрит и Патрик Хаггард, психологи Пол Блум и Дэниел Вегнер, а также несколько весьма уважаемых физиков, включая Альберта Эйнштейна и Стивена Хокинга. Может ли ошибаться такое количество светил? Множество – не все, а возможно, и не большинство – философов утверждает, что может. Они называют это задачей философии! Правы ли они? Думаю, да.

Как правило, ученые совершают глупую ошибку: они путают манифестную картину мира с тем, что можно назвать *народной идеологией* манифестной картины мира. Давайте признаем, что народная идеология цвета – просто чепуха. Цвет не таков, каким его представляет большинство людей, но это не значит, что манифестная картина мира лишена цветов: это значит, что цвета – настоящие цвета – не такие, какими их считают люди. Народная идеология сознания – тоже чепуха, намеренно двойственная и таинственная; если бы сознание *таким* и было, то Райт был бы прав (см. главу 59): нам пришлось бы сказать, что сознания не существует. Но мы не должны считать сознание “настоящей магией” – той самой, которой не существует, состоящей из **чудо-ткани**; мы можем признать реальность сознания феноменом, заметив, что народ еще не сформировал его четкой идеологии. Подобным же образом свобода воли представляет собой не то, чем ее провозглашает *народная идеология* манифестной картины мира, и не выходит магическим образом за рамки каузальности. В этом смысле я сравнил свободу воли с левитацией, и один из философских защитников этой абсурдной точки зрения честно назвал свободу выбора “маленьким чудом”. Я полностью согласен с целым сонмом ученых, утверждающих, что *такая* свобода воли иллюзорна, однако это не значит, что свобода воли иллюзорна в моральном смысле. Она не менее реальна, чем цвета и доллары.

К несчастью, некоторые ученые, утверждающие, что наука доказала иллюзорность свободы воли, полагают также, что это “открытие” *имеет* важность с этической точки зрения. Они полагают, что его следствия многое значат для этики и закона: к примеру, никто и никогда *на самом деле* не несет ответственности за свои поступки, а следовательно, никто и никогда *не заслуживает* наказания или похвалы. Их ошибка сродни ошибке людей, которые утверждают, что ничто *на самом деле* нельзя считать твердым телом. Они используют неадаптированную популярную концепцию свободы воли, хотя сперва ее необходимо скорректировать, как делается с концепциями цвета и сознания (а также пространства, времени, твердости и другими концепциями, неправильно представленными в *идеологии* манифестной картины мира).

Описанные в этом разделе насосы интуиции помогут вам отказаться от этой идеологии о свободе воли и увидеть лучшую концепцию – концепцию реальной свободы воли, практической свободы воли, важного феномена манифестной картины мира. Споры о свободе воли не утихают тысячелетиями – они слишком многочисленны и слишком запутанны, чтобы разрешить их одним махом в одном разделе или даже в целой книге, – и все же нам необходимо с чего-то начать, поэтому предлагаемые здесь инструменты мышления должны на манер гвоздодера вырвать вас из привычного гнезда и тем самым открыть вам новые горизонты. Первый из них призван показать, почему это так важно.

## 65. Подлый нейрохирург

Нейрохирургическое лечение тяжелых психических заболеваний уже не за горами. Так, глубокая стимуляция головного мозга имплантированными электродами поразительно эффективна при лечении обсессивно-компульсивного расстройства (ОКР), о чем свидетельствует прорывное исследование нейропсихиатра Дамиана Дени и его коллег (2010) из Амстердама. Это факт, а вот вымысел: однажды блестящий нейрохирург сказала пациенту, которому только что имплантировала электроды в своей прекрасной высокотехнологичной операционной:

Устройство, которое я установила, не контролирует ваше ОКР. Оно контролирует каждое ваше решение благодаря нашей центральной системе контроля, которая круглосуточно поддерживает радиокontakt с вашим микрочипом. Иными словами, я отключила вашу сознательную волю. Отныне ваше ощущение свободы воли будет иллюзией.

На самом деле она ничего подобного не сделала, но решила солгать пациенту, чтобы посмотреть, что из этого выйдет. Несчастный пациент поверил ей и вернулся к жизни, убежденный, что отныне не несет ответственности за свои действия, поскольку он более не агент, а лишь марионетка. Изменения в поведении не заставили себя ждать: он стал безответственным, агрессивным и небрежным и потворствовал всем своим капризам, пока его не поймали и не представили суду. Свидетельствуя в свою защиту, он отчаянно оспаривал свою безответственность, ссылаясь на установленный в его мозге имплант, а когда для дачи показаний пригласили нейрохирурга, та признала свои слова и добавила: “Но я говорила не всерьез – я просто пошутила, не более. Откуда мне было знать, что он мне поверит!”

Неважно, чьим словам поверил суд и кого из них наказал, ведь факт остается фактом: нейрохирург в любом случае сломала пациенту жизнь своей неосторожной шуткой, лишив его целостности и парализовав его способность к принятию решений. Фактически с помощью этого ложного “инструктажа” она нехирургическим путем добила многого из того, чего, по собственному утверждению, добила хирургически: она искалечила пациента. Однако если она несет ответственность за ужасные последствия своего действия, то нейробиологи, которые сегодня на каждом шагу твердят, что их наука доказывает иллюзорность свободы воли, рискуют точно так же навредить целой массе людей, готовых поверить им на слово<sup>[77]</sup>. Нейробиологам, психологам и философам следует относиться к своему моральному долгу обдумывать предпосылки и следствия своих публичных заявлений о подобных вещах с той же осторожностью, которая требуется от людей, рассуждающих о глобальном потеплении или грядущих столкновениях с астероидами. В качестве примера рассмотрим сообщение, которое находит в заявлениях этих нейробиологов въедливый общественный критик и эксперт Том Вулф (2000, p. 100):

За пределами лабораторий люди приходят к следующему выводу:

*“Это заговор! Мы все запрограммированы!”* А еще: *“Не вините меня! Я запрограммирован неправильно!”*

Запрограммирован неправильно? Каково же тогда быть запрограммированным правильно? Или же ученые “открыли”, что никто не может быть правильно запрограммирован, чтобы нести моральную ответственность?

## 66. Детерминистская модель: игра “Жизнь”

Когда физику Ричарду Фейнману доводилось слушать научную лекцию по теме, в которой он не слишком разбирался, он частенько задавал лектору следующий вопрос: “Не могли бы вы привести простой пример того, о чем идет речь?” Если лектор не мог найти простого примера, это давало Фейнману повод насторожиться – и небезосновательно. Действительно ли лектору было о чем рассказать – или же он просто прикрывался сложными словами, выдавая их за научную мудрость? Если вы не можете сделать сложную задачу относительно простой, вероятно, вы неправильно к ней подходите. Упрощение полезно не только новичкам.

В биологии есть “модельные организмы” – виды, которые были тщательно отобраны, чтобы облегчить труд экспериментаторов: они быстро размножаются в лабораторных условиях, не представляют серьезной опасности и не требуют особого ухода, а также – после изучения многими научными группами – хорошо описаны и поняты. К этим видам относятся плодовая мушка, лабораторная крыса, полосатый данио, кальмар (имеющий гигантские нервные аксоны), нематода *Caenorhabditis elegans* и резуховидка Таля (*Arabidopsis thaliana*) – морозоустойчивое, быстрорастущее цветковое растение, родственное горчице, первое растение, геном которого был полностью секвенирован. В сфере искусственного интеллекта (ИИ) есть свои простые случаи, называемые “модельными задачами”: по сути, они представляют собой предельно упрощенные вариации “серьезных” задач реального мира. Многие из наиболее интересных программ, разработанных на ИИ, представляют собой решения модельных задач – например, как заставить компьютерную программу строить простые объекты в *мире кубиков*, особом виртуальном мире, где на столе лежат передвигаемые детские кубики. Игра в шахматы – тоже модельная задача, и решить ее, определенно, гораздо проще, чем приехать на машине из Мэна в Калифорнию, положить конец арабо-израильскому конфликту или даже сделать бутерброд из подходящих ингредиентов, найденных на кухне. У этиков есть свои “проблемы вагонетки”: в простейшем варианте неуправляемая вагонетка несется по рельсам и убьет пять человек, которые стоят у нее на пути, если только ее не развернуть, но вы можете перевести стрелку таким образом, чтобы вагонетка ушла на запасной путь и убила бы только одного человека. Переведете ли вы стрелку?

Вот модельный мир, помогающий людям думать о *детерминизме*, который гласит, что факты – местоположение, масса, направление и скорость каждой частицы, – наблюдаемые в один момент времени, определяют, что случится в следующий момент, и так далее до бесконечности. Физики, философы и другие ученые несколько тысяч лет спорят, является ли наша вселенная детерминистической или же некоторые события на самом деле не предопределены, то есть представляют собой непредсказуемые “случайные” события, которые просто *происходят*, хотя ничто не заставляет их произойти. Возможно, даже опытные мыслители откроют для себя что-то новое, поиграв с “Жизнью” – поразительно простой моделью детерминистического мира, созданной математиком Джоном Хортоном Конвеем и его студентами в 1970 г.

Игра “Жизнь” ведется на двумерном, размеченном на клетки поле, например на шахматной доске, с использованием простых фишек, например камешков или монеток. Можно также обратиться к высоким технологиям и играть в “Жизнь” на компьютерном экране. Цель игры состоит не в победе; если эту модель вообще можно назвать игрой, то она будет ближе всего к пасьянсу. Каждая из квадратных клеток поля в конкретный момент времени может быть либо “живой”, либо “мертвой”. (Если она “живая”, положите в нее монетку; если “мертвая”, оставьте клетку пустой.) Обратите внимание, что у каждой клетки есть восемь соседей: четыре смежные клетки – север, юг, восток и запад – и четыре диагонали – северо-восток, юго-восток, юго-запад и северо-запад.

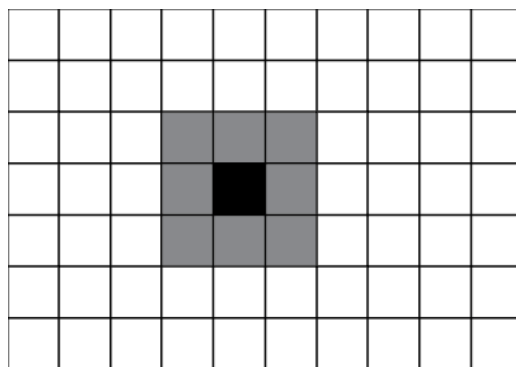


Рисунок 1

Время в мире “Жизни” дискретно, а не непрерывно – оно идет шагами, и между этими шагами конфигурация мира меняется по следующему правилу:

*Физика “Жизни”.* Посчитайте, сколько из восьми соседей каждой клетки “живо” в текущий момент. Если два, в следующий момент клетка остается в текущем состоянии (“живая” или “мертвая”). Если три, в следующий момент клетка становится “живой”, в каком бы состоянии она ни пребывала в текущий момент. При всех остальных обстоятельствах клетка “мертва”.

Это единственное правило игры. Теперь вы знаете, как играть в “Жизнь”. *Вся физика мира “Жизни” описывается этим единственным правилом, из которого нет исключений.* Хотя этот фундаментальный закон относится к “физике” мира “Жизни”, на первых порах эту любопытную физику удобно представлять в биологических терминах: считайте, что клетка становится “живой”, когда в ней зарождается жизнь, и “мертвой”, когда эта жизнь угасает, а каждый шаг времени представляет собой смену поколений. Перенаселение (более трех “живых” соседей) и изоляция (менее двух “живых” соседей) приводят к гибели. Рассмотрим несколько простых примеров.

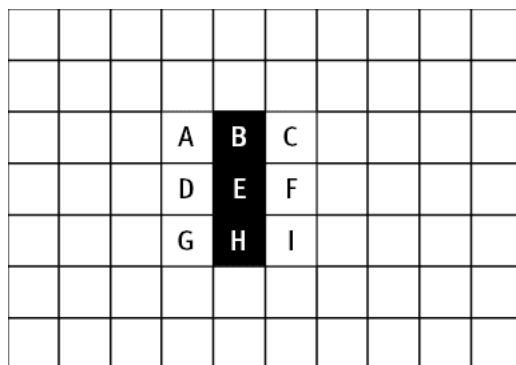


Рисунок 2

В конфигурации, изображенной на рисунке 2, ровно по три “живых” соседа есть только у клеток *D* и *F*, поэтому только с них зародится жизнь в следующем поколении. У клеток *B* и *H* по одному “живому” соседу, поэтому в следующем поколении они погибнут. У клетки *E* два “живых” соседа, поэтому она останется “живой”. Таким образом, в следующий “момент” конфигурация будет такой, как на рисунке 3.

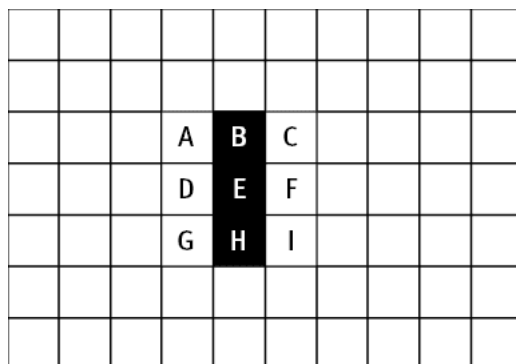


Рисунок 3

Очевидно, в следующий момент восстановится изначальная конфигурация, и это будет повторяться бесконечно, если только в игру каким-то образом не вступят новые “живые” клетки. Такая конфигурация называется *мигалкой* или *светофором*. Что произойдет с конфигурацией, изображенной на рисунке 4?

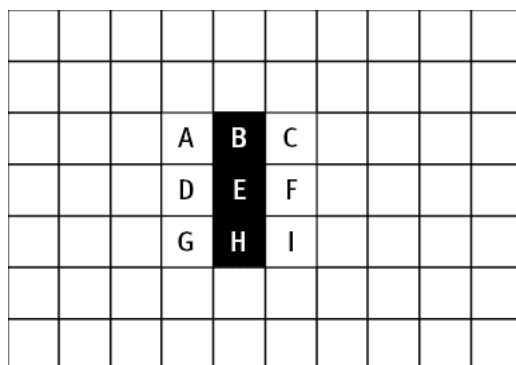


Рисунок 4

Ничего. У каждой “живой” клетки три “живых” соседа, поэтому она перерождается такой же, как была. Ни у одной “мертвой” клетки нет трех “живых” соседей, поэтому нигде не зарождается жизнь. Эта конфигурация называется *натюрморт*. Скрупулезно применяя единственный закон “Жизни”, любой может абсолютно точно предсказать, какой станет любая конфигурация “живых” и “мертвых” клеток в следующее мгновение – и в следующее за ним, и так далее. *Иными словами, мир “Жизни” представляет собой модельный мир, который прекрасно иллюстрирует детерминизм, прославленный в начале девятнадцатого века французским ученым Пьером Лапласом: имея описание состояния этого мира в некоторый момент времени, мы, наблюдатели, можем в точности предсказать его состояние во все будущие моменты времени, применяя единственный закон физики.* Можно сказать иначе: принимая физическую установку в отношении конфигурации в мире “Жизни”, мы получаем совершенные способности к предсказанию – в этом мире нет ни шума, ни неопределенности, ни вероятностей ниже единицы. Более того, двухмерность мира “Жизни” предполагает, что ничто не скрыто из виду. Нет ни закулисья, ни скрытых переменных – физика объектов в мире “Жизни” видна как на ладони.

Если вам скучно следовать простому правилу, можно прибегнуть к компьютерной модели “Жизни”, где вы генерируете изначальную конфигурацию на экране, после чего компьютер выполняет алгоритм за вас, снова и снова изменяя конфигурацию в соответствии с единственным правилом. В лучших моделях можно менять скорость течения времени и масштаб, то приближая конфигурацию, то наблюдая за ней с высоты птичьего полета.

Довольно быстро становится очевидно, что одни простые конфигурации интереснее других. Рассмотрим диагональный отрезок из клеток, изображенный на рисунке 5.

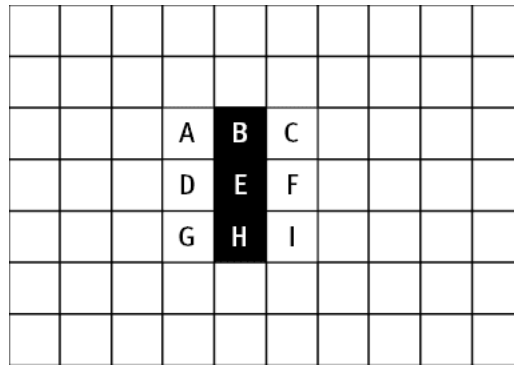


Рисунок 5

Эта конфигурация не мигалка: в каждом поколении две “живых” клетки на концах отрезка умирают в изоляции, а новая жизнь не зарождается. Вскоре весь отрезок исчезает. Помимо неизменных конфигураций – натюрмортов – и конфигураций, которые полностью исчезают, подобно диагональному отрезку, существуют конфигурации различной периодичности. Период мигалки, как мы видели, составляет два поколения, которые будут повторяться до бесконечности, если только не произойдет вторжения другой конфигурации. Вторжения и делают “Жизнь” интересной: среди периодических конфигураций есть такие, которые плывут по полю, подобно амебе. Простейшая из них – *планер*, пятিকлеточная конфигурация, пошагово продвигающаяся на юго-восток на рисунке 6.

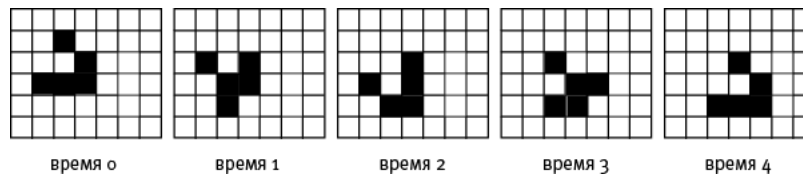


Рисунок 6

Существуют также пожиратели, паровозы и космические корабли, а также множество других метко названных обитателей мира “Жизни”, которые становятся узнаваемыми объектами на новом уровне (аналогичном *конструктивному уровню*). На этом уровне используется свой язык, на котором кратко формулируются точные описания, чересчур затянутые на физическом уровне. Например:

Пожиратель может поглотить планер за четыре поколения. Что бы ни поглощалось, принцип всегда один. Между пожирателем и его жертвой формируется мост. В следующем поколении зона моста вымирает из-за перенаселения, откусывая по куску от пожирателя и его жертвы. Затем пожиратель восстанавливается. Жертва обычно восстановиться не может. Если остаток жертвы умирает, как происходит с планером, жертва считается поглощенной. [Poundstone 1985, p. 38]

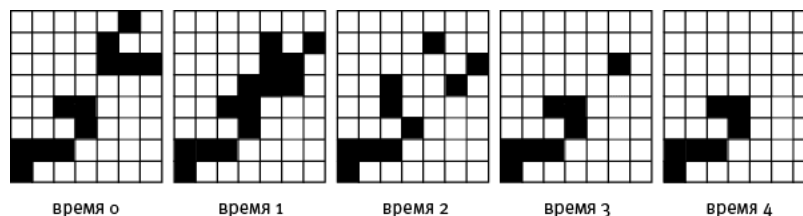


Рисунок 7

Обратите внимание, какая любопытная вещь происходит с нашей “онтологией” – нашим каталогом сущего, – по мере того как мы переходим с уровня на уровень. На физическом уровне нет движения, только “живое” и “мертвое” состояние, а единственные существующие объекты, клетки, определяются фиксированным положением в пространстве. На конструктивном уровне неожиданно появляется

движение устойчивых конфигураций: на рисунке 6 на юго-восток движется один и тот же планер (хотя в каждом поколении он состоит из разных клеток), в процессе меняющий форму, а на рисунке 7 в мире становится одним планером меньше, поскольку его поглощает пожиратель.

Также обратите внимание, что на физическом уровне нет никаких исключений из общего правила, а на конструктивном уровне наши обобщения требуют оговорок вроде “обычно” или “если не произойдет вторжения”. Осколки прошлых конфигураций могут “сломать” или “убить” один из объектов онтологии этого уровня. Их *салиентность в качестве реальных объектов* высока, но не гарантирована. Высокая салиентность предполагает, что можно без особенного риска подняться на этот уровень, принять его онтологию и предсказать – в общих чертах, с оглядкой на риски – поведение крупных конфигураций и систем конфигураций, не просчитывая при этом происходящее на физическом уровне. Например, можно поставить перед собой задачу создать интересную сверхсистему из доступных на конструктивном уровне “элементов”.

Именно этим занялись Конвей и его студенты, которые добились невероятных успехов. Они сконструировали и доказали жизнеспособность самовоспроизводящейся сущности, состоящей целиком из клеток “Жизни”. Работая в соответствии с принципом детерминизма на бесконечном поле, она идеально копировала саму себя, после чего эта копия копировала себя и так далее. Она также (вдобавок) представляла собой универсальную машину Тьюринга – двумерный компьютер, который теоретически может вычислить любую вычислимую функцию! Что же вдохновило Конвея и его студентов создать сначала этот мир, а затем и такого удивительного его обитателя? Они пытались на очень абстрактном уровне ответить на один из главных вопросов биологии: какова минимальная сложность самовоспроизводящегося организма? Они развивали блестящие ранние рассуждения Джона фон Неймана, который работал над этим вопросом до самой своей смерти в 1957 г. Фрэнсис Крик и Джеймс Уотсон открыли ДНК в 1953 г., однако принцип ее работы многие годы оставался тайной. Фон Нейман довольно подробно описал плавающего робота, который подбирает в воде обломки, чтобы сконструировать свою копию, способную впоследствии повторить тот же самый процесс. Его описание того (опубликованное посмертно, в 1966 г.), как автоматон читает собственную схему и копирует ее при создании нового робота, во многом предвосхитило целый ряд позднейших открытий о механизмах экспрессии и репликации ДНК, но чтобы сделать свое доказательство возможности существования самовоспроизводящегося автомата математически точным и понятным, фон Нейман переключился на простые, двумерные абстракции, которые теперь называют *клеточными автоматами*. “Жизнь” Конвея представляет собой особенно покладистый клеточный автомат.

Конвей и его студенты хотели подтвердить доказательство фон Неймана, создав двумерный мир с простой физикой, где подобная самовоспроизводящаяся конструкция оказалась бы стабильной, *работающей* конфигурацией. Как и фон Нейман, они хотели получить как можно более общий ответ, который был бы максимально независим от фактической (земной? местной?) физики и химии. Им хотелось, чтобы модель получилась предельно простой для расчетов и визуализации, поэтому они не только перешли от трехмерного мира к двумерному, но и “оцифровали” пространство и время: как мы видели, время и расстояние в “Жизни” всегда выражается целыми числами “мгновений” и “клеток”. Именно фон Нейман взял абстрактную концепцию механического компьютера Алана Тьюринга (теперь ее называют машиной Тьюринга) и сделал из нее универсальный серийный компьютер с хранимой в памяти программой (теперь его называют машиной фон Неймана), а также в ходе блестящих исследований пространственных и структурных требований для создания такого компьютера обнаружил – и доказал, – что универсальную машину Тьюринга (см. раздел IV) теоретически можно “сконструировать” в двумерном мире<sup>[78]</sup>. Конвей и его студенты также решили подтвердить это, поупражнявшись в двумерной инженерии<sup>[79]</sup>.

Им пришлось нелегко, но они показали, как “сконструировать” рабочий компьютер из простых форм “Жизни”. Поток планеров могут, к примеру, служить “лентой” ввода и вывода, а в качестве управляющего устройства может использоваться крупная конфигурация пожирателей, планеров и других элементов. Как выглядит эта машина? Паундстоун (1985) вычислил, что вся конструкция займет около  $10^{13}$  клеток, или пикселей.

Для отображения  $10^{13}$ -пиксельного паттерна потребуется видеозэкран шириной, по меньшей мере, около 3 миллионов пикселей... [Представьте высокочеткий экран вашего ноутбука или планшета, но шириной около километра.] Перспектива скроет из виду пиксели самовоспроизводящегося паттерна. Если вы достаточно далеко отойдете от экрана, чтобы увидеть



паттерн целиком, пиксели (и даже планеры, пожиратели и ружья) окажутся слишком маленькими, чтобы их различить. Самовоспроизводящийся паттерн будет напоминать туманность или галактику. [pp. 227–228]

Иными словами, когда достаточное количество элементов составит способную к самовоспроизводству конфигурацию (в двумерном мире), эта конфигурация будет примерно во столько же раз больше каждого из элементов, во сколько организм больше атома. Вероятно, создать способную к самовоспроизводству систему значительно меньшей сложности невозможно, однако это пока не доказано.

Игра “Жизнь” иллюстрирует многие важные принципы и может быть использована для построения множества аргументов и мысленных экспериментов, однако я с ее помощью подчеркну лишь три момента, чтобы остальное вы открыли сами. Во-первых, обратите внимание, как в ней размывается различие между физической и конструктивной установкой. Считаются ли планеры, к примеру, сконструированными или природными объектами (подобно атомам и молекулам)? Если на то пошло, управляющее устройство, составленное Конвеем и его студентами из планеров, пожирателей и тому подобных элементов, уж точно должно считаться сконструированным, но его элементы представляют собой сырье – простейшие “объекты” мира “Жизни”. Никому не пришлось конструировать или изобретать планер – он был *открыт* как подразумеваемый физикой мира “Жизни” объект. Впрочем, это относится к *любому* объекту мира “Жизни”. В этом мире не происходит ничего, что не подразумевалось бы – то есть не было бы логически выводимо посредством доказательства теорем – физикой и изначальной конфигурацией клеток. Одни объекты мира “Жизни” просто более удивительны и неожиданны (для нас, с нашим жалким интеллектом), чем другие. В определенном смысле самовоспроизводящаяся компьютерная пиксельная галактика Конвея – “лишь” еще одна макромолекула “Жизни”, для которой характерна особенно долгая и сложная периодичность. Таким образом иллюстрируется параллельный тезис о биологии и происхождении жизни: можно сказать, что аминокислоты просто есть – их не нужно было конструировать. Но белки, состоящие исключительно из аминокислот, слишком сложны – они, по крайней мере, **вроде как** сконструированы. И снова мы сталкиваемся с градуализмом Дарвина.

Во-вторых, мир “Жизни”, будучи детерминистическим, располагает в полной мере предсказуемым будущим для любой возможной конфигурации, но, как ни странно, его прошлое часто бывает совершенно непостижимо! Рассмотрим натюрморт из четырех “живых” пикселей в квадрате. При взгляде на него – и даже при взгляде на него и его соседей – невозможно сказать, каким было его прошлое. Чтобы заметить это, обратите внимание, что любые три из четырех “живых” пикселей в следующем поколении приведут к формированию натюрморта из этих четырех “живых” пикселей. Была ли хоть одна из этих клеток в прошлом “мертвой” – **инертный исторический факт**.

В-третьих, вспомните, как важен “шум” и столкновения для создания мутаций, которыми питается эволюция – и другие творческие процессы. Гигантская конфигурация Конвея воспроизводила себя, но не могла мутировать. Она всегда создавала бы свою *совершенную* копию, а чтобы включить в картину мутации, конфигурацию пришлось бы многократно увеличить. Почему? Потому что в детерминистическом мире “Жизни” “случайная” мутация может произойти, только если появится (псевдослучайным образом) фрагмент, который что-то ломает. Но самым маленьким движущимся объектом остается планер, так что можете представить его в виде отдельного фотона, или космического луча, движущегося со скоростью света (физики мира “Жизни”). Единственный планер может причинить немало вреда – если он должен только “подправить” что-то в геноме самовоспроизводящейся конфигурации, не разрушая генома, этот геном должен быть огромным в сравнении с планером и довольно крепким. Возможно, если бы выяснилось, что гигантские конфигурации размером с галактику слишком хрупки, чтобы пережить редкий дождь из планеров, нашелся бы и способ доказать, что эволюция в мире “Жизни” невозможна, какими бы большими ни были эти конфигурации.

## 67. Камень, ножницы, бумага

Вероятно, всем вам знакома игра “Камень, ножницы, бумага”. Два человека встают лицом друг к другу и считают: “Камень, ножницы, бумага – раз, два, три!” На счет три оба они одновременно показывают один из символов: либо кулак (камень), либо два пальца (ножницы), либо всю руку ладонью вниз (бумага). Камень ломает (побеждает) ножницы, ножницы режут (побеждают) бумагу, а бумага накрывает (побеждает) камень. Если оба игрока не показывают одинаковый символ – и не объявляется ничья, – один из них побеждает, а другой проигрывает. Обычно игра весьма волнующа, потому что, если вы в силах перехитрить оппонента, будет казаться, словно вы читаете его мысли, снова и снова показывая правильный символ, чтобы выиграть. Когда такое случается, всем становится не по себе. Может, одним людям эта игра дается лучше, чем другим? Похоже, что да, ведь проводятся соревнования с солидными денежными призами, определяются национальные и международные чемпионы, а лучшие игроки – и это важно, ведь в каждом турнире должен быть победитель, даже если в игре не задействуются никакие специальные навыки, – имеют на своем счету немало побед.

Как они это делают? Возможно, они обращают внимание на едва заметные перемены в мимике и жестах оппонентов. Игроки в покер утверждают, что читают других игроков, чувствуя, когда те блефуют, а когда нет, а сами при этом не забывают сохранять бесстрастное выражение лица. Возможно, большинство игроков в “камень, ножницы, бумагу” не могут контролировать свою мимику и жестикуляцию, благодаря чему лучшие игроки в последний момент и разгадывают их намерения. Какую же стратегию стоит использовать, чтобы ваш противник не получил никакой подсказки из вашей манеры держаться? Лучше всего играть случайным образом, поскольку при случайной последовательности движений ваш противник просто не может выявить никакой закономерности, ведь ее и вовсе нет. (Если вы играете случайным образом и идете ноздря в ноздю с противником, попробуйте найти закономерность в его действиях, построить на ее основании неслучайную стратегию и нанести решающий удар.)

Людам очень сложно создавать по-настоящему случайные последовательности. Как правило, они переключаются слишком часто, например, стараясь не выбирать один и тот же символ два-три раза подряд (что в поистине случайных последовательностях должно случаться довольно часто). Поскольку вы знаете, что попытка создать действительно лишенную закономерностей последовательность будет обречена на провал, вам стоит рассмотреть другую стратегию: найдите в библиотеке (или в интернете) таблицу случайных чисел. “Случайным” образом поставьте палец в одну из ячеек таблицы и скопируйте следующие сто чисел. Выбросите все нули (скажем), замените все единицы, двойки и тройки буквой “К” (камень), все четверки, пятерки и шестерки – буквой “Б” (бумага), а все семерки, восьмерки и девятки – буквой “Н” (ножницы). Так у вас получится последовательность примерно из девяноста символов (поскольку вы выбросили около десяти нулей), которой должно хватить на одну партию.

Теперь, когда вы подготовились к игре, пора узнать главное правило: *никому не показывайте свой список*. Если ваш противник сумеет в него заглянуть, вы окажетесь в его власти. Как говорится, он станет выкачивать из вас деньги. Если же он не сумеет заглянуть в ваш список, ему придется попотеть, чтобы разгадать вашу стратегию. (Иными словами, ему придется оценивать вас из **интенциональной установки**, рассуждая о ходе ваших мыслей, вместо того чтобы рассматривать вас в качестве простого механизма, поведение которого полностью предсказуемо на основании списка.)

Простой принцип скрывать свои намерения от противника оказывается одним из главных поворотных моментов в многолетних спорах о свободе воли. Фактически изобретение фон Нейманом и Моргенштерном (1944) *теории игр* началось с осознания, что при попытке одиночного агента (или интенциональной системы) предсказать будущее на основании собираемой информации ему достаточно рассчитать *ожидаемую полезность*, используя теорию вероятности (того или иного типа), но при появлении второго агента – второй интенциональной системы в среде – обстоятельства меняются радикальным образом. Теперь каждому агенту приходится учитывать попытки предсказания другого агента, включая наблюдения другого агента о его собственном поведении и попытки предугадать и использовать это поведение, что приводит к созданию петель обратной связи неограниченной сложности<sup>[80]</sup>.

Непроницаемый туман непостижимости и непредсказуемости поведения агентов другими агентами создает условия для процветания теории игр. Эволюции это известно, и принципы теории игр в своих взаимодействиях применяют многие виды (**компетентность без понимания!**). **Смотровые прыжки**

газелей – лишь один простой пример. Другим примером может служить беспорядочный полет бабочек, траектории движения которых трудно предсказать насекомоядным птицам, хотя эволюция и помогла птицам, наделив их более высокой “частотой слияния мельканий”, чем, например, наша (они видят больше “кадров в секунду”, чем мы, поэтому обычный фильм показался бы им слайд-шоу). Фактически “понимание” этого принципа наблюдается и в типичном инстинкте животных, склонных считать любой сложный движущийся объект агентом – “Кто здесь, и что вам надо?”, а не “Что это?”, – чтобы не подвергать себя опасности, на случай если это действительно агент, который хочет съесть животное, спариться с ним или вступить в схватку за что-то. (Я утверждаю [2006а], что этот инстинктивный ответ объясняет появление в процессе эволюции всех невидимых эльфов, гоблинов, лепреконов, фей, огров и богов, которые в итоге выливаются в единого Бога, непревзойденную невидимую интенциональную систему.)

Подобно животным, которые используют эту стратегию, хотя им и не нужно понимать почему, мы, люди, по достоинству ценим тактику сохранения своей непредсказуемости, хотя и не имеем необходимости понимать, почему это важно. Но обычно это вполне очевидно. Заметив в антикварном магазине симпатичную вещицу, вы не выражаете своего восхищения, пока не узнаете ее цену. Если вы поступите по-другому, продавец возьмет с вас втридорога. Выставляя что-либо на продажу, вы называете цену, которая кажется вам разумной, потому что не можете предугадать, готов ли покупатель заплатить больше, и надеетесь, что он не поймет, на сколько вы готовы ее снизить. (Само собой, отказаться от торга невозможно.) Аукционы дают прекрасный шанс исследовать эту неизведанную область – заранее называя аукционисту свою цену, вы полагаетесь на его порядочность и надеетесь, что он не раскроет ваш максимум другим потенциальным покупателям.

Точно так же, без памяти влюбляясь в кого-то с первого взгляда, вы стараетесь не терять голову и не выражать своего восторга, оставаясь как можно более сдержанным и бесстрастным, ведь вам вовсе не хочется отпугнуть такого удивительного человека или наоборот дать ему преимущество и сразу позволить вить из вас веревки. Бесстрастность полезна не только в покере. Как правило, вы повышаете свои шансы добиться желаемого, держа конкурентов – других агентов в окрестностях – в подвешенном состоянии и не позволяя им предугадать, какой шаг вы сделаете, и подготовиться соответствующим образом. (Подготовка среды другого агента весьма затратна, поэтому противники не будут пытаться предугадать ваши действия, не имея достаточных оснований для этого.)

Фокусники умело пользуются “психологической силой”, чтобы заставлять вас (“на ваше усмотрение”) вытаскивать из колоды именно ту карту, которая им нужна. Способов много, а распознать их очень сложно, поэтому хороший фокусник почти всегда справляется с задачей. Таким образом он ограничивает вашу субъектность, манипулирует вами, превращая вас в инструмент, в пешку, в исполнителя воли фокусника и лишая вас статуса свободного агента.

Вопреки античной идеологии, мы не хотим, чтобы наша свобода выбора ничем не обуславливалась. Мы хотим – и должны хотеть – действовать, имея в своем распоряжении достаточное количество информации о доступных нам вариантах. Если бы только среда *обуславливала* наличие у нас множества необходимых истинных убеждений об окружающем мире, а также *подталкивала* нас выбирать наиболее разумный курс действий на основании этой информации! Так мы получили бы *почти* все, чего хотим в качестве агентов, за исключением одного: нам не хотелось бы, чтобы в среде присутствовал агент-манипулятор, который забирает бы у нас контроль, поэтому мы не хотели бы, чтобы среда делала наши лучшие действия слишком очевидными для других агентов, поскольку в таком случае они стали бы эксплуатировать нас, зная слишком много о том, что мы хотим и насколько сильно. Итак, к нашим желаниям нужно добавить способность хранить в тайне ход мыслительного процесса и принимаемые решения, даже если для этого нам придется время от времени выбирать второй по качеству вариант, просто чтобы держать остальных в неведении. (Первый формальный анализ этого см. Clegg 2012.)

Некоторые люди, смутно осознавая важность этой непредсказуемости, полагают, что “на всякий случай” им стоит стремиться к *абсолютной* непредсказуемости, достичь которой можно, только если в основе нашего мозга лежит физическая индетерминистичность. Вот как однажды с характерной живостью описал это философ Джерри Фодор (2003):

Человеку хочется быть таким, какой традиция рисует Еву в тот момент, когда она кусает яблоко. Совершенно свободной поступить иначе. Настолько свободной, что даже Бог был не в силах сказать, какой шаг она сделает. [p. 18]

Но почему “человек хочет” этого? Разве такая абсолютная непредсказуемость лучше практической непредсказуемости? Многие философы на протяжении нескольких тысяч лет утверждали, что без абсолютной непредсказуемости невозможна истинная свобода воли. Может, они знают что-то, чего не знаем мы? Если и так, они не спешат раскрывать свои карты. Для большинства из них мысль о том, что свобода воли несовместима с детерминизмом, кажется слишком очевидной, чтобы ее подкреплять. Я же считаю, что они обязаны доказать свою точку зрения. Объяснить нам, почему мы должны страдать, если нам не дана *абсолютная* непредсказуемость. Я показал, почему нам – как и эволюции – выгодно как можно большая *практическая* непредсказуемость. Прошу вас, объясните, почему ее недостаточно.

Я приведу одну причину: если вы планируете играть в “камень, ножницы, бумагу” с Богом, а ставки высоки (например, спасение души), тогда у вас действительно, как замечает Фодор, есть основание желать “совершенной” свободы. Лично я такой игры не ожидаю, поэтому мне достаточно и практической свободы, которую я получаю, когда держу язык за зубами и сторонюсь мошенников и других манипуляторов, если на карту поставлено многое.

## 68. Две лотереи

Сравним две следующих лотереи на справедливость. В лотерее П – от слова “после” – осуществляется продажа билетов, их корешки кладутся в подходящий барабан и перемешиваются как можно более случайным образом, *после* чего вслепую выбирается выигрышный билет. (По такому принципу проводится большинство знакомых нам лотерей.) В лотерее Д – от слова “до” – перемешивание корешков и случайный выбор победителя осуществляется *до* продажи билетов (и выигрышный корешок кладется в сейф), но в остальном лотереи ничем друг от друга не отличаются. Кому-то может показаться, что вторая лотерея несправедлива, поскольку выигрышный билет определяется до того, как люди купят билеты. Один из этих билетов *уже* выиграл (даже если никто не знает, какой именно), а остальные билеты представляют собой ничемные куски бумаги, и продавать их ничего не подозревающим людям – значит, в некотором роде жульничать. Но на самом деле обе лотереи в равной степени справедливы. Все покупатели билетов имеют равные шансы выиграть, и совершенно неважно, в какой момент осуществляется выбор победителя.

В большинстве лотерей победитель определяется после продажи билетов, чтобы люди своими глазами видели, что их не обманывают. Никакой мошенник, имеющий доступ к внутренней информации, не манипулировал распространением билетов, потому что ни один агент не знал (и не мог знать), какой билет победит, прежде чем билеты были проданы. Любопытно, что не все лотереи работают по этому принципу. Компания *Publishers Clearing House* каждый год рассылала миллионы конвертов, на которых было жирным шрифтом напечатано “ВОЗМОЖНО, ВЫ УЖЕ ВЫИГРАЛИ” – миллион долларов или какой-нибудь другой приз. (Теперь эта организация проводит лотерею онлайн.) Такие дорогие кампании основаны на маркетинговом исследовании, которое показывает, что люди, как правило, считают лотереи с заранее определенными победителями справедливыми, если они проводятся честно. Однако, возможно, люди не жалуются на такие лотереи, потому что получают билеты бесплатно. Много ли людей *купило бы* билеты лотереи, выигрышный корешок которой с самого начала поместили бы в специальном конверте на хранение в банковскую ячейку? Люди миллионами покупают билеты моментальной лотереи, хотя в момент покупки уже определено, выигрывает ли каждый конкретный билет. Очевидно, эти люди считают, что у них есть реальная возможность выиграть. На мой взгляд, они правы, но, даже если это не так, их спокойная уверенность в справедливости подобных лотерей и в реальной возможности выиграть должна опровергать уверенность философов (восходящую к Демокриту и Лукрецию, которые жили две тысячи лет назад), которые почему-то считают, что возможность нельзя признать *реальной*, если только исход не остается неопределенным до последней секунды. Эти философы утверждают, что в отсутствие постоянного притока поистине случайных, неопределенных *точек разветвления*, разрывающих ткань каузальности, нет ни свободы выбора, ни *реальной* возможности поступать правильно.

Две лотереи помогают нам по-новому взглянуть на проблему детерминизма. Если мир предопределен, то в нас стоят генераторы псевдослучайных чисел, а не истинные (квантовомеханические) рандомизаторы. Если наш мир предопределен, то все наши лотерейные билеты, по сути, были разыграны около четырнадцати миллиардов лет назад, в момент Большого взрыва, после чего их положили в конверт и стали раздавать нам по мере необходимости. Всякий раз, когда вы “подбрасываете монетку” или принимаете рискованное решение менее претенциозным образом, ваш мозг открывает конверт и вытаскивает оттуда следующее “случайное” число, позволяя ему *определять*, что вы сделаете, подобно тому как список жестов определяет вашу стратегию при игре в “камень, ножницы, бумагу”. “Но это несправедливо! – воскликнет кто-нибудь. – Ведь одни будут выигрывать чаще других”. И правда, в конкретном случае у некоторых людей шансы выиграть выше, но не стоит забывать, что в долгосрочной перспективе удача усредняется. “Но если все определяется до нашего рождения, одни люди *обречены* быть удачливее других!” Так будет и в том случае, если лотерея будет проводиться не до нашего рождения, а периодически, по требованию, в течение всей нашей жизни. Даже в совершенно случайной и беспристрастной лотерее, в поистине непредопределенной лотерее, одни люди все равно обречены выигрывать чаще других. Даже в совершенно справедливом, совершенно случайном турнире по подбрасыванию монетки кто-то *обречен* победить, в то время как все остальные участники проигрывают. Победитель не может с полным правом сказать, что победа была предначертана ему “судьбой”, и все же побеждает именно он – а разве что-то может быть справедливее? Справедливость не предполагает, что победителями становятся все.

Вероятно, чаще всего надежда на неопределенность объясняется тем, что в отсутствие этой неопределенности, если мы решили действовать, “мы не могли поступить иначе”, а это, **безусловно** (динь!), не может нас не волновать. Это тоже не так очевидно, как часто кажется, а потому, чтобы понять, как нас сбивает с толку эта знакомая идея, давайте рассмотрим любопытную категорию инертных исторических фактов.

## 69. Инертные исторические факты

Инертный исторический факт – это любой факт о совершенно заурядном положении вещей в мире в некоторый момент прошлого, более не поддающийся восприятию и *вообще* не оставивший следа в современном мире. Вот мой любимый инертный исторический факт:

А. Часть золота моих зубов когда-то принадлежала Юлию Цезарю.

Вот еще один:

Б. Утверждение о том, что часть золота моих зубов когда-то принадлежала Юлию Цезарю, неверно.

Итак (согласно логике), фактом можно назвать лишь одно из этих утверждений. (Погодите-ка, разве в главе об операторе “вроде как” не говорилось, что нам не стоит доверять таким “очевидным” дизайнкам? Давайте проведем проверку. В каком случае ни утверждение А, ни утверждение Б не будет истинным? Что, если во времена Цезаря *право собственности* было сформулировано нечетко, а потому часть золота Цезаря принадлежала ему лишь вроде как – например, как королеве Англии до сих пор принадлежат все лебеди в стране?<sup>[81]</sup> Если понятие собственности определено четко, как понятие золота, одно из двух утверждений должно быть фактом, но понять, какое из них истинно посредством физического исследования почти наверняка невозможно, каким бы сложным ни было исследование и сколько бы времени на него ни ушло.

Правда? Можно представить случаи, в которых мы могли бы почти наверняка сказать, какое из утверждений верно. Если бы выяснилось, что ряд хорошо задокументированных исторических процессов позволяет проследить “провенанс” части золота моих зубов, который скрупулезно контролировался и записывался на протяжении тысячелетий (как “цепь сохранности” вещественных доказательств в делах об убийствах), мы могли бы быть вполне уверены в истинности утверждения А. Скажем, мой стоматолог купил в музее золотое кольцо, знаменитое “кольцо с мизинца Цезаря”, которое, как свидетельствует множество документов, веками передавалось от монарха к монарху, пока не оказалось в музее, а затем записал на видео процесс переплавки этого кольца и изготовления коронки с помощью гипсовой формы. Да, это маловероятно, но явно не выходит за рамки физически возможного. Или же допустим, что я на досуге мою золото и съездил на Аляску к основанию тающего ледника, покрывавшего землю на протяжении десяти тысяч лет, и по крупницам собрал там все золото, которое впоследствии пошло на изготовление моих коронок. В таком случае истинным даже с большей вероятностью будет утверждение Б. Но если ни один из этих крайних случаев не подкреплён надёжными свидетельствами, мы почти наверняка так и не сможем установить, какое из утверждений истинно. Истинное утверждение – какое бы из двух им ни было – представляет собой инертный исторический факт.

Квантовая физика приводит к возникновению любопытной проблемы: даже если бы существовала возможность проследить многовековую траекторию каждого атома золота из моих коронок, если за это время два или более атома золота – один из которых происходил из наследия Цезаря, а другой из внешних источников – сталкивались (или хотя бы оказывалось *очень* близко друг к другу), определить, какой из атомов какой, после такого “столкновения” было бы в принципе невозможно. Атомы и более мелкие частицы не имеют никаких опознавательных признаков вроде отпечатков пальцев и не могут отслеживаться *непрерывно*, поэтому не всегда можно говорить о постоянной тождественности частиц, и это становится очередным препятствием к установлению фактов о золоте.

Будь наша вселенная хоть детерминистической, хоть нет, компьютеры созданы быть детерминистическими, несмотря на субмикроскопический шум и даже квантовую случайность, и нейтрализуют эти флуктуации, будучи цифровыми, а не аналоговыми. (Прекрасным примером этого служит описанная в главе 66 **игра “Жизнь” Конвея**, но цифровой детерминизм повсеместен.) Фундаментальная идея цифровизации с целью детерминизма заключается в том, что мы можем специально *создавать* инертные исторические факты. Принудительная разбивка всех ключевых событий на две категории – высоко- и низковольтные, ВКЛ и ВЫКЛ, 0 и 1 – гарантирует, что микроразличия (между разными значениями высокого напряжения, разными типами ВКЛ, разными вариантами 0) безжалостно отбрасываются. От них ничто не зависит, поэтому они исчезают без следа, как факты о подлинных исторических вариациях, *никак не влияющих* на последующие цепочки состояний

компьютера. К примеру, ваш друг скачивает песню с сайта и записывает ее на два CD-диска. Два этих диска – мы назовем их А и Б, но не будем подписывать, – само собой, представляют собой цифровые дубликаты. Попросите своего друга “скопировать” один из них на ваш ноутбук в совершенно темной комнате. Не уточняйте, какой из дисков использовать, а затем попросите его подержать оба диска в руках (чтобы невозможно было установить, какой из них использовался, по отпечаткам его пальцев или следам ДНК), бросить их в мешок с другими дисками и хорошенько его потрясти. Теперь у нас есть два возможных инертных исторических факта:

- (а) На вашем ноутбуке хранится копия с диска А.
- (б) На вашем ноутбуке хранится копия с диска Б.

На микроскопическом уровне структура любой физической кодировки двоичного потока песни отличается от структуры любой другой кодировки. Если “скопировать” один из дисков в оперативную память (ОЗУ) компьютера, картина распределения напряжения в ОЗУ также будет обладать уникальной структурой, а если затем “скопировать” файл из ОЗУ на жесткий диск или на флешку, для новых копий также будут характерны микроскопические отличия. Для удобства этот процесс называется копированием, но на самом деле он представляет собой создание нового непрерывного или аналогового физического сигнала, обладающего собственной уникальной структурой, поскольку именно так устроен мир на уровне электронов и протонов. Но гений цифровизации в том, что вся эта мелкая структура “игнорируется”, стирается в процессе “корректировки к норме”. Когда существует норма наподобие алфавита, неважно, как иМЕННо выглядят отдельные символы, ведь все они читаются одинаково. Для компьютера это по-прежнему нули и единицы.

Таким образом, если только на одном из дисков не содержится “ошибки”, различимой на уровне цифровой идеализации (то есть *переключенного бита* – нуля на месте единицы или наоборот), определить, с какого диска произошло копирование, будет невозможно. Цифровизация не позволяет индивидуальности двух дисков *распространяться* на последующие версии и проявляться при конверсии цифрового сигнала в аналоговый, осуществляемой колонками или наушниками. Некоторые любители музыки могут похвастаться “абсолютным слухом” и по звучанию отличают виниловые пластинки от лучших цифровых дисков, а сжатые цифровые музыкальные форматы (например, MPEG) от несжатых, но ни один меломан не скажет вам, с одного или с разных дисков проиграла две записи одной и той же музыки. Это инертный исторический факт, который не поддается определению не только на слух, но и при электронной микроскопии структуры “копий” этих дисков в оперативной памяти. Если бы нашелся человек, способный различать такие “копии”, он стал бы первым кандидатом на обладание сильнейшим экстрасенсорным восприятием, поскольку цифровизация создает физический барьер, предотвращающий передачу информации, которая могла бы дать подсказку гадающему.

Компьютер – это цифровое устройство, а потому не составляет труда заставить его выполнить несколько триллионов шагов, затем привести его обратно в *точно* то же самое (цифровое) состояние, в котором он пребывал до этого, и наблюдать, как он исполняет *точно* те же самые несколько триллионов (цифровых) шагов снова и снова.

Погодите-ка! Вы утверждаете, что компьютеры детерминистичны? Что их можно заставить повторять один и тот же триллион шагов снова и снова? Да ладно! Почему же тогда мой ноутбук то и дело меня подводит? Почему мой текстовый процессор завис во вторник, хотя я делал то же самое, что прекрасно работало в понедельник?

Но вы делали *не совсем* то же самое. Он завис не потому, что он индетерминистичен, а потому, что во вторник он пребывал не в *точно* том же состоянии, что и в понедельник. Должно быть, в промежутке ваш ноутбук сделал что-то, что подняло скрытый “флаг” или задействовало ту часть текстового процессора, которую вы никогда прежде не активировали, и это привело к переключению какого-то бита, после чего он сохранился в новом положении при выключении компьютера, а теперь текстовый процессор споткнулся об это крошечное изменение и завис. И если вы каким-то образом сумеете вернуть его в *точно* то же самое состояние, в котором он пребывал во вторник утром, он зависнет снова.

Что насчет “генератора случайных чисел”? Я думал, в мой компьютер встроено устройство для создания хаотичности по требованию.



В каждый компьютер сегодня встроен генератор “случайных чисел”, к которому при необходимости может обращаться любая работающая на компьютере программа. (До появления компьютеров люди покупали для своих исследований книги, все страницы которых были заполнены расставленными в случайном порядке числами, сгенерированная последовательность которых проходила все проверки на случайность, разработанные математиками. Само собой, каждый экземпляр конкретного издания такой книги содержал одну и ту же случайную последовательность. Одну из лучших книг в 1955 г. опубликовала организация *RAND Corporation* – в этой книге содержался миллион случайных цифр.) Последовательность чисел, генерируемых так называемым генератором случайных чисел, на самом деле не случайна, а лишь псевдослучайна: она “математически сокращаема” в том смысле, что эту бесконечно длинную последовательность можно заключить в конечно определенный механизм, который поставит ее генерацию на поток. Допустим, например, что ваш генератор случайных чисел – это программа, описание которой уложится, скажем, в один мегабайт (восемь миллионов бит), но генерируемая ею последовательность (всегда одна и та же) на самом деле бесконечна. Если бы вы захотели переслать кому-то эту бесконечную последовательность, вам бы не пришлось для этого отправлять бесконечно длинное электронное письмо, записывая последовательность *буквально*. Вы могли бы просто переслать свой алгоритм размером в мегабайт, и ваш визави получил бы доступ ко всей бесконечной последовательности. В этом суть генератора псевдослучайных чисел. При каждом перезапуске генератора случайных чисел – например, при каждой перезагрузке компьютера – он всегда начинает выдавать *точно* такую же последовательность цифр, но эта последовательность столь же *очевидно* лишена закономерностей, как если бы ее генерировали подлинно случайные квантовые флуктуации. Можно сказать, что этот генератор – встроенная “таблица случайных чисел”, подобная той, что вы получите на очень длинной петле видеопленки, записывая результаты миллионов вращений рулетки при честной игре. Петля всегда возвращается “к началу” – к моменту запуска компьютера. Иногда это важно. Компьютерные программы, обращающиеся к генератору случайных чисел в различных точках “выбора”, все равно будут проходить одну и ту же последовательность состояний, если их снова и снова будут перезапускать, а если вам захочется проверить программу на баги, вы всегда будете тестировать одну и ту же “случайную выборку” состояний, если только не предпримете необходимые шаги (достаточно простые), чтобы заставить программу время от времени выбирать следующее “случайное” число из нового фрагмента потока цифр.

## 70. Компьютерный шахматный марафон

Ясно мыслить о детерминизме и свободе выбора чрезвычайно сложно. Если детерминизм реален, существует ли вообще *реальная* свобода выбора? Если агент, якобы обладающий свободой воли, на самом деле детерминистичен и живет в детерминистическом мире, исключает ли это свободу выбора и *случай*? Далее описывается насос интуиции, который помогает изучить этот вопрос, взглянув на упрощенный мир – игру в шахматы – в искусственно сконструированном детерминистическом мире: мире работы компьютера.

Допустим, вы установили на свой компьютер две разные шахматные программы и связали их с небольшой управляющей программой, которая заставляет их играть друг с другом партию за партией в потенциально бесконечной серии. Станут ли они играть одну и ту же партию снова и снова, пока вы не выключите компьютер? Вы *могли бы* настроить программы А и Б таким образом, но в таком случае не узнали бы о них ничего интересного. Допустим, А побеждает Б в этой часто повторяемой партии. На основании этого нельзя сказать, что программа А в целом лучше программы Б или что программа А победит программу Б в другой партии, а повторение одной и той же партии не позволило бы вам узнать ничего о сильных и слабых сторонах каждой программы. Гораздо более информативным стал бы турнир, в котором программы А и Б сыграли бы последовательность разных партий. Организовать его несложно. Если любая из шахматных программ в ходе расчетов обращается к генератору случайных чисел (если, например, она периодически “подбрасывает монетку”, чтобы найти выход из ситуаций, где нет очевидного основания для выбора того или иного действия в процессе поиска удачного хода), то в следующей партии состояние генератора случайных чисел будет иным (если только вы не настроили его перезагрузку), а следовательно, будут рассматриваться другие варианты в другом порядке, что время от времени будет приводить к “выбору” других ходов. В результате вторая партия сложится по-другому, а третья будет отличаться иным образом – и в итоге все партии получившейся серии окажутся уникальны, как снежинки. Тем не менее, если вы выключите компьютер, а затем перезапустите его с той же программой, на ваших глазах развернется точно та же серия не похожих друг на друга партий, поскольку та же самая псевдослучайная последовательность чисел будет определять результаты “подбрасывания монетки” обеими шахматными программами.

Допустим, мы создали такую шахматную вселенную с двумя программами, А и Б, и изучаем результаты серии из, скажем, тысячи партий. Мы найдем множество заслуживающих доверия закономерностей. Допустим, мы установим, что программа А всегда побеждает программу Б в тысяче *разных* партий. Нам захочется объяснить эту закономерность, а фраза “поскольку программа детерминистична, программа А *обречена* всегда побеждать программу Б” не удовлетворяет нашего небезосновательного любопытства. Нам захочется понять, что в структуре, методах и диспозициях программы А объясняет ее превосходство в шахматах. Программа А имеет компетенцию или способность, которой не хватает программе Б, и нам необходимо изолировать этот любопытный фактор. Возможно, объяснение следует искать на низком уровне: может выясниться, например, что программы А и Б на самом деле одинаковы и на уровне исходного кода представляют собой *идентичные* оценщики шахматных ходов, но программа А скомпилирована эффективнее программы Б, а потому может лучше анализировать партию за одинаковое количество машинных циклов. Фактически программа А “думает [о шахматах] ровно то же самое”, что и программа Б, а программа Б “знает” о шахматах все, что известно программе А, но программа А просто думает быстрее. (На серьезном уровне, на шахматных турнирах, игра всегда ведется с использованием часов: не успев сделать все ходы за отведенное время, вы проиграете.) Более вероятно, что превосходство программы А над программой Б потребует объяснения на более высоком уровне – на уровне типичных элементов, которые задействуются в процессе принятия решений в шахматах: *представлений* шахматных позиций, *оценок* возможных продолжений, *решений* о выборе конкретных продолжений и так далее. Таким образом, программа А может корректировать относительную ценность фигур по ходу партии, лучше оценивать шахматные позиции либо раньше или позже прерывать определенные изыскания. Она не “думает то же самое”, что и программа Б, она “думает лучше, а ее мысли сложнее”. (Само собой, она лишь **вроде как** думает. В отличие от человека она не наделена сознанием.)

Пожалуй, случай будет показательнее, если одна из программ не будет побеждать всегда. Допустим, программа А *почти* всегда побеждает программу Б и при этом оценивает ходы, используя другой набор принципов. В таком случае закономерность окажется еще любопытнее. Чтобы исследовать *этот*

каузальный вопрос, нам нужно будет изучить историю тысячи разных игр, выявляя прочие закономерности. Без сомнения, их найдется много. Некоторые из них будут характерны для шахмат в принципе (например, высокая вероятность поражения программы Б в любой партии, где программе Б недостает одной ладьи), а некоторые будут специфичны для программ А и Б как конкретных шахматистов (например, склонность Б слишком рано вводить в игру ферзя). Мы выявим также стандартные закономерности шахматной стратегии, например тот факт, что, когда время программы Б истекает, она начинает исследовать оставшиеся узлы дерева игры не так внимательно, как исследует их в такой же позиции при наличии большего количества времени. Иными словами, мы обнаружим множество *пояснительных* закономерностей – как не допускающих исключений (в нашей серии из тысячи игр), так и статистических.

Такие различные шахматные закономерности сразу бросаются в глаза на фоне торжества детерминизма, которое с позиции микрокаузальности представляется весьма монотонным. Со своей позиции мы наблюдаем за напряженным поединком двух шахматных программ, но при взгляде через “микроскоп” (при наблюдении за инструкциями и данными, проходящими через ЦПУ компьютера) увидим единственный детерминистический автоматон, работающий единственным доступным ему способом. Предугадать его шаги не составляет труда – достаточно изучить, в каком именно состоянии пребывает его генератор псевдослучайных чисел, а также проанализировать структуру программы и данные. В его будущем нет “настоящих” разветвлений и развилок: все “выборы”, совершаемые программами А и Б, заранее предопределены состоянием компьютера и его памяти. Казалось бы, в таком мире по-настоящему *возможным* нельзя считать ничего, за исключением того, что на самом деле происходит. Допустим, например, что над программой А в момент времени *t* нависает зловещая *матовая сеть* (гарантированная победа, распознать которую порой непросто), но все меняется, когда у программы Б заканчивается время и она прерывает свой поиск важного хода чуть раньше, чем необходимо. Получается, что эта матовая сеть *не должна была возникнуть*. (Если у нас возникают сомнения, мы можем доказать это, проведя точно такой же турнир в другой день. На том же этапе серии у программы Б снова закончится время, из-за чего ей снова придется прервать свои поиски ровно в тот же момент.)

Что же это значит? Неужели в этом модельном мире ничего нельзя предотвратить и ничего нельзя избежать? Неужели в нем нет ни нападений, ни защит, ни упущенных возможностей, ни пикировок истинной субъектности, ни подлинных возможностей? Следует признать, что наши шахматные программы, подобно насекомым и рыбам, слишком просты в качестве агентов, чтобы быть подходящими кандидатами на обладание морально значимой свободой воли, но детерминизм их мира не лишает их разноплановых талантов и *способностей* пользоваться имеющимися *возможностями*. Если мы хотим понять, что происходит в таком мире, мы можем – и должны – обсудить, как их осознанный *выбор* приводит к изменению обстоятельств, а также что они *могут* и *не могут* делать. Если мы хотим выявить *каузальные закономерности*, чтобы объяснить повторяющиеся мотивы, замеченные нами в тысяче партий, нам нужно всерьез рассмотреть позицию, в соответствии с которой в этом мире действуют два агента, А и Б, пытающиеся обыграть друг друга в шахматы.

Допустим, мы настроим турнирную программу таким образом, что при каждой победе А будет звенеть колокольчик, а при каждой победе Б – раздаваться гудок. Когда мы запустим марафон, наблюдатель, не знающий ничего о программе, заметит, что колокольчик звенит довольно часто, а гудок не раздается почти никогда. Наблюдателю захочется узнать, чем объясняется эта закономерность. Закономерность побед А над Б можно выявить и *описать*, не принимая **интенциональную установку**, но она все равно потребует объяснения. Единственным – и верным – объяснением может быть тот факт, что генерируемые А “убеждения” о том, что сделает Б, если... оказываются лучше, чем генерируемые Б “убеждения” о том, что сделает А, если... В таком случае, чтобы найти объяснение, *необходимо* принять интенциональную установку (см. главы 33 и 42, где приводятся примеры других каузальных связей, которые не поддаются объяснению, пока вы не принимаете интенциональную установку).

Пока все идет хорошо, но эти “решения” и “выборы” кажутся лишь **вроде как** решениями и выборами. Такое впечатление, что им недостает чего-то, что свойственно подлинной свободе выбора: возможности поступить иначе. Но давайте внимательнее изучим конкретный пример, ведь внешность бывает обманчива. Для этого добавим в нашу турнирную программу третью шахматную программу, программу В. Допустим, программа В лучше программ А и Б и побеждает их почти всегда. Также допустим, что первые двенадцать ходов в паре таких партий в точности повторяются и программа В выигрывает обе партии, побеждая и программу А, и программу Б, но после двенадцатого хода партии идут не совсем одинаково. По завершении партий эксперты приходят к выводу, что программа В с большой вероятностью проиграла бы, если бы 12-м ходом, последним общим ходом обеих партий,

программа А или программа Б провела бы рокировку. Рокировка 12-м ходом давала ключ к победе, который не заметила ни программа А, ни программа Б.

Пожимая плечами, разработчик программы А говорит: “Программа А могла бы провести рокировку”. Разработчик программы Б добавляет: “Моя программа, программа Б, тоже могла бы провести рокировку”. Но разработчик программы А прав, а разработчик программы Б ошибается! Как такое может быть? Турнирная программа Т детерминистична, и если мы снова сыграем те же партии в *точно* том же самом состоянии, ни программа А, ни программа Б не проведет рокировку! Разве разработчик программы А не заблуждается? Не обязательно. Что мы пытаемся выяснить, когда спрашиваем, могла ли программа А поступить иначе? Снова и снова рассматривая в *точности* такой же случай, мы не получим никакой информации, но рассматривая *похожие* случаи, мы увидим более полную картину. Если мы выясним, что во многих подобных обстоятельствах в других партиях программа А все же *продолжает* процесс оценки, замечает плюсы подобных ходов и совершает их, это подтвердит убежденность разработчика, что программа А могла бы провести рокировку.

Как минимум, мы можем обнаружить, что переключение единственного бита в генераторе (псевдо)случайных чисел заставило бы программу А провести рокировку. Допустим, разработчик программы А проанализирует фактическое исполнение программы и выяснит, что в этом случае программа А перестала “думать” на мгновение раньше нужного. (Любая шахматная программа, какой бы прекрасной она ни была, вынуждена в какой-то момент принудительно прерывать свои поиски.) Программа А рассматривала возможность рокировки и начала анализировать ее исход, но время поджимало, а потому программа А обратилась к своему генератору случайных чисел, то есть, по сути, подбросила монетку, и выбрала ход, который сочла лучшим на тот момент – и этим ходом стала не рокировка. Но если бы псевдослучайным числом стала единица, а не ноль, программа А подумала бы над своим ходом немного дольше и в итоге провела бы рокировку. “Просто переключите один бит в случайном числе – и программа А победит!” – заявляет разработчик. Мы же скажем, что в таком случае просчет программы А с рокировкой оказался непредвиденным результатом неудачного обращения к генератору случайных чисел.

Когда мы обратимся к разработчику программы Б, он не сможет подобным образом подтвердить свое заявление, что программа Б могла бы провести рокировку. Программа Б действительно “знает”, что рокировка в сложившейся ситуации допустима, и, возможно, даже некоторое время “рассматривала” возможность рокировки, но выбрать рокировку в этом случае даже не собиралась. Рокировка была для нее *сложным* ходом – одним из тех ходов, которые в газетных шахматных задачах помечают символом “!”, – и выходила далеко за пределы ограниченных аналитических способностей программы Б. Итак, у нас есть полностью детерминистический мир – программа Т, – в котором программа А могла провести рокировку, а программа Б провести ее не могла. Разница между программами А и Б реальна и объяснима – это разница в компетентности или способностях. Можно сформулировать это очевидно парадоксальным образом:

Программа А могла провести рокировку в момент времени  $t$ , но во вселенной в момент времени  $t$  рокировки произойти не могло.

Что дает нам право так описывать ситуацию? Все просто: если мы считаем программу А *отделенной от ее непосредственной среды* – куда входит и генератор случайных чисел, – то не предопределено, проведет ли программа А рокировку. Это зависит от того, что, строго говоря, находится за пределами программы А. В момент времени  $t$  вселенная пребывала в таком состоянии, которое не допускало проведение рокировки программой А, но программа А в этом “не виновата”. Программа Б, напротив, провести рокировку не могла, потому что рокировка была не в ее природе. Чтобы представить, как программа Б проводит рокировку, нам пришлось бы внести слишком много поправок в реальность.

Это полезное открытие – мы нашли различие между тем, что “могли бы” сделать программы А и Б, которое *не зависит* от индетерминизма. Даже в детерминистическом мире мы видим, что программа А *может* делать вещи, которые *не может* делать программа Б, и это различие отчасти *объясняет*, почему программа А побеждает программу Б. Тот факт, что в этом мире царит истинный детерминизм, а потому программы А и Б могут делать только то, что они на самом деле делают в конкретном случае (и делали бы снова и снова при повторении *точно* тех же самых обстоятельств), просто не интересен и не имеет отношения к получаемому нами объяснению совершенно объективной и наглядной закономерности: программа А побеждает программу Б.

Шахматная программа не является моральным агентом и не несет моральной ответственности за свой выбор – ее мир абсолютно аморален, и нарушение одного из шахматных правил для шахматной программы просто немыслимо, а следовательно, не требует штрафов. Однако, как мы только что увидели, даже в простом детерминистическом мире компьютерных шахмат мы можем найти реальное и важное различие между программами А и Б. Порой, когда программа А делает глупость или наоборот поступает мудро, мы можем сказать: “Программа А могла бы поступить иначе, но программа Б иначе поступить не могла”. Если вы считаете это ошибкой, “потому что ни программа А, ни программа Б *вообще* не могли бы поступить иначе, поскольку мир детерминистичен”, ошибку на самом деле совершаете вы.

Способности программ А и Б к шахматам различаются, и то, “могла ли [каждая из программ] поступить иначе”, прекрасно описывает один из аспектов этого различия. Что насчет их *моральной компетентности*? Когда люди говорят о других людях, поступивших плохо, что “они могли бы поступить иначе”, и используют это в качестве оправдания, чтобы не прощать их, но при этом соглашаются, что другие люди в подобных обстоятельствах *не могли* поступить иначе, и тоже оказываются правы – это не зависит от того, истинен ли детерминизм. Они указывают на *реальное различие в моральной компетентности*, которое не зависит ни от индетерминизма, ни от детерминизма, и этим может объясняться различие нашей реакции.

Чтобы лучше понять это, встанем на место программиста, который разработал программу Б. Он хочет узнать, не нашел ли он слабое место программы Б. Вот партия, в которой не проведенная вовремя рокировка стоила программе Б победы; помогла ли программа Б провести рокировку в нужный момент? Если для этого нужно было только переключить один-единственный бит в генераторе случайных чисел, то, возможно, улучшить программу не требуется. В половине случаев в подобных обстоятельствах программа Б будет проводить рокировку, и, возможно, на большее не стоит и надеяться. Программа всегда должна время от времени использовать случайные числа (будто бы подбрасывая монетку), чтобы прерывать поиск и продолжать партию, а следовательно, всегда будут случаи, в которых случайное число будет обрывать поиск прямо накануне открытия. Обратите также внимание, что ситуация не улучшится, если мы дадим программе Б (или программе А) квантовый генератор случайных чисел, скажем счетчик Гейгера, который выдает биты на основании неопределенных траекторий субатомных частиц. Что бы мы сказали о программе Б, если бы она не провела рокировку из-за единственного нуля, на месте которого могла бы быть единица? Если квантовый генератор чисел выдает ноль, программа Б проводит рокировку; если же он выдает единицу, программа Б не проводит рокировку. “Программа Б могла бы сделать рокировку”, – замечает наблюдатель, когда появляется единица. Да, но программа Б не вольна поступать как ей вздумается. В серии партий, где возникает подобная возможность, в половине случаев программа Б будет проводить рокировку, а в другой половине – не будет, и неважно, использует она при этом генератор “истинно” или “псевдо-” случайных чисел. Философ Дэвид Уиггинс (1973, р. 54) однажды написал о “вселенской несправедливости” детерминизма, но наш насос интуиции о компьютерном шахматном турнире демонстрирует такую же “вселенскую несправедливость” индетерминизма. Программа Б пребывает “во власти” своего генератора случайных чисел *или* своего генератора псевдослучайных чисел. (Само собой, как и программа А – и как и все мы.) Нет никаких оснований делать выбор в пользу генератора истинно случайных чисел – если только вы не собираетесь играть в шахматы со всеведущим Богом, который насквозь видит ваш генератор псевдослучайных чисел и планирует свои ходы соответствующим образом!

Итак, мы все еще не видим причин желать, чтобы индетерминизм был реален. Возможно, нам достаточно и той свободы воли, которую мы можем получить без вмешательства индетерминизма. Вот другая возможная причина:

Я не могу изменить прошлое, но если индетерминизм реален, я могу изменить будущее!

Нет. Как именно изменить будущее? Сделать его не таким, каким оно *собиралось* стать, а таким, каким оно *собирается* стать? Вы не можете *изменить будущее* точно так же, как не можете *изменить прошлое*. Эта мысль не выдерживает критики. Тогда:

Если детерминизм реален, я не могу изменить будущее, а если детерминизм нереален, я все равно не могу изменить будущее. Следовательно, я не могу изменить будущее.

Почему кажется, что мы хотим изменить будущее? Потому что мы хотим иметь возможность предвидеть катастрофы и предпринимать какие-то шаги для их предотвращения. И мы *можем* делать это

вне зависимости от индетерминизма. Если кто-то швырнет в вас кирпич, а вы заметите это и пригнетесь, вы сможете избежать удара кирпичом. И это прекрасно. Должно ли было случиться это столкновение? В некотором роде да, ведь кирпич явно летел прямо вам в голову, но вы заметили его (вас заставил заметить его отраженный свет, который вы уловили глазами, после чего ваш мозг рассчитал риск и был вынужден предпринять меры) и потому избежали удара. Само собой, если бы вы хотели *избежать избегания* (если вы нашли какую-то причину, по которой вам показалось выгоднее принять удар), вы могли бы поступить именно так. Какой-то наблюдатель, возможно, не сумел бы сказать до последней секунды, примете вы удар или нет. И он бы проиграл, если бы сделал ставку на то, что вы пригнетесь. Мы вернулись к нашей причине стремиться к непредсказуемости, которая не требует индетерминизма.

Что дает нам этот насос интуиции? Он берет знакомую фразу “не мог поступить иначе” и показывает, что, вопреки широко распространенному, но плохо взвешенному мнению, ее значимая вариация *не зависит* от реальности индетерминизма. Даже если какое-то из значений фразы “не мог поступить иначе” несовместимо с детерминизмом и при этом имеет важность с моральной точки зрения – а не представляет собой исключительно метафизический курьез, – его только предстоит обнаружить, после чего бремя доказывания будет лежать на тех, кто так считает. Еще одно “очевидное” соображение при ближайшем рассмотрении оказалось не таким уж и очевидным.

## 71. Конечная ответственность

Пока что мы рассматривали ситуации элементарного выбора, не предполагающего никакой моральной ответственности: игру в “камень, ножницы, бумагу”, шахматные ходы и уход от удара кирпичом. Возможно, индетерминизм действительно желателен только тогда, когда мы подвергаем анализу наши попытки быть моральными агентами, а не просто интенциональными системами наподобие шахматных компьютеров и выпрыгивающих газелей. Так полагали многие мыслители. По их мнению, описанные ранее упражнения лишь отвлекают от главной проблемы. Ниже приводится четкое изложение того, что ряд мыслителей считает первостепенным. Формулировкой мы обязаны философу Галену Стросону (2010):

1. Вы делаете то, что делаете, в любой рассматриваемой ситуации, поскольку вы такой, какой есть.
2. Следовательно, чтобы нести конечную ответственность за свои действия, вы должны нести конечную ответственность за то, какой вы есть, – по крайней мере, в некотором критически значимом психическом отношении.
3. Однако вы не можете нести конечную ответственность за то, какой вы есть, ни в каком отношении.
4. Следовательно, вы не можете нести конечную ответственность за свои действия.

Первое утверждение неоспоримо: в “то, какой вы есть”, включается и ваше состояние в конкретный момент времени, как бы вы в него ни вошли. Каким бы ни было это состояние, ваши действия вытекают из него обычным, не чудесным образом. Во втором утверждении сообщается, что вы не можете нести “конечную” ответственность за свои действия, если только не несете “конечную” ответственность за то, как вы вошли в это состояние, по крайней мере в некотором отношении. Однако, согласно пункту (3), это невозможно.

Таким образом, кажется, что пункт (4), вывод, логически проистекает из вышесказанного. Несколько мыслителей сочли этот аргумент решающим и важным. Но так ли это на самом деле? Давайте внимательнее рассмотрим пункт (3). Почему вы не можете нести конечную ответственность за то, какой вы есть, хотя бы в *некотором* отношении? В повседневной жизни мы делаем именно такое различие, и это важно с моральной точки зрения. Допустим, вы проектируете и конструируете робота и отправляете его в мир, где он остается без присмотра и без контроля. Вы прекрасно знаете, на что он способен, и в какой-то момент он наносит кому-то серьезное увечье. Разве вы не несете ответственности за это – хотя бы в некотором отношении? Большинство людей назовут вас ответственным. Это вы сделали робота и должны были предвидеть опасности – и даже *предвидели* ряд опасностей, – а потому винить за причиненный ущерб, по крайней мере отчасти, тоже следует вас. Мало кто проникся бы к вам сочувствием, если бы вы стали утверждать, что *вообще* не несете ответственности за ущерб, причиненный вашим роботом.

Теперь рассмотрим немного другой случай: вы проектируете и конструируете человека (себя в будущем) и отправляете себя в полный опасностей мир, прекрасно понимая, с какими угрозами вам предстоит столкнуться. Вы напиваетесь в баре, садитесь в машину и нажимаете на газ. Разве вы не несете ответственности – хотя бы частично – за то, “каким вы были”, когда врезались в школьный автобус? Здравый смысл подсказывает: конечно, вы несете за это ответственность. (Ответственность с вами может делить бармен или чересчур радушный хозяин, в гостях у которого вы пили.) Но как это возможно перед лицом категоричного аргумента Стросона? Не забывайте, Стросон говорит, что вы не можете нести *абсолютную* ответственность за то, какой вы есть. И что тогда? Кто вообще считает, что важно нести *абсолютную* ответственность? Такое состояние уж точно невозможно, *даже если индетерминизм реален!* (Хватит полагать, что этот аргумент дает нам основание надеяться на реальность индетерминизма.) Вот что говорит сам Стросон (2010):

Чтобы нести абсолютную ответственность за свои действия, необходимо быть *causa sui*, причиной самого себя, а это невозможно (и не было бы возможно, даже если бы мы обладали нематериальными душами, а не были полностью материальны).

*Абсолютная* ответственность – фальшивка. Это благо, к которому никто не должен стремиться. Стросон (2003) считает иначе и критикует меня за пренебрежение ею:

Он не доказывает существование абсолютной свободы воли и моральной ответственности, в которую хочет верить и верит большинство людей. Доказать это невозможно, и он это знает.

Он совершенно прав: я не доказываю существование такой свободы воли, в которую хочет верить большинство людей, и я это знаю. Но я полагаю, что они ошибаются, если желают верить в ее существование или если действительно верят в него. Стросону и его сторонникам нужно доказать, почему нам в нашей повседневной жизни должно быть дело до конечной ответственности – или до проблемы детерминизма/индетерминизма. Они могут предложить определение такого типа свободы воли, который несопоставим с детерминизмом, и показать, что многие считают это важным, но в то же время им необходимо доказать, что эти многие не заблуждаются. Какое кому до этого дело? (Обратите внимание на мой риторический вопрос. Я специально лезу на рожон, но буду счастлив, если Стросон или еще кто-нибудь вступит в дискуссию и попытается ответить на мой вопрос. Впрочем, пока никто не вызвался.)

Прежде чем закончить с аргументом Стросона, позвольте поинтересоваться, заметили ли вы его поразительное сходство с другим аргументом, рассматривавшимся ранее. Я переформулирую его по аналогии с аргументом Стросона, чтобы сходства стали очевиднее.

1. Млекопитающее является млекопитающим в любом рассматриваемом контексте, поскольку оно такое, какое есть.
2. Чтобы быть млекопитающим, вам необходимо стать таким, какой вы есть, родившись у матери, которая является млекопитающим.
3. Но то же самое должно быть истинным и для вашей матери, и для ее матери – и так далее до бесконечности, а это невозможно.
4. Следовательно, вы не можете быть млекопитающим, поскольку существование млекопитающих абсолютно невозможно.

С такими “наследственными” аргументами всегда надо держать ухо востро. Почти наверняка они оказываются замаскированными вариациями древних ложных заключений, называемых *соритами* (или “парадоксом кучи”):

1. Одно зернышко пшеницы не является кучей.
2. Если добавить к одному зернышку еще одно зернышко, кучи не образуется.
3. Если добавлять к тому, что кучей не является, по одному зернышку, кучи не получить.
4. Следовательно, куч вообще не существует!

Философы на протяжении тысяч лет писали о парадоксе кучи и проблеме неопределенности терминов (которая и лежит в основе этого парадокса), но до сих пор не найдено способа диагностировать ошибку и научиться ее избегать. (Прекрасное и актуальное исследование по теме см. в онлайн-версии “Стэнфордской философской энциклопедии”.) Немногочисленные отважные философы даже признают сориты состоятельными и пытаются смириться с “фактом”, что в мире нет ни лысых мужчин, ни мужчин, которые лысыми не являются. Отстоять такую точку зрения непросто! Однако, как показано в главе 43, Дарвин научил нас отказываться от соритов: нам нет нужды искать “принципиальную” линию раздела между категориями, выстроенными в наследственном порядке.

Насколько мне известно, я первый обратил внимание на сходство аргумента Стросона – и других аргументов подобного рода, приводимых в литературе о свободе воли, – с парадоксом кучи, но это сходство существует. Думаю, тот факт, что люди принимают на себя моральную ответственность постепенно, в процессе взросления, столь же очевиден, как и постепенное, растянувшееся на целые эпохи перерождение рептилий и терапсид в млекопитающих. Чтобы быть млекопитающим, вам не нужно быть *абсолютным* млекопитающим; чтобы нести ответственность, вам не нужно нести *абсолютную* ответственность; а чтобы обладать желаемой свободой воли, вам не нужно обладать *абсолютной* свободой воли. На самом деле, поскольку абсолютная свобода воли относится к разряду чудес, кто-то должен привести неоспоримый аргумент, показывающий, зачем вообще к ней стремиться. Может, чтобы стать Богом? К несчастью, этого не достичь, но доступного человеку и так вполне достаточно.



## 72. Сфексовость

Даг Хофштадтер (1982) предложил термин “сфексовость” для знакомой всем механистичной, роботоподобной бездумности, которую часто принимают за великую смышленость. Прекрасный пример сфексичности демонстрируют сфексы – осы, которые ведут себя весьма любопытным образом (название их рода и легло в основу нового термина). Мы с Дагом независимо друг от друга восхитились фрагментом научно-популярной книги “Механика мозга”, написанной Дином Вулдриджем (1963), который следующим образом описал поведение сфексов:

Когда приходит время откладывать яйца, сфекс строит для этой цели гнездо и подыскивает сверчка, которого жалит таким образом, чтобы парализовать, но не убить. Она затаскивает сверчка в гнездо, откладывает рядом яйца, закрывает вход в гнездо и улетает навсегда. В надлежащий срок из яиц вылупляются новые сфексы. Они начинают питаться парализованным сверчком, который не разложился, поскольку содержался в осином эквиваленте глубокой заморозки. С человеческой точки зрения, сложность и кажущаяся целесообразность этой процедуры убедительно свидетельствует о логике и дальновидности сфексов, но только если не вдаваться в детали. К примеру, оса действует следующим образом: она подтаскивает парализованного сверчка к гнезду, оставляет его у входа, заходит внутрь, чтобы удостовериться, что там все в порядке, выходит обратно и затем затаскивает сверчка в гнездо. Если сдвинуть сверчка на несколько сантиметров, пока оса проводит предварительную проверку в гнезде, она подтащит его обратно к входу, но не станет втаскивать внутрь, а вместо этого повторит подготовительную процедуру проверки состояния гнезда. Если сверчка снова сдвинуть, пока оса будет в гнезде, она опять подтащит его к входу и войдет в гнездо, чтобы проверить его в последний раз. Оса не думает, что можно сразу втащить сверчка в гнездо. Один раз эта процедура повторялась сорок раз – и результат оставался неизменным. [р. 82]

Казалось бы, это идеальный пример неидеальной компетентности без понимания, которую мы обнаруживаем, когда разоблачаем поверхностное псевдопонимание второсортной компьютерной программы. Однако недавно мы узнали, что Вулдридж описал феномен упрощенно – как часто делают авторы научно-популярных книг. Психолог Ларс Читтка в письме ко мне привел цитату из работы Жан-Анри Фабра (1879 г.!), очевидно, ставшей источником для Вулдриджа, который узнал бы, что сфексовость характерна лишь для некоторых сфексов, если бы прочитал Фабра дальше! Фабр подробно все описал. Если на первый взгляд кажется, что сфексы умны, на второй – что они глупы, стоит взглянуть на них в третий раз и обнаружить, что некоторые сфексы вообще не проявляют особой сфексовости. Читтка прислал мне немецкий перевод Фабра (я до сих пор не нашел этот текст на французском), в котором есть следующая фраза: *Nach zwei oder drei Malen... packt ihre Fühler mit den Kieferzangen und schleift sie in die Höhle. Wer war nun der Dummkopf?* (“После двух-трех повторений... она схватила [жертву] за усики своими жвалами и втащила ее в нору. И кто теперь дурак?”)

Итак, прилагательное “сфексовый” немного неточно, но оно уже прижилось, а потому сфексам придется смириться с этим оскорблением. В некотором роде им повезло оказаться в центре внимания неспециалистов, а это, вероятно, может хорошо сказаться на их выживаемости! (Чью среду обитания вы захотите сохранить – среду сфексов или среду заурядниц?) Может, сфексы и не так “харизматичны”, как слоны, тигры или волки, но они довольно известны – благодаря своей проблематичной сфексовости.

Сфексовость важна не столько потому, что ее проявляют (в разной степени) многие интактные, простые животные – насекомые, черви, рыбы, – сколько потому, что она дает нам термин для обозначения ограниченных, автоматизированных, недальновидных компетенций, из которых можно сконструировать более сложное, более разностороннее, способное к пониманию сознание. Кирпичики любой модели сознания должны быть сфексовыми! Или, как я заметил ранее, кирпичики должны быть **вроде как** сознанием, бледной тенью нашего сознания. Сфексовость также полезна, чтобы отличать морально компетентное сознание от морально некомпетентного. В тех случаях, когда человек проявляет сфексовость, будь то из-за опухоли мозга, травмы мозга, серьезного дисбаланса нейромодуляторов, психической болезни, простого невежества или незрелости, этот человек не мог бы поступить иначе в интересующем нас смысле.

Упорный биолог, мешающий сфексу, чтобы продемонстрировать его сфексовость, представляет собой образец агента-манипулятора, которого мы по праву боимся. Мысленные эксперименты многих философов задействуют именно такого кукловода или подлого нейрохирурга, который втайне запрограммировал кого-то играть себе на руку. По-видимому, мораль этих страшилок в том, что, даже если кукловода на самом деле не существует, тот факт, что наше поведение обуславливается различными характеристиками окружающей среды, анализируемыми нашими системами восприятия и мозгом, показывает, что кукловод с таким же успехом мог бы и существовать. (На обложке небольшой книжечки Сэма Харриса “Свобода воли” [2012] изображены нити марионеток.) Но этот напрашивающийся вывод явно нелогичен. Если “контроль” среды осуществляется с помощью наших отлаженных систем восприятия и незамутненного сознания, бояться нам нечего. На самом деле нам даже в высшей степени желательно, чтобы объекты и события вокруг нас провоцировали наше сознание генерировать истинные убеждения о них, которые мы могли бы использовать для корректировки своего поведения в свою пользу! Фотон, отскакивающий от полостей в приливной зоне прямо ко мне в глаза, могут заставить меня схватить корзину и грабли и выйти на охоту за моллюсками. Если в этом случае мои действия контролирует среда, я ничуть не возражаю. Как и большинство людей, я не пугаюсь и не чувствую, что мной манипулируют, когда друзья предлагают мне роскошные блюда, прекрасно зная, что я не сумею устоять перед искушением полакомиться.

Встречая в литературе о свободе воли примеры с кукловодом и нейрохирургом, вы также заметите, что вмешательство всегда – всегда – происходит тайно. Почему? Потому что сделать вывод о том, что наша воля не свободна, можно только в том случае, если другой, тайный агент заставляет нас действовать определенным образом, хотя мы сами этого *не осознаем*. Причину этого долго искать не надо – она отсылает нас к наблюдению, лежащему в основе теории игр: если агент знает о попытке манипуляции со стороны другого агента, он начинает принимать ответные меры и как минимум корректирует свое поведение, чтобы справиться с этим открытием. Конкурентное взаимодействие двух агентов предполагает предоставление обратной связи на множестве уровней, а следовательно, ослабление “контроля” потенциального манипулятора. Если же вмешательство не только не является тайным, но и затребовано “куклой”, ситуация полностью меняется.

Мы можем продемонстрировать это, просто скорректировав стандартные примеры и покрутив регуляторы насосов интуиции. Философ Гарри Франкфурт (1969) сконструировал насос интуиции, где за человеком следит тайно имплантированное устройство нейростимуляции, контролируемое нейрохирургом, который обеспечивает, чтобы человек принимал решения в соответствии с желаниями нейрохирурга. Если вы, будучи этим человеком, выбираете вариант А вместо варианта Б и если именно этого и хотел нейрохирург, он ничего не делает; если же его приборы показывают, что вы собираетесь сделать выбор в пользу варианта Б, он нажимает на кнопку, не позволяя вам выбрать вариант Б, и вы в результате выбираете вариант А. Вы ничего не чувствуете. Был ли ваш выбор свободным в каждом из случаев? За годы философы исписали сотни страниц, рассуждая о многочисленных вариациях “случая Франкфурта”, но, насколько мне известно, один поворот регуляторов никогда не исследовался, хотя он существенно меняет дело. Вот он:

Желая подавить свою страсть к калорийным десертам, вы обратились к хорошему врачу, который имплантировал вам описанное устройство. Вы щедро платите ему, чтобы он следил за каждым вашим приемом пищи и страховал вас от заказа мороженого с шоколадным соусом и чизкейков. Вы оба надеетесь, что ему ни разу не придется нажать на кнопку, и вскоре вы почти забываете, что он или его ассистент всегда находятся рядом с вами в электронном виде. Вы излечились: устройство фиксирует стабильную последовательность сотен случаев выбора в пользу варианта А (“Нет, спасибо, мне только чашку черного кофе”). Если это не примеры *вашего* ответственного и свободного выбора, то почему? Разве вы не поступили мудро, когда решили подстраховаться своей не слишком надежной силой воли?

### 73. Мальчики из Бразилии: еще один упорный башмак

В 2004 г. психологи Джошуа Грин и Джонатан Коэн написали статью “В плазах закона нейробиология меняет все и не меняет ничего”, которая была опубликована в престижном журнале *Philosophical Transactions of the Royal Society*. В этой авторитетной статье авторы воображают правовую революцию, вызванную научными открытиями.

Закон *гласит*, что он допускает не более чем метафизически скромное представление о свободе воли, которое полностью совместимо с детерминизмом. Однако мы утверждаем, что интуитивная поддержка закона в конечном счете основывается на метафизически амбициозном, либертарианском [индетерминистическом] представлении о свободе воли, которому угрожает детерминизм и – более демонстративно – когнитивная нейронаука будущего. [р. 1776]

Их позиция нетривиальна – но так и должно быть, ведь они признают, что существует множество аргументов в поддержку компатибилизма (отстаиваемого мною на этих страницах), но при этом хотят показать, что на самом деле в вопросе о свободе воли “все мы стоим на перепутье”. Они предлагают мысленный эксперимент, который должен продемонстрировать зависимое положение повседневного практического представления об индетерминизме. Обратите внимание, что именно об этом я только что и просил: покажите нам, пожалуйста, что нам должно быть дело до индетерминизма и почему. Их мысленный эксперимент вдохновлен фильмом “Мальчики из Бразилии”, в котором рассказывается о нацистских ученых, растящих клонов Гитлера (созданных с применением сохранившейся ДНК). Вот описание эксперимента:

Допустим, группа ученых сумела создать человека – назовем его Марионеткой, – который по умолчанию втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков. [р. 1780]

Вот что они говорят о своем эксперименте:

Да, он столь же рационален, как другие преступники, и да, его действия основаны на его желаниях и убеждениях. Однако эти убеждения и желания были сформированы извне, а потому на интуитивном уровне мы понимаем, что он заслуживает скорее сострадания, чем морального осуждения... Чем мы отличаемся от Марионетки? Очевидно, что Марионетка – жертва коварного заговора, а большинство людей, как мы полагаем, жертвами не являются. Но важно ли это? Мысль о том, что Марионетка не несет полной ответственности, основана на представлении, что его действия были предопределены извне... Но тот факт, что эти силы связаны с желаниями и намерениями коварных ученых не относится к делу, верно? Важно лишь то, что эти силы неподвластны контролю Марионетки и ему не принадлежат. [р. 1780]

Что вы думаете? Подходит ли этот насос интуиции для обозначенной цели? Забавно, что авторы также считают этот вопрос достойным рассмотрения: “Дэниел Деннет может возразить, что история Марионетки – некорректный «насос интуиции»”. О да. На мой взгляд, это **упорный башмак**. Но они отмахиваются от этого и продолжают: “Нам кажется, что чем больше становится известно о Марионетке и его жизни, тем меньше хочется считать его в полной мере ответственным за свои действия и признавать его наказание самоцелью”.

Давайте внимательнее присмотримся к этому насосу интуиции и покрутим его регуляторы, чтобы разобраться, как он работает. Я предлагаю повернуть четыре регулятора. Во-первых, давайте избавимся от коварного заговора – тем более что сами авторы настаивают, что он не имеет значения. Я удивлен, что они столь беспечно вводят “подлого нейрохирурга” и считают свой шаг совершенно невинным, ведь этот ход оспаривается годами, однако мы можем проверить их убежденность, заменив “группу ученых” “нейтральной средой”:

до: Допустим, группа ученых сумела создать человека – назовем его Марионеткой, – который по умолчанию втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков.

после: Допустим, *нейтральная среда* сумела создать человека – назовем его Марионеткой, – который по умолчанию втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков.

Второй регулятор: поскольку заговорщиков больше нет, нужно заменить “по умолчанию” на “с высокой вероятностью”:

до: Допустим, группа ученых сумела создать человека – назовем его Марионеткой, – который *по умолчанию* втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков.

после: Допустим, *нейтральная среда* сумела создать человека – назовем его Марионеткой, – который *с высокой вероятностью* втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков.

Третий регулятор: я хочу изменить мотивацию преступления – пусть это будет убийство, но в других обстоятельствах. (Это ведь не должно играть роли?)

до: Допустим, группа ученых сумела создать человека – назовем его Марионеткой, – который по умолчанию втягивается в преступную деятельность: скажем, совершает убийство *во время неудачной сделки по покупке наркотиков*.

после: Допустим, *нейтральная среда* сумела создать человека – назовем его Марионеткой, – который с высокой вероятностью втягивается в преступную деятельность: скажем, совершает убийство, *чтобы замести следы растраты имущества*.

Четвертый регулятор: давайте изменим имя злоумышленника. В конце концов, это всего лишь имя.

до: Допустим, группа ученых сумела создать человека – назовем его Марионеткой, – который по умолчанию втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков.

после: Допустим, *нейтральная среда* сумела создать человека – назовем его Автономом, – который с высокой вероятностью втягивается в преступную деятельность: скажем, совершает убийство во время неудачной сделки по покупке наркотиков.

На какие мысли вас теперь наталкивает этот насос интуиции? На те же самые? Вы склоняетесь к состраданию или к осуждению? Возможно, нам помогут дополнительные детали. Вот моя редакция насоса интуиции Грина и Коэна:

Автоном изучал экономику в Гарварде и после получения диплома устроился на работу в банк *Lehman Brothers*, где оказался в среде людей, которые обманывали систему и сколачивали огромные состояния. Он влюбился в падкую на роскошь сердцеедку, которая пригрозила бросить его, если он не разбогатеет в ближайшем будущем. Он заметил, что можно организовать практически незаметную растрату и почти наверняка не попасться на этом, а потому осознанно рискнул, но – увы! – вопреки всем ожиданиям нашелся свидетель, который совершенно напрасно встал слишком близко к ограждению сада на крыше... а потом “оступился” – ой! – и разбился насмерть при падении с высоты. Возникли подозрения, и вскоре Автонома взяли под арест.

Вы по-прежнему склоняетесь к мысли, что он “не в полной мере ответственен” за свои действия, *потому что эти действия были предопределены извне*? Даже если вам по-прежнему хочется считать Автонома *жертвой* своей (роскошной) среды, я полагаю, вы согласитесь, что это желание сильно притупилось, а возможно, и вовсе стало лишь отголоском прошлого. (А может, я просто умело поиграл с вашими чувствами, рассчитывая на ваше кровожадное стремление наказать алчных типов с Уолл-стрит за их роль в экономическом спаде.) Я не утверждаю, что мои вариации доказывают, что люди несут или могут нести ответственность за свои действия, даже если эти действия были предопределены извне; я просто говорю, что конкретно этому насосу интуиции верить ни в коем случае нельзя, поскольку (доступные, разрешенные) повороты регуляторов слишком сильно влияют на наши суждения. Возможно, его создали не для того, чтобы напустить туману, однако он явно не способствует ясности мышления.

## Резюме

Людам очень важно обладать свободой воли, но в то же время складывается впечатление, что они не слишком хорошо понимают, что такое свобода воли и какой она может быть (подобно тому, как они не слишком хорошо понимают, что такое цвет и сознание). Наши решения не маленькие чудеса, которые происходят в мозге, нарушая законы физики и химии, управляющие остальными процессами в наших телах, хотя многие и считают, что только таким образом наши решения могут быть истинно свободными. Однако мы не можем на основании этого заключить, что *не обладаем* свободой воли, поскольку такое дикое представление о свободе воли все же не единственное. Закон – и здравый смысл – различает подписание договора “по собственной воле” с подписанием договора под давлением или под влиянием галлюцинаций или других психических расстройств. Это знакомое нам представление о свободе воли, обусловленное множеством практик и взглядов, составляющих нашу манифестную картину мира, не проявляет зависимости от описанного выше дикого представления.

Сотни лет некоторые философы настаивали, что важно именно такое представление о свободе воли, что только оно для нас и значимо, ведь оно полностью совместимо с детерминизмом, материализмом, а также законами физики и химии. Описанные в этом разделе насосы интуиции и другие инструменты мышления призваны поддержать и углубить понимание этого учения, именуемого *компатибилизмом*. За годы было предложено немало вариаций на тему компатибилизма – и, пожалуй, с ним согласны не только философы, но и судьи, юристы и другие люди, вынужденные определять, кто и за что несет ответственность, а также кто от ответственности освобождается, поскольку действовал не по собственной воле. Некоторые ученые сегодня бросают вызов этим представлениям – и, возможно, не зря. Давайте внимательно рассмотрим их аргументы.

Возможно, наука учит нас чему-то радикальному, даже революционному: что никто и никогда не несет ответственности за свои действия и нет никаких оснований считать одни поступки достойными похвалы, а другие – заслуживающими порицания. Но столь революционный вывод требует гораздо больше добросовестного внимания к деталям, чем проявляют оглашающие его ученые. Подлый нейрохирург искалечила пациента, не используя ничего, кроме ложной идеи, а ошибочные взгляды влиятельного ученого могут лишить людей обоснованного и улучшающего жизнь представления о свободе воли. В этом вопросе всем следует проявлять особенную осторожность.

Несмотря на свою популярность среди философов, компатибилизм всегда вызывал подозрения. Как известно, Иммануил Кант назвал его “гнусной уловкой”, и сегодня многие авторы сомневаются в искренности тех, кто его отстаивает. На самом деле все так и должно быть. Наука учит нас с особенной осторожностью относиться к самообману, и многие правила научных изысканий созданы как раз для того, чтобы не позволять нам предаваться надеждам, когда свидетельства кажутся нам вполне убедительными. Представьте, что одни астрономы объявят, что через десять лет на нашу планету упадет гигантский астероид, который уничтожит все живое, а затем другая группа астрономов скажет, что мы можем спать спокойно, поскольку они заново проанализировали данные и пришли к выводу, что астероид пройдет в непосредственной близости от Земли, но столкновения не случится. Конечно, это хорошая новость, но как нам знать, что вторая группа астрономов не заблуждается – или не вводит нас в заблуждение своей ложью во спасение? Снова и снова проверяйте их расчеты, стройте собственные модели – только не принимайте на веру их вывод просто потому, что он не содержит очевидных ошибок и приходится вам по душе. Но также не забывайте, что они могут оказаться правы. Не совершайте другую ошибку и не ставьте под сомнение – на “общих основаниях” – то, что кажется “слишком хорошим, чтобы быть правдой”.

Не слишком ли хорош компатибилизм, чтобы быть правдой? На мой взгляд, нет. Я считаю, он *и есть* правда, а потому мы можем с полным правом отмахнуться от паникеров, в то же время корректируя и пересматривая наше понимание основ концепции моральной ответственности. Но это задача на будущее – и в одиночку с ней не справиться. Насколько я вижу, это самая сложная и самая важная философская задача, стоящая перед нами сегодня. Ставки высоки, вопросы щекотливы, а чувства нередко мешают нам мыслить здраво. Нам понадобятся все имеющиеся у нас инструменты мышления – и не только. Новые инструменты придется создавать по ходу дела.

## IX.

### Что значит быть философом?

Стоя слишком близко, сложно понять, что перед нами. Философы, ученые и художники давно стараются “делать привычное странным”. В последние годы я меньше, чем раньше, общался с другими философами, но при этом больше общался с учеными и прочими мыслителями. Но я по-прежнему философ (что бы ни говорили некоторые философы!) и с удовольствием объясняю нефилософам, в чем польза философии. Тем, кто совсем далек от философии, она часто кажется нелепостью – прекрасным образчиком бесполезного ума. Кое-чего они не замечают (вспомните **закон Старджона**: 90 процентов чего угодно – полное дерьмо, а значит, они не замечают оставшихся 10 процентов). За пятьдесят лет я успел неплохо познакомиться с философией, а теперь, проводя немало времени вдали от философии, я также хорошо вижу ее странности. Некоторые из моих друзей и коллег-ученых признаются, что не понимают, почему я не брошу все и не пополню их ряды. Краткий ответ таков: раздвигая границы, я сумел взять лучшее от обоих миров. Работая с учеными, я купаюсь в изобилии любопытных и проблематичных *фактов*, о которых можно мыслить, но оставаясь философом без лаборатории и гранта на исследование, я размышляю обо всех теориях и экспериментах, не пачкая рук. Как поется в одной из моих любимых песен Гершвина, это “непыльная работенка”.

Мне кажется, что в последнее время ученые, к счастью, уделяют философам больше внимания и относятся к ним с большим уважением, чем раньше, и особенно это видно в сфере моей специализации – философии сознания, где когнитивная наука исследует почти те же самые феномены, над которыми веками размышляли философы, а именно восприятие, память, значения, волеизъявление и сознание. И философы – некоторые из них – заслужили внимание, изучая соответствующие научные открытия и помогая прояснять и углублять научные исследования, а также находить лучшие способы объяснять получаемые результаты далеким от науки людям. Однако между лагерями ученых и философов по-прежнему нередко возникает недопонимание, в связи с чем я хочу коснуться нескольких различий в подходах, чтобы наладить лучшее взаимодействие в будущем.

## 74. Сделка с дьяволом

Несколько лет я задавал своим коллегам-философам такой вопрос: если бы Мефистофель предложил вам два следующих варианта, что бы вы выбрали?

(А) Вы *решаете* серьезную философскую проблему по своему выбору так убедительно, что пресекаете все дальнейшие разговоры на эту тему (благодаря вам этот сектор философии закрывается навсегда, и ваше имя входит в историю).

(Б) Вы пишете книгу такой поразительной запутанности и противоречивости, что она на века остается в списке литературы для обязательного чтения.

Некоторые философы неохотно признаются, что выбрали бы вариант (Б). Если бы им пришлось выбирать, они предпочли бы, чтобы их читали, а не чтобы они оказались правы. Подобно композиторам, поэтам, писателям и другим людям искусства, они хотят, чтобы к их работе снова и снова обращались миллионы (а по возможности и миллиарды!) людей. Но они также тянутся в сторону научного познания. В конце концов, *предполагается*, что философы должны стремиться к истине.

Когда я задаю тот же вопрос ученым, они без колебаний склоняются к варианту (А) – им на этот счет и думать нечего. Узнавая, что многим философам, ряд которых несколько смущенно склоняется к варианту (Б), этот выбор дается не без труда, ученые с удивлением (или отвращением?) качают головой. Но такая реакция ученых показывает, что они упускают из виду важную мысль, озвученную Николасом Хамфри (1987) (см. главу 48):

В “Двух культурах” Ч. П. Сноу восславил великие открытия науки как “научного Шекспира”, но в одном он в корне ошибся. Пьесы Шекспира принадлежали Шекспиру – и никому больше, в то время как научные открытия, напротив, не принадлежат – в конечном счете – никому.

Если бы Шекспира не существовало, никто бы не написал “Гамлета”, “Ромео и Джульетту” и “Короля Лира”. Если бы Ван Гог не существовало, никто бы не написал “Звездную ночь”. Возможно, я несколько преувеличиваю, но в этом что-то есть. В творениях великих людей искусства есть индивидуальность, которая не только редко встречается, но и совершенно не имеет значения в науке. Знаменитые споры о первоочередности научных открытий и сражения за Нобелевскую премию так жестоки именно потому, что кто-то еще *мог бы* сделать ровно то же открытие, к которому стремились вы, и если бы вы стали вторым, о признании вам не стоило бы и мечтать. В искусстве таких столкновений не случается, поскольку им правят другие принципы.

Некоторые ученые мечтают добиться высокой читаемости и угодить читателям. Работы лучших из них имеют немалую литературную ценность. Здесь на ум приходят книги Дарвина. Но на первом месте всегда стоит задача разложить все по полочкам и убедить читателей в открытой истине – это очевидно, если сравнить “Путешествие натуралиста вокруг света на корабле «Бигль»” Дарвина с “Моби Диком” Мелвилла. Из “Моби Дика” можно многое узнать о китах и китобойном промысле, однако перед Мелвиллом не стояла задача написать интересный и убедительный справочник китобоя.

Держа в уме различие между задачами науки и искусства, я предлагаю ученым ответить на вопрос, который действительно сопоставим с вопросом, обращенным к моим коллегам-философам: если бы Мефистофель предложил вам два следующих варианта, что бы вы выбрали?

1. Вы побеждаете в гонке за совершение открытия, которое становится основой для огромной экспансии научного знания (и получаете за это Нобелевскую премию), но впоследствии становится понятно, что это открытие, как выразился Хамфри, на самом деле никому не принадлежит. (Здесь вспоминаются Крик и Уотсон: если бы они не победили в гонке в тот момент, несомненно, вскоре это удалось бы Лайнусу Полингу или еще кому-нибудь.)

2. Вы предлагаете такую оригинальную, такую немыслимую до вас теорию, что ваша фамилия становится нарицательной, но ваша теория оказывается по большей части неверной, хотя и продолжает годами – и даже столетиями – рождать важные споры. (Здесь мне на ум

приходят картезианский дуализм о сознании, теория эволюции Ламарка, бихевиоризм Скиннера и представления Фрейда обо всем – от младенческой сексуальности и неврозов до искусства, музыки и литературы.)

Более удачным, хотя и менее известным примером варианта (2) могут служить амбициозные труды Декарта по физике, которые были настолько весомы и настолько неверны, что стали в итоге провокацией для Исаака Ньютона, намеренно назвавшего свой революционный труд (написанный в 1687 г.) “Математическими началами натуральной философии”, чтобы его заглавие перекликалось с “Первоначалами философии” Декарта (1644 г.) и не возникало сомнений, какое представление о мире Ньютон намеревается изменить. Не стоит забывать и о лингвистике Хомского. Она явно проходит проверку на оригинальность. Как было при победе яхты “Америка”, которая дала имя Кубку “Америки”, когда Хомский появился на сцене, соперников на горизонте было не разглядеть. В последующие годы изначальное теоретическое зерно – “трансформационная” теория, изложенная в “Синтаксических структурах” Хомского (опубликованных в 1957 г.), – было в основном забыто и сменилось разнообразными потомственными теориями, столь же отличавшимися от своего общего предка, как страусы, колибри и альбатросы отличаются от динозавров, от которых они произошли. Значит ли это, что Хомский в 1957 г. совершил плодотворную ошибку, или же он **скорее** (динь!) открыл великую истину? Ответ “да” подходит здесь как нельзя лучше.

Мы уважаем ученых, совершающих полезные ошибки, – вспомните Вольфганга Паули, который, как известно, презрительно отозвался о работе одного теоретика, сказав, что она “даже не ошибочна”. Но, если бы вам пришлось выбирать, променяли ли бы вы возможность стать первым и правым на шанс оказаться оригинальным и провокационным? Решить не так-то просто, правда?



## 75. Философия как наивная автоантропология

Исследователь искусственного интеллекта Патрик Хейс однажды занялся проектом по аксиоматизации наивной (или народной) физики жидкостей. Идея заключалась в том, чтобы наделить робота установками, которые понадобились бы ему в качестве основных убеждений, если бы ему пришлось взаимодействовать с людьми (которые каждый день опираются на народную физику). Это оказалось сложнее, чем предполагал Хейс, и он написал о проекте любопытную статью под названием “Манифест наивной физики” (Hayes 1978). В наивной физике жидкостей все, что кажется наивным людям противоречащим здравому смыслу, конечно же, отвергается: сифоны “невозможны”, как и пипетки, но жидкость можно вытереть махровым полотенцем, а воду из колодца – выкачать с помощью вакуумного насоса. Наделенный таким набором “знаний” робот, как и любой из нас, удивился бы, впервые увидев сифон в действии. Я бы назвал проект Хейса *продвинутой* наивной физикой, поскольку Хейс не питал иллюзий и прекрасно понимал, что теория, которую он пытается аксиоматизировать, неверна, но весьма полезна в повседневной жизни. Можно также назвать его попытку упражнением в аксиоматической антропологии: рассматривая представления, разделяемые людьми, в качестве аксиом и теорем, он пытался упорядочить получаемый набор данных и разрешить все противоречия. Само собой, он не привлекал к сотрудничеству информаторов, отталкиваясь от мысли, что сам знает наивную физику жидкостей не хуже любого обычного человека, а следовательно, сам был своим единственным информатором и занимался аксиоматической *автоантропологией*<sup>[82]</sup>.

Теперь сравним проект Хейса с философскими проектами по аналитической метафизике, которые часто кажутся мне *наивной* наивной автоантропологией, поскольку складывается впечатление, что участники этих исследований убеждены, что их проект действительно открывает истину, а не просто озвучивает то, что считает истинным конкретный подкласс людей (англоязычные философы со склонностью к аналитической метафизике). В остальном проекты кажутся идентичными: собранные представления, разделяемые разными людьми, проверяются в ходе глубоких размышлений, а после этого результаты складываются в связную “теорию”, основанную на “полученных” принципах, которые в идеале приравниваются к аксиомам. Я спросил нескольких специалистов по аналитической метафизике, могут ли они отличить свой проект от наивной наивной автоантропологии своего клана, но пока не получил ни одного убедительного ответа.

Альтернативой ей служит *продвинутая* наивная антропология (как авто-, так и гетеро-) – то есть антропология, которая не выносит суждений о том, заслуживают ли доверия выведенные в ходе работы теоремы, – и этот проект вполне осуществим и часто полезен. На мой взгляд, именно к нему стоит обратиться специалистам по аналитической метафизике, поскольку он требует лишь минимальной корректировки их методов и лишь одного существенного изменения их *raison d'être*: они должны отказаться от своих претензий и признать, что их исследования лучше всего считать рекогносцировкой ландшафта манифестной картины мира, не предполагающей ни веры, ни неверия, как это происходит, когда антропологи изучают экзотическую культуру (“давайте на время сделаем вид, что аборигены правы, и посмотрим, что из этого выйдет”). Поскольку, на мой взгляд, во многом задача философии состоит в налаживании взаимодействия между **манифестной и научной картинами мира**, философам не помешает проанализировать, с какими народными представлениями они имеют дело, прежде чем приступать к построению и критике своих теорий.

Одна из уникальных черт продвинутой наивной антропологии – ее открытость для парадоксальных открытий. Пока вы занимаетесь наивной антропологией, парадоксальность (для аборигенов) играет против вашей реконструкции, но когда вы переключаетесь и начинаете спрашивать, какие из аспектов наивной “теории” истинны, парадоксальность перестает быть недостатком и даже при случае становится признаком значительного прогресса. В конце концов, парадоксальные результаты приветствуются и в остальной науке.

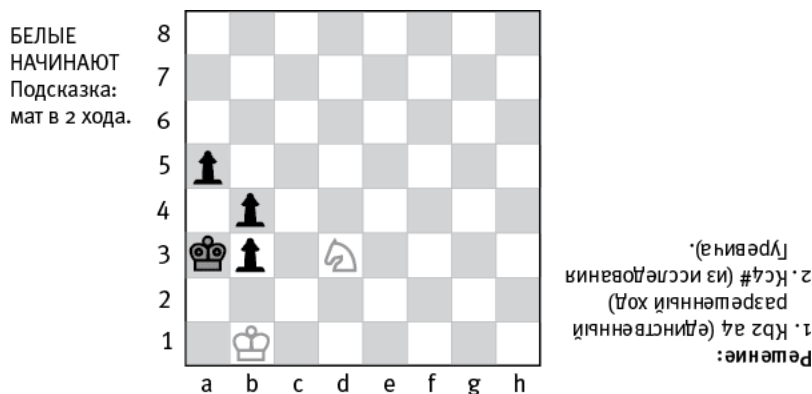
Одна из слабостей автоантропологии состоит в том, что собственные представления человека подвержены искажению под влиянием его теоретических склонностей. Лингвистам прекрасно известно, что они так погружены в свои теории, что не могут быть надежными источниками интуитивных лингвистических представлений. Правда ли можно сказать по-английски: *The boy the man the woman kissed punched ran away* (“Мальчишка, которого ударил мужчина, которого поцеловала женщина, убежал”), – или же мои представления о построении предложений обманывают мое “ухо”? Сырые, неискушенные представления лингвистов обременены таким количеством теории, что лингвисты понимают, что искать

интуитивные лингвистические представления следует у нелингвистов. В последнее время к тому же приходят и философы, проявляющие энтузиазм по отношению к экспериментальной философии (см. Knobe and Nichols 2008). Это направление только зарождается – и первые шаги не слишком впечатляющи, – однако философы хотя бы начинают понимать, что больше нельзя объявлять утверждения очевидно истинными, просто потому что они кажутся совершенно очевидными *им самим*. (В том же ключе Хейс, *возможно*, удивился бы, познакомившись с основными положениями народной физики, если бы провел интервью со случайной выборкой людей, вместо того чтобы считать свой случай показательным.)

Итак, вот какой продвинутой наивной антропологией стоит заняться философам, чтобы провести исследование здравого смысла или манифестной картины мира, прежде чем строить свои теории о знании, правосудии, красоте, истине, нравственности, времени, каузации и подобных вещах, чтобы удостовериться, что они анализируют и аргументируют темы, которые действительно значимы для остального мира – как в научном, так и в обывательском представлении. В результате такого систематического анализа появится своеобразный каталог нереформированного концептуального ландшафта, ставящего задачи теоретикам, – если угодно, метафизика манифестной картины мира. Именно здесь философам и нужно установить соответствие с последними инновациями научной картины мира, а потому наличие подробной карты этого народного ландшафта совсем не повредит. Можно сказать, что это вторая половина реформы, которая превратила философию науки из кабинетной фантазии в серьезное партнерство с настоящей наукой, когда философы науки решили, что им действительно нужно познать современную науку изнутри. Размышляя о наших философских задачах с такой позиции, мы видим, что немалая доля неформального труда, лавирования, приведения контрпримеров и подпитки интуиции на страницах философских журналов представляет собой – в лучшем случае – попытку прийти к приемлемому консенсусу об этой территории.

## 76. Махматные истины высшего порядка

Рассмотрим следующую шахматную задачу<sup>[83]</sup>. Мат белыми в два хода.



Эта задача была недавно опубликована в *Boston Globe* и привлекла мое внимание, поскольку я был уверен, что доказана невозможность поставить мат одиноким конем (и королем, конечно). Я ошибался: как недавно заметил в письме ко мне Дэвид Мусяловски, доказано, что нельзя поставить мат, если на доске остались только король противника и ваши король и конь. Тот факт, что утверждение о *невозможности поставить мат одиноким конем и королем не является шахматной истиной*, представляет собой шахматную истину *высшего порядка*.

Традиционно философия считается априорной дисциплиной наравне с математикой или хотя бы опирается на априорную методологию, и это имеет свои плюсы и минусы. С одной стороны, это позволяет философам не просиживать часами в лаборатории и не проводить полевых исследований, а также не требует от них знакомства с техниками сбора данных, статистическими методами, географией, историей, иностранными языками, эмпирической наукой и прочими областями знания, что дает им достаточно времени для оттачивания своих философских навыков. С другой стороны, как часто отмечается, философию можно создать из чего угодно, и это не всегда хорошо. Для молодых читателей, подумывающих о карьере в этой сфере – а я надеюсь, такие среди вас есть, – эта глава послужит предупреждением, что свобода и абстрактность философии могут быть ее слабостями. Для неспециалистов эта глава также станет путеводителем по особенностям и ловушкам философии.

Рассмотрим в качестве образца априорной истины шахматную истину. Люди играют в шахматы – и это эмпирический факт. Существует также множество других эмпирических фактов о шахматах, о том, как люди играли в них веками, как часто они использовали красивые резные фигуры, расставленные на инкрустированных досках, и так далее. Никакое знание этих эмпирических фактов не играет незаменимой роли в установлении априорных истин о шахматах, которых также немало. Вам необходимо знать лишь правила игры. Существует ровно двадцать разрешенных первых ходов (шестнадцать ходов пешек и четыре хода слонов); королем и одиноким слоном – а также королем и одиноким конем – не поставить мат одинокому королю, и так далее. Устанавливать эти априорные истины о шахматах порой непросто. Задача доказать, что в шахматах возможно, а что невозможно, весьма сложна, и ошибки вполне вероятны. К примеру, несколько лет назад компьютерная шахматная программа обнаружила матовую сеть – гарантированную или вынужденную победу, – состоящую из более чем двухсот ходов без взятий. В результате оказалась опровергнута устоявшаяся шахматная “теорема” и пришлось менять правила игры. Ранее ничья (пат) объявлялась после пятидесяти ходов без взятий с каждой из сторон, но после обнаружения этой длинной непрерывной матовой сети, приводящей к победе, правило о пате после пятидесяти ходов стало безосновательным. (До того как компьютеры начали играть в шахматы, никто и представить себе не мог, что такая длинная серия ходов вообще может привести к гарантированной победе.) Все это весьма интересно, и многие умные люди посвятили себя изучению системы априорных шахматных истин<sup>[84]</sup>.

Некоторые философские исследовательские проекты – или проблемы, если вам угодно, – напоминают попытку установить шахматные истины. Существует – и редко обсуждается – набор общепринятых

правил, а из этих правил выводятся следствия, которые затем формулируются, выставляются на дебаты и корректируются. Пока все понятно. Шахматы – серьезный и важный человеческий артефакт, о котором написано немало ценных трудов. Однако некоторые философские проекты напоминают попытку установить махматные истины. *Махматы* во всем похожи на шахматы, только король в них может двигаться на две, а не на одну клетку в любом направлении. Я только что придумал эту игру, хотя и не сомневаюсь, что кто-нибудь уже успел ее изучить и выяснить, заслуживает ли она внимания. Вероятно нет. Вероятно, ее называют иначе. Я не стал задаваться этими вопросами, потому что на них можно найти ответы, но это не стоит моего времени и сил. По крайней мере, мне так кажется. Априорных истин в махматах столько же, сколько и в шахматах (целая бесконечность), и установить их столь же нелегко. Это значит, что если люди действительно займутся установлением махматных истин, то ошибок им будет не избежать, а эти ошибки нужно будет исправлять, что откроет целую новую область априорных исследований, область махматных истин *высшего порядка*, например:

1. Предложенное Джонсом (1989) доказательство махматной истины *p* несовершенно, поскольку он не учитывает следующей возможности...

2. Утверждение Смита (2002) о несовершенстве доказательства Джонса (1989) предполагает истинность леммы Брауна (1975), которая была недавно поставлена под сомнение Гарфинклем (2002)...

И это не игрушки. При установлении махматных истин высшего порядка можно продемонстрировать выдающиеся способности. Здесь уместно вспомнить афоризм психолога Дональда Хебба:

Если нет смысла браться за работу, нет смысла и делать ее хорошо.

Пожалуй, любой философ сразу вспомнит об идущем в философии споре, участники которого лишились бы работы, если бы неукоснительно следовали завету Хебба, но не стоит и сомневаться, что мы не сумеем принять единогласное решение о том, какие из споров пора прекратить. Вероятно, в нашей широкой дисциплине не найдется такого исследования, которое хотя бы одна школа мысли не считала бы бесполезной растратой гениальности по пустякам. Голосованием ничего решить не получится, а диктаторский подход будет и того хуже, так что пусть цветет вся тысяча цветов. Но не забывайте, что 995 цветов из этой тысячи рано или поздно завянут. Предупреждаю вас: постарайтесь не посвящать ценные годы в начале своей карьеры исследованиям, которые ведутся не слишком долго. Философские поветрия меняются очень быстро, а простое правило не лишено достоверности: чем горячее тема, тем быстрее она выгорает.

Чтобы проверить, не является ли философский проект попыткой установить махматные истины высшего порядка, нужно посмотреть, интересуются ли им специалисты из других областей. Можно ли заставить кого-нибудь, кто не имеет отношения к академической философии, *озаботиться* вопросом, опровергает ли контрпример Джонса принцип Смита? Кроме того, можно попробовать объяснить концепцию непосвященным студентам. Если они ее не “поймут”, вам стоит задуматься, не идете ли вы в артефактную ловушку вслед за замкнутым сообществом экспертов.

Вот как может работать эта ловушка. Философия в некоторой степени противоестественна, и чем умнее вы, тем больше сомневаетесь, понимаете ли вы ее, “правильно ли все делаете”, есть ли у вас талант к этой дисциплине и даже стоит ли ею вообще заниматься. Подающий надежды студент Джонс, *как полагается*, не уверен, стоит ли ему обратиться к философии. Заинтересовавшись лекцией профессора Брауна, Джонс решает попробовать себя в этом деле, пишет статью на горячую тему *H* и получает пятерку у профессора Брауна. “У вас талант, Джонс”, – говорит Браун. Похоже, Джонс только что нашел себе дело жизни. Он начинает изучать правила этой конкретной игры и увлеченно играет в нее с другими молодыми учеными. “Смотри-ка, у нас получается!” – говорят они, подначивая друг друга. Сомнения в допущениях, лежащих в основе теории, подавляются или отбрасываются “чисто теоретически”. Публикуются статьи.

Не рассчитывайте, что одобрение однокурсников *или* любимых преподавателей поможет вам решить этот вопрос. Все они заинтересованы в том, чтобы исследования продолжались. Они знают, что делать, и добиваются успехов. Такая проблема возникает и в других сферах, причем решить ее порой еще сложнее. Бывает, что экспериментаторы, которые овладевают одной техникой и оснащают дорогую лабораторию всем необходимым для проведения конкретных экспериментов, в итоге заполняют пробелы в матрицах данных, уже не вызывающих ни у кого интереса. Что им делать? Выбросить всю дорогую технику? Это может стать проблемой. Переквалификация философов гораздо проще и дешевле. В конце

концов, наша “подготовка”, как правило, не сопряжена с высокими технологиями. В основном мы учимся ориентироваться в различной литературе и овладеваем техниками, которые уже использовались и проверялись ранее. Здесь стоит избегать такой опасности: вы замечаете, что авторитетный ученый опубликовал неаргументированный или сомнительный довод – любопытная, но не безупречная статья профессора Весельчака кажется вам легкой мишенью, идеальной для первой публикации. Дерзайте. Вы заметили это наряду с десятком других читателей статьи, и вам следует быть осторожным, поскольку к тому моменту, когда вы все сошлетесь друг на друга и ответите на ответы, вы станете экспертом по реакциям на незначительное преувеличение Весельчака. (И не забывайте, что если бы Весельчак не переборщил со смелостью своего заявления, он вообще не привлек бы к своей статье достаточного внимания, а к провокациям склонны не только студенты, мечтающие заявить о себе.)

Некоторым людям достаточно найти группу умных единомышленников, чтобы делить с ними “радость открытий, сотрудничества и согласия”, как однажды выразился в опубликованной лекции философ Джон Остин (1961, р. 123). Этим людям неважно, стоит ли изучения та тема, которой они занимаются вместе. Если же интерес к ней проявляет достаточное количество человек, это само по себе становится феноменом, достойным изучения. Философ Бертон Дребен говорил студентам Гарварда: “Философия – это мусор, но история мусора – это наука”. Однако один мусор важнее другого, и тяжело решить, какой именно мусор заслуживает изучения. В другой лекции, опубликованной в той же книге, Остин (1961) выдал такой исполненный презрения перл:

Среди присутствующих на лекции нередко бывают те, кто предпочитает, чтобы вещи имели важность, и теперь специально для них – если таковые есть в этом зале – я сделаю выжимку. [р. 179]

Остин был блестящим философом, но большинство подающих надежды философов, которые вращались рядом с ним и, без сомнения, посмеивались над этой ремаркой, исчезли без следа, а их ну очень умные труды по *философии обыденного языка* (это направление фактически основал Остин) сначала исправно публиковались, а через несколько лет стали намеренно игнорироваться. Такое случалось не раз.

Так что же вам делать? Предложенные мною проверки – посмотреть, можно ли заинтересовать вопросом специалистов из других областей или непосвященных студентов, – дают лишь предупредительные сигналы, но не окончательные ответы. Определенно, существовали – и будут существовать – крайне невразумительные и сложные области философии, исследовать которые стоит, несмотря на то что непосвященные ими не интересуются. Я вовсе не хочу препятствовать исследованиям, которые бросают вызов стандартным представлениям об интересном и важном. Напротив, совершая смелые ходы, вы почти всегда сначала будете сталкиваться с недоверием и презрением, но это не повод опускать руки. Я лишь хочу сказать, что не стоит примыкать к какой-либо стороне просто потому, что вы нашли прекрасных попутчиков, которые считают ваши труды по теме столь же важными, сколь важными вы считаете их работы. Возможно, вы все водите друг друга за нос.

## 77. 10 процентов хорошего

Итак, если **закон Старджона** работает в философии точно так же, как и в любой другой области, что же я считаю хорошим? Прежде всего, классика стала классикой не без причины. Стандартный набор теорий, которые изучаются в рамках истории философии, от Платона до Рассела, стабильно выдерживает проверку временем, а лучшая литература об этих первоисточниках тоже имеет немалую ценность. Самостоятельно читать Аристотеля, Канта и Ницше без всякой подготовки весьма полезно, но гораздо больше вы получите, если обратитесь за поддержкой к тем, кто всю жизнь занимался изучением философии этих мыслителей.

Не все историки философии ставят перед собой одни и те же цели и руководствуются одинаковыми принципами, и лично я не вижу оснований давать отвод хоть кому-то из них. Некоторые считают, что мыслителей следует помещать в исторический контекст, в котором они писали, а это предполагает, к примеру, изучение науки семнадцатого века, если вы действительно хотите понять Декарта, и политической истории семнадцатого и восемнадцатого веков, если вы действительно хотите понять Локка и Юма, а также, конечно, философии их менее прославленных современников. Зачем изучать неудачников? Причина есть. Я не ценил по достоинству многих художников шестнадцатого и семнадцатого века, пока не посетил европейские музеи и не увидел целые залы второсортных картин тех же жанров. Если вы видите только лучшее – так случается во вводных курсах и ведущих музеях, – вам очень сложно понять, насколько оно хорошо. Чем отличается хорошая библиотека от прекрасной? В хорошей библиотеке собраны все хорошие книги. В прекрасной библиотеке собраны все книги. Если вы действительно хотите понять великого философа, вам придется выделить время на изучение идей его не столь великих современников и предшественников, которые остались в тени настоящих мастеров.

Другие специалисты лишь слегка касаются исторического контекста, в котором работали их герои, и вместо этого показывают, как применить их идеи сегодня. В конце концов, Лейбниц написал “Монадологию” не чтобы создать образчик рационализма семнадцатого века, а чтобы *докопаться до истины*. Если уж на то пошло, вы не воспринимаете философа всерьез, пока не зададитесь вопросом, *прав* ли он. Студенты философии – как и профессора – порой забывают об этом и слишком много внимания уделяют навешиванию ярлыков и “выявлению сходств и различий” между теориями, что нередко приходится делать на экзаменах. Бывает, такую цель перед собой ставят целые кафедры философии. Но это не философия, а *оценка философии*. Вот как я помогаю моим студентам отказаться от этой привычки:

Вы раскрыли страшную тайну – скажем, заговор с целью уничтожить статую Свободы или вывести из строя национальную электросеть. Вы изо всех сил стараетесь собрать и упорядочить как можно больше свидетельств, а затем сочиняете письмо, используя все свое красноречие. Вы отправляете копии этого письма в полицию, ФБР, *The New York Times* и на CNN, а в ответ получаете: “Очередная теория заговора – сколько их расплодилось после 11 сентября!” А еще: “Увлекательное чтение – и в некотором роде правдоподобное, а детали просто великолепны”. И наконец: “Напоминает Дона Делилло с элементами Пинчона”. А-а-а! Сосредоточьтесь! Я пытаюсь сказать вам *правду*! Уважайте философа, труд которого вы читаете, спрашивая о каждом предложении и абзаце: “Верю ли я этому, а если нет, то почему?”

Помимо истории философии, внимания заслуживают великолепные труды по философии науки – математики, логики, физики, биологии, психологии, экономики, политологии. Работ по философии химии, астрономии, геологии и инженерии почти не существует, однако есть хорошие исследования о концептуальных проблемах, возникающих в этих областях. Есть также этика. В 1971 г. Джон Ролз опубликовал “Теорию справедливости” – выдающуюся работу, которая открыла плодотворную эпоху для философов, подходящих к традиционным вопросам этики с огидкой на социальные науки, в частности экономику и политологию, но также биологию и психологию. Во многом благодаря Ролзу философы, занимающиеся этикой, вышли на новый уровень и создали множество ценных философских работ, заслуживающих и получающих внимание исследователей других дисциплин, а также политиков и общественных критиков.

Наконец, есть философы, которые далеки от междисциплинарности и лишь косвенно опираются на историю науки, специализируясь на современных проблемах, возникающих в работах других

современных философов. Некоторые из их трудов, как я уже заметил, попадают под правило Хебба: если нет смысла браться за работу, нет смысла и делать ее хорошо. Но другие труды прекрасны и ценны. На страницах этой книги я упоминал немало современных философов, а я не стал бы упоминать о них, если бы не считал их идеи достойными внимания, особенно если я утверждаю, что они совершают ошибку. Я также восхищаюсь работой нескольких десятков других философов, но перечислять их поименно я не буду! Несколько раз в своей карьере я полагался на суждение коллеги, который говорил мне не тратить время на работу *X*, потому что это полная бессмыслица, а позднее оказывалось, что я зря не уделял внимания мыслителю, имеющему ценные идеи, тем самым не позволяя ему вовремя направить мою мысль в нужном направлении из-за неудачного совета. Я прекрасно понимаю, с какой легкостью заинтересованные мыслители могут отказаться от изучения идей философа, который не попадет в мой список, а потому прошу вас считать эту книгу введением в *некоторые* способы философствования. Если же они покажутся вам бесполезными, используйте их в качестве трамплина для начала собственных исследований вопросов и ответов, которые так долго занимают такое множество мыслителей.

## **X.**

### **Используйте инструменты. Старайтесь.**

“Это непостижимо!” – восклицают некоторые, сталкиваясь с “тайной” сознания или, скажем, с утверждением, что жизнь зародилась на этой планете более трех миллиардов лет назад без помощи какого-либо разумного творца. Когда я слышу это, мне так и хочется сказать: “Само собой, *для вас* это непостижимо. Вы отбросили все инструменты мышления и даже не стараетесь ничего постичь”. Вспомните заявление Уильяма Бэтсона о непостижимости материальной основы генов. Сегодня даже школьники без особого труда постигают ДНК – и дело не в том, что они умнее Бэтсона. Дело в том, что за последнее столетие мы создали и доработали инструменты мышления, которые существенно облегчают эту задачу. Само собой, есть люди, которые просто не хотят постигать такие вещи. Они стараются защитить все тайны от любой попытки объяснения, боясь, что это объяснение уничтожит их прелесть.

Проявляя любознательность, другие люди выясняют, что фраза “пути Господни неисповедимы” – прекрасный антиинструмент мышления. Надеясь на то, что вопрошающий высокомерен и несдержан, она мгновенно подавляет любое любопытство. Когда-то она прекрасно работала и по-прежнему работает в сообществах, где незнание науки простительно или даже предпочтительно. Я полагаю, что нам пора перестать считать это “богоугодное” наблюдение мудростью и признать его откровенно оборонительной пропагандой, какой оно и является. Конструктивный ответ на него может быть таким: “Прекрасно! Я люблю загадки! Может, сумеем решить и эту? Есть идеи?”

Постигать новое непросто – для этого недостаточно быстро ознакомиться с идеей и тотчас расписаться в ее правильности. То, что кажется нам непостижимым, вполне может стать довольно очевидным, если мы поработаем подольше. Когда мы с уверенностью называем некоторые вещи поистине невозможными – самое большое простое число, плоский треугольник, сумма углов которого составляет более ста восьмидесяти градусов, или женатого холостяка, – дело не в том, что мы не можем их постичь, а в том, что мы постигли их компоненты так хорошо, что невозможность их соединения стала для нас очевидной.

Пока мы не сумели полностью постичь, как смысл существует в материальном мире, как зародилась и развивалась жизнь, как работает сознание и наделены ли мы свободой воли, но мы делаем успехи: вопросы, которые мы задаем и обдумываем сегодня, гораздо лучше вопросов прошлых лет. Ответы на них не за горами.



## XI.

### Что осталось за кадром

Некоторые читатели ранних версий этой книги удивились и расстроились, не обнаружив в ней ряда самых известных из моих насосов интуиции. На этих страницах не описано несколько десятков насосов, включая некоторые из моих любимых. В нескольких случаях это заслуживает объяснения.

Пожалуй, самым известным из моих насосов интуиции можно признать мысленный эксперимент “Где я?”», но именно из-за его известности я и не стал включать его в эту книгу. Он впервые был описан в книге “Мозговые штурмы” в 1978 г., а в 1981 г. был включен в книгу “Глаз разума”. Эти работы были переведены на десяток языков, а сам эксперимент входил во многие антологии. В фильме “Жертва мозга” (1988) содержится получасовая инсценировка эксперимента (и сам я играю тело будущего Деннета). Небольшая сценка также включена в документальный фильм BBC о сознании и мозге, выпущенный в 1981 г. В 1984 г. эксперимент также был инсценирован куклами яванского театра теней под управлением прославленного кукольника Линн Джеффрис в Гарвардском *Loeb Theater*. Google не оставляет сомнений, что найти описание этого эксперимента не представляет труда – вместе со множеством комментариев.

“Баллада о пиццерии Шейки” (см. Dennett 1982a) неплохо справляется с развенчанием некоторых представлений о “понимании *de re* и *de dicto*”, которые некогда господствовали среди философов, работающих над интенциональностью, но более не были знакомы никому. Если бы я включил ее в эту книгу, мне пришлось бы сначала расписать любопытную, но неверную логику мышления, чтобы затем искоренить ее при помощи моего насоса интуиции. Некоторые философы должны прекрасно в этом разбираться, но остальные могут без потерь оставаться в блаженном неведении.

В разделе об эволюции я неохотно оставил за кадром мой любимый новый насос интуиции “Дарвинистские пространства” Питера Годфри-Смита – лучшее, на мой взгляд, применение многомерного пространства в качестве инструмента мышления в философии, – потому что для его эффективной работы потребовалось бы слишком подробное описание эволюционной теории и биологических феноменов. Я объясняю его техническим языком для других философов биологии и привожу прекрасные примеры его использования в своей рецензии “Гомункулы у руля” (2010) на книгу Годфри-Смита “Популяционное мышление и естественный отбор” (2009). См. также ответ Годфри-Смита (2010).

В моем эссе “Куайнирование квалиа” (1988a) я описываю целых четырнадцать насосов интуиции, призванных прояснить и затем опровергнуть безнадежно запутанную философскую концепцию квалиа. В эту книгу вошел лишь один из них, “**Проклятие цветной капусты**”. Он помогает мне ввести понятие и познакомить читателей с серьезной проблемой квалиа. Я рекомендую прочитать “Куайнирование квалиа” всем тем, кто по-прежнему считает концепцию квалиа (как ее называют философы) состоятельной. Эссе не раз включалось в разные антологии и доступно в интернете на нескольких языках. В моей книге “Сладкие сны” (2005b) содержатся другие аргументы и насосы интуиции по теме. Среди других насосов интуиции о сознании можно выделить “амнестический кураре” в эссе “Почему нельзя создать компьютер, чувствующий боль” (Dennett 1978c), “Болотная Мэри и робот Мэри” в эссе “Что знает робот Мэри” (Dennett 2007d), а также “оруэлловскую и сталинскую” модели сознания в книге “Объясненное сознание” (Dennett 1991a). Все они требуют более тщательной подготовки мизансцены, чем я мог себе позволить, ибо мне не хотелось, чтобы эта книга не стала слишком длинной.

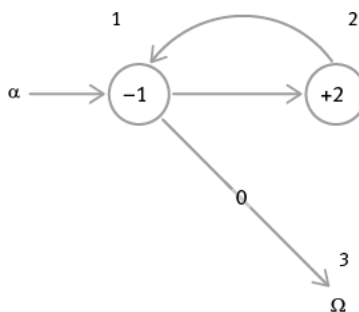
Я также не включил сюда насосы интуиции для мышления о религии, которые предложил в эссе “Рассеивая чары” (Dennett 2006a), и мой пример с Суперменом, подправляющим сланцы Бёрджесс, приведенный в книге “Наука и религия. Совместимы ли они?” (Dennett, Plantinga 2011).

## Приложение

### Решения задач о регистровых машинах

#### упражнение 1

программа 1:



а. Сколько шагов потребуется регистровой машине, чтобы сложить  $2 + 5$  и получить  $7$ , выполняя программу 1 (считая Кон отдельным шагом)?

**Ответ:** шесть шагов. Три декремента, два инкремента, одно окончание программы (последний декремент – переход на ноль).

б. Сколько шагов потребуется машине, чтобы сложить  $5 + 2$ ?

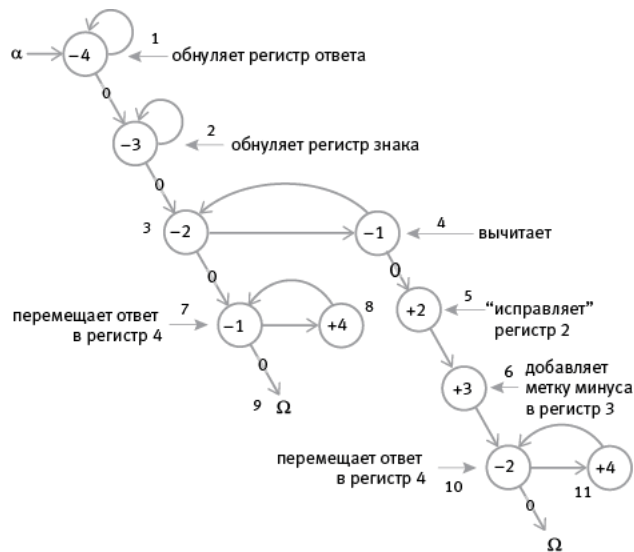
**Ответ:** двенадцать шагов. Шесть декрементов, пять инкрементов и одно окончание программы.

1(Какой из этого можно сделать вывод?)

**Ответ:** порядок содержимого может иметь большое значение, так что вам может показаться необходимым ввести правило всегда помещать меньшее число в регистр 1, но, если для этого вам сначала придется сравнивать два числа, чтобы выяснить, какое из них меньше, в итоге вам потребуется больше шагов, чем для сложения!

#### упражнение 2

а. Напишите РПА-программу для этого графа потока. (Обратите внимание: поскольку программа разветвляется, вы можете пронумеровать шаги несколькими способами. Неважно, какой из них вы выберете: главное, чтобы на верные следующие шаги указывали команды безусловного перехода.)



| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ [ПЕРЕХОД НА ШАГ] |    |
|-----|------------|---------|--------------------------------|----|
| 1.  | Деп        | 4       | 1                              | 2  |
| 2.  | Деп        | 3       | 2                              | 3  |
| 3.  | Деп        | 2       | 4                              | 10 |
| 4.  | Деп        | 1       | 3                              | 5  |
| 5.  | Инк        | 2       | 6                              |    |
| 6.  | Инк        | 3       | 7                              |    |
| 7.  | Деп        | 1       | 8                              | 9  |
| 8.  | Инк        | 4       | 7                              |    |
| 9.  | Кон        |         |                                |    |
| 10. | Деп        | 1       | 11                             | 9  |
| 11. | Инк        |         | 4                              | 10 |

б. Что происходит, когда программа пытается вычесть 3 из 3 или 4 из 4?

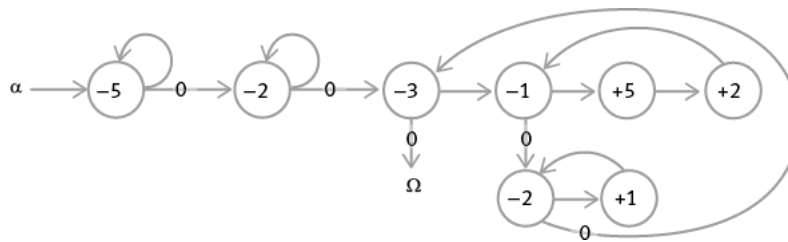
**Программа останавливается с нулем в регистре 4.**

в. Какая возможная ошибка предотвращается обнулением регистра 3 перед попыткой вычитания на шаге 3 вместо шага 4?

Если на старте и в регистре 1, и в регистре 2 были нули, в конце программа могла выдать абсурдный ответ (либо – 0, либо число, отличное от 0 и 1, в регистре знака).

### упражнение 3

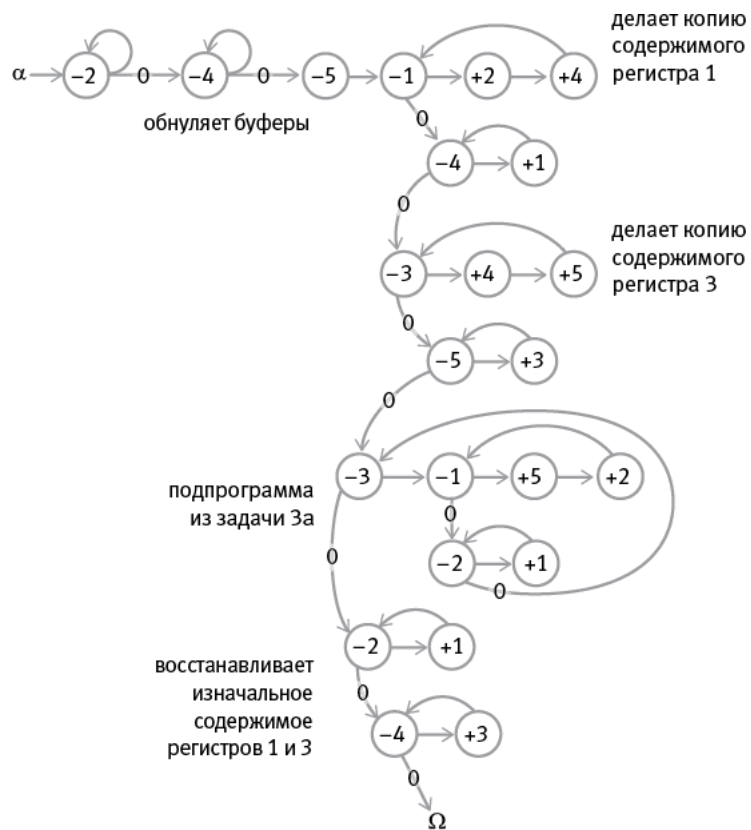
а. Нарисуйте граф потока (и напишите РПА-программу) для умножения содержимого регистра 1 на содержимое регистра 3, поместив ответ в регистр 5.



б. (По желанию) Используя копирование и перемещение, улучшите программу умножения, созданную в задаче а: когда она закончит работу, изначальное содержимое регистра 1 и регистра 3 восстановится, так что вы сможете легко проверить исходные данные и ответы на правильность по завершении программы.

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ]    |
|-----|------------|---------|---------------|---------------------|
| 1.  | Деп        | 5       | 1             | 2 [обнуляет буферы] |
| 2.  | Деп        | 2       | 2             | 3                   |
| 3.  | Деп        | 3       | 4             | 9                   |
| 4.  | Деп        | 1       | 5             | 7                   |
| 5.  | Инк        | 5       | 6             |                     |
| 6.  | Инк        | 2       | 4             |                     |
| 7.  | Деп        | 2       | 8             |                     |
| 8.  | Инк        | 1       | 7             |                     |
| 9.  | Кон        |         |               |                     |

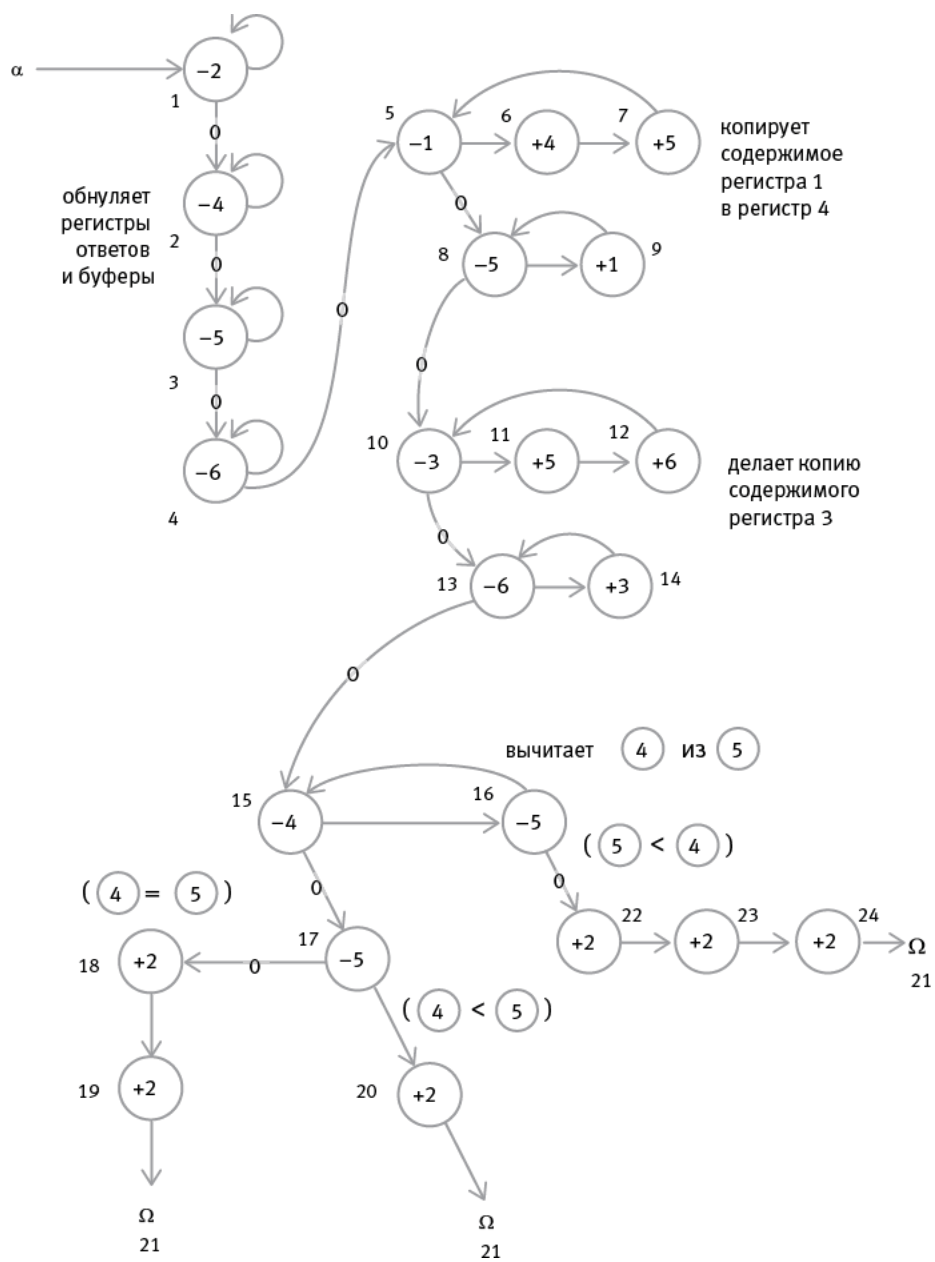
| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ]      |
|-----|------------|---------|---------------|-----------------------|
| 1.  | Деп        | 2       | 1             | 2 [обнуляет 3 буфера] |
| 2.  | Деп        | 4       | 2             | 3                     |
| 3.  | Деп        | 5       | 3             | 4                     |



| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ]                         |
|-----|------------|---------|---------------|------------------------------------------|
| 4.  | <i>Деп</i> | 1       | 5             | 7 [делает копию содержимого регистра 1]  |
| 5.  | <i>Инк</i> | 2       | 6             |                                          |
| 6.  | <i>Инк</i> | 4       | 4             |                                          |
| 7.  | <i>Деп</i> | 4       | 8             | 9                                        |
| 8.  | <i>Инк</i> | 1       | 7             |                                          |
| 9.  | <i>Деп</i> | 3       | 10            | 12 [делает копию содержимого регистра 3] |

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ | [ПЕРЕХОД НА ШАГ]                                            |
|-----|------------|---------|---------------|-------------------------------------------------------------|
| 10. | Инк        | 4       | 11            |                                                             |
| 11. | Инк        | 5       | 9             |                                                             |
| 12. | Деп        | 5       | 13            | 14                                                          |
| 13. | Инк        | 3       | 12            |                                                             |
| 14. | Деп        | 3       | 15            | 20 [это код из упражнения 3а с другими номерами строк]      |
| 15. | Деп        | 1       | 16            | 18                                                          |
| 16. | Инк        | 5       | 17            |                                                             |
| 17. | Инк        | 2       | 15            |                                                             |
| 18. | Деп        | 2       | 19            | 15                                                          |
| 19. | Инк        | 1       | 18            |                                                             |
| 20. | Деп        | 2       | 21            | 22 [восстанавливает изначальное содержимое регистров 1 и 3] |
| 21. | Инк        | 1       | 20            |                                                             |
| 22. | Деп        | 4       | 23            | 24                                                          |
| 23. | Инк        | 3       | 22            |                                                             |
| 24. | Кон        |         |               |                                                             |

в. (По желанию) Нарисуйте граф потока и напишите РПА-программу, которая изучает содержимое регистра 1 и регистра 3 (не разрушая их!) и записывает адрес (1 или 3) регистра с большим содержимым в регистр 2 или помещает 2 в регистр 2, если содержимое регистров 1 и 3 равно. (После выполнения этой программы содержимое регистра 1 и регистра 3 должно остаться неизменным, а регистр 2 должен показывать, равно ли их содержимое, а если нет, то в каком из регистров содержимое больше.)



| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ                                 | [ПЕРЕХОД НА ШАГ]                                                             |
|-----|------------|---------|-----------------------------------------------|------------------------------------------------------------------------------|
| 1.  | <i>Деп</i> | 2       | 1                                             | 2 [обнуляет регистр ответа]                                                  |
| 2.  | <i>Деп</i> | 4       | 2                                             | 3 [обнуляет буферы]                                                          |
| 3.  | <i>Деп</i> | 5       | 3                                             | 4                                                                            |
| 4.  | <i>Деп</i> | 6       | 4                                             | 5                                                                            |
| 5.  | <i>Деп</i> | 1       | 6                                             | 8 [копирует содержимое регистра 1 в регистр 4]                               |
| 6.  | <i>Инк</i> | 4       | 7                                             |                                                                              |
| 7.  | <i>Инк</i> | 5       | 5                                             |                                                                              |
| 8.  | <i>Деп</i> | 5       | 9                                             | 10                                                                           |
| 9.  | <i>Инк</i> | 1       | 8                                             |                                                                              |
| 10. | <i>Деп</i> | 3       | 11                                            | 13 [копирует содержимое регистра 3 в регистр 5]                              |
| 11. | <i>Инк</i> | 5       | 12                                            |                                                                              |
| 12. | <i>Инк</i> | 6       | 10                                            |                                                                              |
| 13. | <i>Деп</i> | 6       | 14                                            | 15                                                                           |
| 14. | <i>Инк</i> | 3       | 13                                            |                                                                              |
| 15. | <i>Деп</i> | 4       | 16                                            | 17 [вычитает содержимое регистра 4 из содержимого регистра 5]                |
| 16. | <i>Деп</i> | 5       | 15                                            | 22                                                                           |
| 17. | <i>Деп</i> | 5       | 20                                            | 18 [проверяет, не равняется ли содержимое регистра 4 содержимому регистра 5] |
| 18. | <i>Инк</i> | 2       | 19 [содержимое равно; помещает 2 в регистр 2] |                                                                              |
| 19. | <i>Инк</i> | 2       | 21                                            |                                                                              |



| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ                                                                    | [ПЕРЕХОД НА ШАГ] |
|-----|------------|---------|----------------------------------------------------------------------------------|------------------|
| 20. | <i>Инк</i> | 2       | 21 [содержимое регистра 4 меньше содержимого регистра 5; помещает 1 в регистр 2] |                  |
| 21. | <i>Кон</i> |         |                                                                                  |                  |
| 22. | <i>Инк</i> | 2       | 23 [содержимое регистра 5 меньше содержимого регистра 4; помещает 3 в регистр 2] |                  |
| 23. | <i>Инк</i> | 2       | 24                                                                               |                  |
| 24. | <i>Инк</i> | 2       | 21                                                                               |                  |

**упражнение 4 (по желанию)**

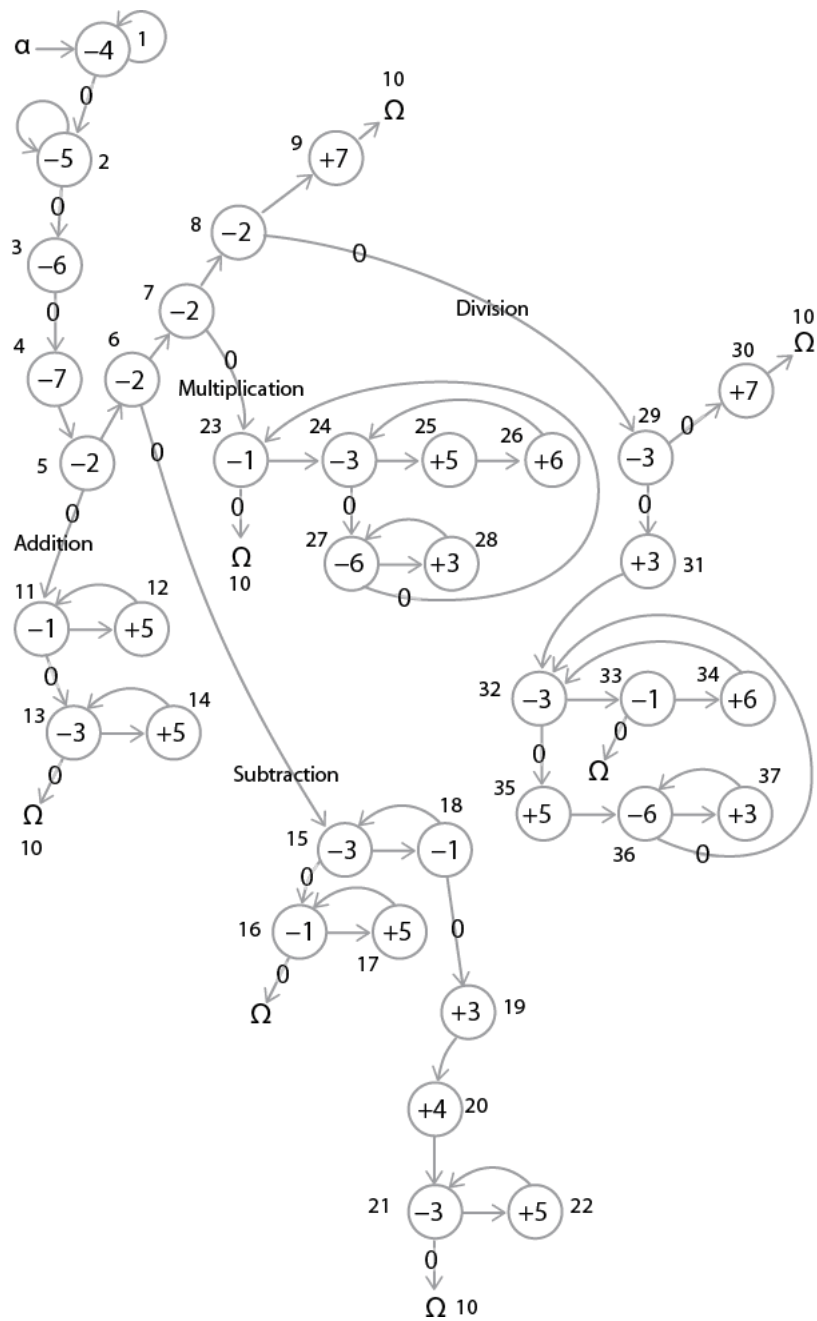
Нарисуйте граф потока и напишите РПА-программу, которая превращает регистровую машину в простой карманный калькулятор, следующим образом:

а. Используйте регистр 2 для операции:

0 = *ADD*  
 1 = *SUBTRACT*  
 2 = *MULTIPLY*  
 3 = *DIVIDE*

б. Поместите числа, с которыми будут производиться манипуляции, в регистры 1 и 3.

(Таким образом, 3 0 6 будет означать  $3 + 6$ ; 5 1 3 будет означать  $5 - 3$ ; 4 2 5 будет означать  $4 * 5$ ; а 9 3 3 будет означать  $9 \div 3$ .) Затем поместите результаты операции в регистры 4–7, используя регистр 4 для знака (где 0 означает +, а 1 означает –), регистр 5 для численного ответа, регистр 6 для возможного остатка в случае деления, а регистр 7 для сообщения об ошибке ввода (либо требовании делить на ноль, либо неопределенной операции в регистре 2).



| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ                                                                                                | [ПЕРЕХОД НА ШАГ]                            |
|-----|------------|---------|--------------------------------------------------------------------------------------------------------------|---------------------------------------------|
| 1.  | <i>Деп</i> | 4       | 1                                                                                                            | 2 [обнуляет<br>регистр ответа]              |
| 2.  | <i>Деп</i> | 5       | 2                                                                                                            | 3                                           |
| 3.  | <i>Деп</i> | 6       | 3                                                                                                            | 4                                           |
| 4.  | <i>Деп</i> | 7       | 4                                                                                                            | 5                                           |
| 5.  | <i>Деп</i> | 2       | 6                                                                                                            | 11 [осуществляет<br>переход<br>к сложению]  |
| 6.  | <i>Деп</i> | 2       | 7                                                                                                            | 15 [осуществляет<br>переход<br>к вычитанию] |
| 7.  | <i>Деп</i> | 2       | 8                                                                                                            | 23 [осуществляет<br>переход<br>к умножению] |
| 8.  | <i>Деп</i> | 2       | 9                                                                                                            | 28 [осуществляет<br>переход к делению]      |
| 9.  | <i>Инк</i> | 7       | 10 [замечает код<br>некорректной<br>операции,<br>помещенный<br>в регистр 2,<br>и останавливает<br>программу] |                                             |
| 10. | <i>Кон</i> |         |                                                                                                              |                                             |
| 11. | <i>Деп</i> | 1       | 12                                                                                                           | 13 [начинает<br>сложение]                   |
| 12. | <i>Инк</i> | 5       | 11                                                                                                           |                                             |
| 13. | <i>Деп</i> | 3       | 14                                                                                                           | 10                                          |

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР                                    | СЛЕДУЮЩИЙ ШАГ                                                              | [ПЕРЕХОД НА ШАГ]                                          |
|-----|------------|--------------------------------------------|----------------------------------------------------------------------------|-----------------------------------------------------------|
| 14. | <i>Инк</i> | 5                                          | 13                                                                         |                                                           |
| 15. | <i>Деп</i> | 3                                          | 18                                                                         | 16                                                        |
| 16. | <i>Деп</i> | 1                                          | 17                                                                         | 10 [начинает<br>и заканчивает<br>вычитание]               |
| 17. | <i>Инк</i> | 5                                          | 16                                                                         |                                                           |
| 18. | <i>Деп</i> | 1                                          | 15                                                                         | 19                                                        |
| 19. | <i>Инк</i> | 3 [восста-<br>навливает<br>регистр 3]      |                                                                            |                                                           |
| 20. | <i>Инк</i> | 4 [помещает<br>знак минуса<br>в регистр 4] |                                                                            |                                                           |
| 21. | <i>Деп</i> | 3                                          | 22                                                                         | 10                                                        |
| 22. | <i>Инк</i> | 5                                          | 21                                                                         |                                                           |
| 23. | <i>Деп</i> | 1                                          | 24                                                                         | 10 [начинает<br>и заканчивает<br>умножение]               |
| 24. | <i>Деп</i> | 3                                          | 24                                                                         | 27                                                        |
| 25. | <i>Инк</i> | 5                                          | 26                                                                         |                                                           |
| 26. | <i>Инк</i> | 6                                          | 24                                                                         |                                                           |
| 27. | <i>Деп</i> | 6                                          | 28                                                                         | 23                                                        |
| 28. | <i>Инк</i> | 3                                          | 27                                                                         |                                                           |
| 29. | <i>Деп</i> | 3                                          | 31                                                                         | 30 [начинает<br>деление, проверяет<br>на деление на ноль] |
| 30. | <i>Инк</i> | 7                                          | 10 [замечает<br>ошибку деления<br>на ноль<br>и останавливает<br>программу] |                                                           |

| ШАГ | ИНСТРУКЦИЯ | РЕГИСТР | СЛЕДУЮЩИЙ ШАГ [ПЕРЕХОД НА ШАГ]    |                             |
|-----|------------|---------|-----------------------------------|-----------------------------|
| 31. | <i>Инк</i> | 3       | 32 [восстанавливает<br>регистр 3] |                             |
| 32. | <i>Деп</i> | 3       | 33                                | 35                          |
| 33. | <i>Деп</i> | 1       | 34                                | 10 [заканчивает<br>деление] |
| 34. | <i>Инк</i> | 6       | 32                                |                             |
| 35. | <i>Инк</i> | 5       | 36                                |                             |
| 36. | <i>Деп</i> | 6       | 37                                | 32                          |
| 37. | <i>Инк</i> | 3       | 36                                |                             |

## Источники

### I. Введение

Мысленные эксперименты об эксцентричном тюремщике и драгоценностях в урне взяты с переработками из *Elbow Room* (Dennett 1984a).

### II. Дюжина универсальных инструментов мышления

1. Глава “Совершая ошибки” сначала была статьей “*How to Make Mistakes*” (Dennett 1995b).
2. Глава “На основе пародии” ранее не публиковалась.
3. Глава “Правила Рапопорта”. В печатном виде я впервые перечислил эти правила в своих рассуждениях о книге Докинза “Бог как иллюзия” в работе *Free Inquiry* (Dennett 2007a).
4. Глава “Закон Старджона”. Закон обсуждался в статье “*Holding a Mirror up to Dupré*” (Dennett 2004).
5. Глава “Бритва Оккама” ранее не публиковалась.
6. Глава “Метла Оккама”. На многих сайтах создание этого термина приписывается Сиднею Бреннеру. Судя по всему, он не публиковал ничего на эту тему.
7. Глава “Использование обывателей в качестве ложной аудитории” ранее не публиковалась.
8. Глава “Выпрыгивание” включает материал из статьи “*I Could Not Have Done Otherwise – So What?*” (Dennett 1984b).
9. Глава “Три вида гулдинга” основана на работах “*Confusion over Evolution: An Exchange*” (Dennett 1993) и “*Shall We Tango? No, But Thanks for Asking*” (Dennett 2011b).
10. Глава “Оператор «безусловно»” выросла из ремарок в работе “*Get Real*” (Dennett 1994a).
11. Глава “Риторические вопросы” ранее не публиковалась.
12. Глава “Что такое плубокость?” основана на материале из работы “*With a Little Help from My Friends*” (Dennett 2000).

### III. Инструменты мышления о значении или содержании

13. Глава “Убийство на Трафальгарской площади” взята из работы “*Three Kinds of Intentional Psychology*” (Dennett 1981).
14. Глава “Старший брат из Кливленда” взята из работы “*Brain Writing and Mind Reading*” (Dennett 1975).
15. Глава “Мой папа – врач” взята из книги *Content and Consciousness* (Dennett 1969).
16. Глава “Манифестная и научная картина мира” содержит материал из работ “*Expecting Ourselves to Expect*” (Dennett 2013b); “*Sakes and Dints*” (Dennett 2012) и “*Kinds of Things*” (Dennett 2013c). Дискуссия о муравьях и птицах взята из книги *Elbow Room* (Dennett 1984a).
17. Глава “Народная психология” взята из работы “*Three Kinds of Intentional Psychology*” (Dennett 1981).
18. Глава “Интенциональная установка” взята из работы “*Intentional Systems*” (Dennett 1971) и книги *The Intentional Stance* (Dennett 1987), а также содержит материалы из множества других книг и статей.
19. Глава “Личностное и субличностное” взята из книги *Content and Consciousness* (Dennett 1969).
20. Глава “Каскад гомункулов” основана на материалах из книги *Brainstorms* (Dennett 1978) и статьи на Edge.org (Dennett 2008), впоследствии опубликованной в сборнике *What Have You Changed Your Mind About* (Brockman 2009).
21. Глава “Оператор «вроде как»” ранее не публиковалась, но затрагиваемые в ней темы развиваются в работе “*Turing’s ‘Strange Inversion of Reason’*” (Dennett 2013e).
22. Глава “Чудо-ткань”. Термин был впервые использован в печати в статье “*Natural Freedom*” (Dennett 2005a). Спасибо Кристофу Коху, который обратил мое внимание на цитату из Уильяма Бэтсона.
23. Глава “В диспетчерской работа” взята из статьи “*Current Issues in the Philosophy of Mind*” (Dennett 1978b).

#### IV. Интерлюдия о компьютерах

24. Глава “Семь секретов силы компьютеров” основана на материалах вводного курса информатики, который мы с Джорджем Смитом и Дэвидом Айлзом преподавали в Университете Тафтса, в чем нам помогал студент-программист Стив Барни. Стив стал ведущим программистом Мастерской учебных программ Университета Тафтса – проекта, который был профинансирован Фондом Слоуна, после того как его представители увидели виртуальный компьютер AESOP, созданный Стивом для нашего курса. AESOP также послужил примером для создания машины *RodRego*, которую я спрограммировал на *Logo* в 1986 г. и назвал *Rego*, после чего она была усовершенствована и улучшена Родом Да Силвой (получив название *RodRego*), а затем обновлена Николаем Швертнером в Мастерской учебных программ.

25. Глава “Виртуальные машины”. Я обсуждал эти машины во многих своих статьях, включая “*Notes on Prosthetic Imagination*” (Dennett 1982d) (манифест, который привел к основанию Мастерской учебных программ), *Consciousness Explained* (Dennett 1991a) и “*The Practical Requirements for Making a Conscious Robot*” (Dennett 1994b).

26. Глава “Алгоритмы” основана на материалах из книги *Darwin’s Dangerous Idea* (Dennett 1995a).

27. Глава “Автоматизация лифта” впервые была представлена в рамках лекции “*Varieties of Content*” на симпозиуме в Копенгагенском университете в 2007 г. (Dennett 2007c). Она ранее не публиковалась.

В резюме содержатся идеи из работы “*Turing’s ‘Strange Inversion of Reasoning’*” (Dennett 2013e).

#### V. Дополнительные инструменты мышления о значении

28. Глава “Пунктик насчет рыжих” взята из работы “*Things about Things*” (Dennett 2001e).

29. Глава “Странствующий четвертачник, Земля-Двойник и гигантский робот” взята с изменениями из работы “*Evolution, Error and Intentionality*” (Dennett 1988c).

30. Глава “Радикальный перевод и кроссворд Куайна” была впервые опубликована в работе “*With a Little Help from My Friends*” (Dennett 2000).

31. Глава “Семантические и синтаксические движки” взята с изменениями из работы “*Three Kinds of Intentional Psychology*” (Dennett 1981).

32. Глава “Болотный человек и королева” содержит материал, взятый с изменениями из работ “*Features of Intentional Action*” (Dennett 1968) и “*Get Real*” (Dennett 1994a).

33. Глава “Два черных ящика” взята с изменениями из книги *Darwin’s Dangerous Idea* (Dennett 1995a). Резюме ранее не публиковалось.

#### VI. Инструменты мышления об эволюции

34. Глава “Универсальная кислота” взята с изменениями из книги *Darwin’s Dangerous Idea* (Dennett 1995a).

35. Глава “Библиотека Менделя: чрезвычайно большое и исчезающе малое” взята с изменениями из книги *Darwin’s Dangerous Idea* (Dennett 1995a).

36. Глава “Тены как слова и как подпрограммы” ранее не публиковалась. На ее создание меня вдохновила книга Ричарда Докинза “Рассказ предка” (2004), длинная цитата из которой вошла в текст.

37. Глава “Древо жизни” ранее не публиковалась. Рисунок создан Леонардом Эйзенбергом. На его сайте <http://evogeneao.com/tree.html> можно найти подробную инструкцию по использованию диаграммы, а также заказать плакаты, футболки, кофты и другие вещи.

38. Глава “Подъем в проектное пространство: небесные крюки и подъемные краны” взята с изменениями из книги *Darwin’s Dangerous Idea* (Dennett 1995a).

39. Глава “Компетентность без понимания” взята с изменениями из работы “*Darwin’s ‘Strange Inversion of Reasoning’*” (Dennett 2009a).

40. Глава “Плавающие основания”. Плавающие основания были впервые представлены в работе “*Intentional Systems in Cognitive Ethology: The ‘Panglossian Paradigm’ Defended*” (Dennett 1983) и обсуждались во множестве других статей.

41. Глава “Понимают ли цикады простые числа?” взята с изменениями из книги *Darwin’s Dangerous Idea* (Dennett 1995a).

42. Глава “Как объяснить смотровые прыжки” ранее не публиковалась.
43. Глава “Опасайтесь первого млекопитающего” взята с изменениями из книги *Darwin's Dangerous Idea* (Dennett 1995a).
44. Глава “Когда происходит видообразование?” взята с изменениями из работы “*The Multiple Drafts Model*” (Dennett, Akins 2008).
45. Глава “Вдоводелы, Митохондриальная Ева и ретроспективные коронации” взята с изменениями из книги *Darwin's Dangerous Idea* (Dennett 1995a).
46. Глава “Циклы” взята с изменениями из работы “*Cycles*” (Dennett 2011a), в которой я отвечаю на заданный Edge в 2011 г. вопрос “Какая научная концепция обогатила бы любой набор когнитивных инструментов?”. Впоследствии работа была опубликована в сборнике *This Will Make You Smarter* (Brockman 2012).
47. Глава “Что именно глаз лягушки сообщает ее мозгу?” взята с изменениями из книги *Darwin's Dangerous Idea* (Dennett 1995a).
48. Глава “Скачки в пространстве Вавилонской библиотеки” взята с изменениями из книги *Darwin's Dangerous Idea* (Dennett 1995a).
49. Глава “Кто автор «Спамлета»?” взята с изменениями из моего президентского послания Восточному отделению Американской философской ассоциации, озаглавленного “*In Darwin's Wake, Where Am I?*” (Dennett 2001d).
50. Глава “Шум в виртуальном отеле” взята с изменениями из работы “*Collision-Detection, Muselot, and Scribble: Some Reflections on Creativity*” (Dennett 2001a).
51. Глава “Херб, Элис и малыш Хэл” взята с изменениями из работы “*Homunculi Rule: Reflections on Darwinian Populations and Natural Selection by Peter Godfrey-Smith, Oxford University Press, 2009*” (Dennett 2010).
52. Глава “Мемы” основана на ряде публикаций о мемах, перечисленных в самой главе.

## VII. Инструменты мышления о сознании

53. Глава “Два контробраза” содержит материал из книг *Consciousness Explained* (Dennett 1991a) и *Sweet Dreams* (Dennett 2005b).
54. Глава “Чутье на зомби”. Чутье на зомби было описано в книге *Sweet Dreams* (Dennett 2005b).
55. Глава “Зомби и зимбо” основана на материале из книги *Consciousness Explained* (Dennett 1991a).
56. Глава “Проклятие цветной капусты” основана на материале статьи “*Quining Qualia*” (Dennett 1988a).
57. Глава “Жизненная сила: сколько это в «настоящих деньгах»?” взята с изменениями из статьи “*Consciousness: 'How Much Is That in Real Money?'*” (Dennett 2001b).
58. Глава “Печальная история мистера Клапгра” взята с изменениями из книги *Sweet Dreams* (Dennett 2005b).
59. Глава “Настроенная колода” взята с изменениями из работы “*Explaining the 'Magic' of Consciousness*” (Dennett 2001c).
60. Глава “Китайская комната” основана на материале из книг *The Mind's I* (Hofstadter, Dennett 1981) и *The Intentional Stance* (Dennett 1987).
61. Глава “Падение телеклона с Марса на Землю” взята с изменениями из книги *The Mind's I* (Hofstadter, Dennett 1981).
62. Глава “Самость как центр нарративной гравитации”. Этот инструмент был впервые описан в статье “*Why Everyone Is a Novelist*” (Dennett 1988b).
63. Глава “Гетерофеноменология”. Гетерофеноменология была впервые описана в статье “*How to Study Consciousness Empirically: or Nothing Comes to Mind*” (Dennett 1982b). Подробнее тема была раскрыта в одной из глав книги *Consciousness Explained* (Dennett 1991a). Гетерофеноменологии посвящен специальный выпуск журнала *Phenomenology and Cognitive Science* (vol. 6, nos. 1 and 2), приглашенным редактором которого выступил Алва Ноз. В этом выпуске содержится большинство лучших работ на эту тему, а также мой ответ “*Heterophenomenology Reconsidered*” (Dennett 2007b).
64. Глава “Исследовательница цвета Мария: разоблачение упорного башмака” основана на материалах из книги *Consciousness Explained* (Dennett 1991a) и статьи “*What RoboMary Knows*” (Dennett 2007d), которая опубликована во второй из двух антологий, посвященных “Мариелогии”. Первая антология: Ludlow, Nagasawa, and Stoljar 2004.



### **VIII. Инструменты мышления о свободе воли**

65. Глава “Подлый нейрохирург” основана на моей Эразмовской лекции “*Erasmus: Sometimes a Spin-Doctor Is Right*”, прочитанной в 2012 г.

66. Глава “Детерминистская модель: игра «Жизнь»”. Мартин Гарднер познакомил широкую аудиторию с игрой “Жизнь”, посвятив ей две колонки “*Mathematical Games*” в журнале *Scientific American* в октябре 1970 и феврале 1971 г. Паундстоун (1985) великолепно разбирает игру и ее философские следствия. О ней также написана неплохая статья в “Википедии”, где содержится множество сносок. Я часто использовал игру “Жизнь” в своих работах. Эта глава взята с изменениями из книги *Freedom Evolves* (Dennett 2003).

67. Глава “Камень, ножницы, бумага” ранее не публиковалась.

68. Глава “Две лотереи” взята с изменениями из книги *Elbow Room* (Dennett 1984a).

69. Глава “Инертные исторические факты” взята с изменениями из статьи “*Get Real*” (Dennett 1994a).

70. Глава “Компьютерный шахматный марафон” взята из книги *Freedom Evolves* (Dennett 2003).

71. Глава “Конечная ответственность” ранее не публиковалась.

72. Глава “Сфексовость” ранее не публиковалась.

73. Глава “Мальчики из Бразилии: еще один упорный башмак” ранее не публиковалась.

### **IX. Что значит быть философом?**

74. Глава “Сделка с дьяволом” взята с изменениями из моей рецензии на книгу Томаса Нагеля *Other Minds: Critical Essays*, 1969–1994 (Dennett 1996b).

75. Глава “Философия как наивная автоантропология” основана на статьях “*Sakes and Dints*” (Dennett 2012) и “*Kinds of Things*” (Dennett 2013c).

77. Глава “Махматные истины высшего порядка” взята с изменениями из статьи “*The Higher-Order Truths about Chess*” (Dennett 2006c).

78. Глава “10 процентов хорошего” ранее не публиковалась.

## Библиография

- Abbott E. A. (1884) 1983: *Flatland: A Romance in Many Dimensions* (репринтное пятое издание 1963 г. с предисловием Айзека Азимова). New York: HarperCollins.
- Alain Ch. (1908) 1956: *Propos d'un Normand 1906–1914*. Paris: Gallimard. (Цитируется в статье: Rogers D. S., Ehrlich P. R. *Natural Selection and Cultural Rates of Change* // *Proceedings of the National Academy of Sciences*, Vol. 105 [2008], pp. 3416–3420.)
- Austin J. L. 1961: *Philosophical Papers*. Oxford: Oxford University Press.
- Axelrod R. 1984: *The Evolution of Cooperation*. New York: Basic Books.
- Axelrod R. and Hamilton W. 1981: *The Evolution of Cooperation* // *Science*, Vol. 211, pp. 1390–1396.
- Barrett J. 2000: *Exploring the Natural Foundations of Religion* // *Trends in Cognitive Science*, vol. 4, pp. 29–34.
- Bateson W. 1916: *Review of the Mechanisms of Mendelian Heredity by T. H. Morgan* (1914).
- Behe M. J. 1996: *Darwin's Black Box: The Biochemical Challenge to Evolution*. New York: Free Press.
- Bennett M., Dennett D., Hacker P. M. S., and Searle J. 2009: *Neuroscience and Philosophy: Brain, Mind, and Language*. New York: Columbia University Press.
- Bennett M., Hacker P. M. S. 2003: *Philosophical Foundations of Neuroscience*. Malden, Mass.: Wiley-Blackwell.
- Block N. 1978: *Troubles with Functionalism* // Savage W. (ed.). *Perception and Cognition: Issues in the Foundations of Psychology*. Minnesota Studies in the Philosophy of Science, vol. 9. Minneapolis: University of Minnesota Press, pp. 261–326.
- Block N. 1994: What Is Dennett's Theory a Theory of? // *Philosophical Topics*, Vol. 22 (special issue on the philosophy of Daniel Dennett), pp. 23–40.
- Bobrow D. 1985: *Qualitative Reasoning about Physical Systems*. Cambridge, Mass.: MIT Press.
- Borges J. L. 1962: *Labyrinths: Selected Stories and Other Writings*. New York: New Directions.
- Bray D. 2009: *Wetware*. New Haven, Conn.: Yale University Press.
- Brockman J. (Ed.) 2009: *What Have You Changed Your Mind About*. New York: HarperCollins.
- Brockman J. (Ed.) 2012: *This Will Make You Smarter*. New York: Harper Torchbook.
- Brooks R. A. 1987: *Planning Is Just a Way of Avoiding Figuring Out What to Do Next. Technical report* // MIT Artificial Intelligence Laboratory, Cambridge, Mass. Доступно в сети: <http://people.csail.mit.edu/brooks/papers/Planning%20is%20Just.pdf>.
- Brooks R. 1991: *Intelligence without Representation* // *Artificial Intelligence*, Vol. 47, pp. 139–159.
- Cann R. L., Stoneking M., and Wilson A. C. 1987: *Mitochondrial DNA and Human Evolution* // *Nature*, Vol. 325, pp. 31–36.
- Chalmers D. 1995: *Facing Up to the Problem of Consciousness* // *Journal of Consciousness Studies*, Vol. 2, No. 3, pp. 200–219.
- Clark A. 2013: *Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science* // *Behavioral and Brain Sciences*.
- Clegg L. 2012: *Protean Free Will. California Institute of Technology*. Pasadena. Доступно в сети: <http://authors.library.caltech.edu/29887/>.
- Cope D. 2000: *The Algorithmic Composer*. Middleton. Wisc.: A-R Editions.
- Cope D. 2001: *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, Mass.: MIT Press.
- Cronin H. 1991: *The Ant and the Peacock*. Cambridge: Cambridge University Press.
- Damiaan D., Mantione M., Figuee M., Van den Munckhof P., Koerselman F., Westenberg H., Bosch A., and Schuurman R. 2010: *Deep Brain Stimulation of the Nucleus Accumbens for Treatment-Refractory Obsessive-Compulsive Disorder* // *Archives of General Psychiatry*, vol. 67, no. 10, pp. 1061–1068.
- Davidson D. 1987: *Knowing One's Own Mind* // *Proceedings and Addresses of the American Philosophical Association*, Vol. 60, pp. 441–458. Reprinted in Davidson, Donald, 2001, *Subjective, Intersubjective, Objective*. New York: Oxford University Press, pp. 15–38.
- Dawkins R. 1976: *The Selfish Gene*. Oxford: Oxford University Press. (Revised edition: 1989.) [Докинз Р. *Эгоистичный ген* / Перевод: Н. Фоминой. М.: Corpus, 2013.]
- Dawkins R. 1982: *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford: Oxford University Press. [Докинз Р. *Расширенный фенотип: длинная рука гена* / Перевод А. Гопко. М.: Corpus, 2010.]

- Dawkins R. 1986: *The Blind Watchmaker*. London: Longmans. [Докинз Р. Слепой часовщик. Как эволюция доказывает отсутствие замысла во Вселенной / Перевод А. Гопко. М.: Corpus, 2015.]
- Dawkins R. 1996: *Climbing Mount Improbable*. London: Viking Penguin.
- Dawkins R. 2004: *The Ancestor's Tale: A Pilgrimage to the Dawn of Time*. London: Weidenfeld & Nicolson. [Докинз Р. Рассказ предка. Паломничество к истокам жизни / Перевод С. Долотовской. М.: Corpus, 2015.]
- De Vries P. 1953: *The Vale of Laughter*. Boston: Little, Brown.
- Dehaene S., Marquies J. F. 2002: Cognitive Euroscience: Scalar Variability in Price Estimation and the Cognitive Consequences of Switching to the Euro // *Quarterly Journal of Experimental Psychology*, Vol. 55, pp. 705–731.
- Dennett D. C. 1968: *Features of Intentional Action* // *Philosophy and Phenomenological Research*, Vol. 29 (December), pp. 232–244.
- Dennett D. C. 1969: *Content and Consciousness*. London: Routledge & Kegan Paul.
- Dennett D. C. 1971: *Intentional Systems* // *Journal of Philosophy*, Vol. 68, pp. 87–106.
- Dennett D. C. 1975: *Brain Writing and Mind Reading* // Gunderson K. (ed.). *Language, Mind and Knowledge*. Minnesota Studies in the Philosophy of Science, Vol. 7. Minneapolis: University of Minnesota Press, pp. 403–416. (Перепечатано в: Dennett 1978a.)
- Dennett D. C. 1978a: *Brainstorms*. Cambridge, Mass.: MIT Press/A Bradford Book.
- Dennett D. C. 1978b: *Current Issues in the Philosophy of Mind* // *American Philosophical Quarterly*, Vol. 15, pp. 249–261.
- Dennett D. C. 1978c: *Why You Can't Make a Computer That Feels Pain* // *Synthese*, Vol. 38 (August), pp. 415–456.
- Dennett D. C. 1980: *The Milk of Human Intentionality* // *Behavioral and Brain Sciences*, Vol. 3, pp. 428–430.
- Dennett D. C. 1981: *Three Kinds of Intentional Psychology* // Healey R. (ed.). *Reduction, Time and Reality*. Cambridge: Cambridge University Press, pp. 37–61.
- Dennett D. C. 1982a: *Beyond Belief* // Woodfield A. (ed.). *Thought and Object: Essays on Intentionality*. Oxford: Oxford University Press. Reprinted in Dennett, 1987.
- Dennett D. C. 1982b: *How to Study Consciousness Empirically: or Nothing Comes to Mind* // *Synthese*, Vol. 53, pp. 159–180. [Денет Д. Как исследовать человеческое сознание эмпирически / пер. с англ. Н. С. Юлиной // История философии. М.: ИФ РАН, 2005. Вып. 12.]
- Dennett D. C. 1982c: *The Myth of the Computer: An Exchange* // *New York Review of Books*, Vol. 29 (June 24), pp. 56–57.
- Dennett D. C. 1982d: *Notes on Prosthetic Imagination* // *New Boston Review*, Vol. 7 (June), pp. 3–7. (Перепечатано: 30 *Years of Boston Review* // *Boston Review*, Vol. 30, No. 5 (September / October 2005), p. 40).
- Dennett D. C. 1983: *Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended* // *Behavioral and Brain Sciences*, Vol. 6, pp. 343–390.
- Dennett D. C. 1984a: *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Mass.: MIT Press.
- Dennett D. C. 1984b: *I Could Not Have Done Otherwise – So What?* // *Journal of Philosophy*, Vol. 81, pp. 553–565.
- Dennett D. C. 1986: *The Logical Geography of Computational Approaches: A View from the East Pole* // Harnish R. M., Brand M. (eds.). *The Representation of Knowledge and Belief*. Tucson: University of Arizona Press, pp. 59–79.
- Dennett D. C. 1987: *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- Dennett D. C. 1988a: *Quining Qualia* // Marcel A., Bisiach E. (eds.). *Consciousness in Modern Science*. Oxford: Oxford University Press, pp. 42–77.
- Dennett D. C. 1988b: *Why Everyone Is a Novelist* // *Times Literary Supplement*, Vol. 4 (September 16–22), p. 459.
- Dennett D. C. 1988c: *Evolution, Error and Intentionality* // Wilks Y., Partridge D. (eds.). *Sourcebook on the Foundations of Artificial Intelligence*. Albuquerque: University of New Mexico Press, pp. 190–211.
- Dennett D. C. 1990: *Memes and the Exploitation of Imagination* // *Journal of Aesthetics and Art Criticism*, Vol. 48, pp. 127–135.
- Dennett D. C. 1991a: *Consciousness Explained*. Boston: Little, Brown.
- Dennett D. C. 1991b: *Real Patterns* // *Journal of Philosophy*, Vol. 88, pp. 27–51.
- Dennett D. C. 1993: *Confusion over Evolution: An Exchange* // *New York Review of Books*, January 14, pp. 43–44.

- Dennett D. C. 1994a: *Get Real* // Philosophical Topics, Vol. 22, pp. 505–556.
- Dennett D. C. 1994b: *The Practical Requirements for Making a Conscious Robot* // Proceedings of the Royal Society, A, Vol. 349, pp. 133–146.
- Dennett D. C. 1995a: *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.
- Dennett D. C. 1995b: *How to Make Mistakes* // Brockman J., Matson K. (eds.). *How Things Are*. New York: William Morrow, pp. 137–144.
- Dennett D. C. 1996a: *Kinds of Minds: Towards an Understanding of Consciousness*. New York: Basic Books. [Деннет Д. *Виды психики: на пути к пониманию сознания* / пер. с англ. А. Веретенникова; под общ. ред. Л. Б. Макеевой. М.: Идея-Пресс, 2004.]
- Dennett D. C. 1996b: *Review of Other Minds: Critical Essays, 1969–1994 by Thomas Nagel*, 1995 // Journal of Philosophy, Vol. 63, No. 8 (August), pp. 425–428.
- Dennett D. C. 2000: *With a Little Help from My Friends* // Ross D., Brook A., and Thompson D. (eds.). *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge, Mass.: MIT Press, pp. 327–388.
- Dennett D. C. 2001a: *Collision-Detection, Muselot, and Scribble: Some Reflections on Creativity* // Cope 2001, pp. 283–291.
- Dennett D. C. 2001b: *Consciousness: How Much Is That in Real Money?* // Gregory R. (ed.). *The Oxford Companion to the Mind*. 2<sup>nd</sup> ed. Oxford: Oxford University Press.
- Dennett D. C. 2001c: *Explaining the 'Magic' of Consciousness* // Exploring Consciousness, Humanities, Natural Science, Religion, Proceedings of the International Symposium, Milano, November 19–20, 2001 (published in December 2002, Fondazione Carlo Erba), pp. 47–58.
- Dennett D. C. 2001d: *In Darwin's Wake, Where Am I?* // Proceedings and Addresses of the American Philosophical Association, Vol. 75, No. 2 (November), pp. 13–30. (Перепечатано: Hodge J., Radick G. [eds.]. 2003: *The Cambridge Companion to Darwin*. Cambridge: Cambridge University Press, pp. 357–376.)
- Dennett D. C. 2001e: *Things about Things* // Branquinho J. (ed.). *The Foundations of Cognitive Science*. Oxford: Clarendon Press, pp. 133–149.
- Dennett D. C. 2002: *The New Replicators* // Pagel M. (ed.). *The Encyclopedia of Evolution*. Vol. 1. Oxford: Oxford University Press, pp. E83 – E92.
- Dennett D. C. 2003: *Freedom Evolves*. New York: Viking Penguin.
- Dennett D. C. 2004: *Holding a Mirror up to Dupré* // Philosophy and Phenomenological Research, Vol. 69, No. 2 (September), pp. 473–483.
- Dennett D. C. 2005a: *Natural Freedom* // Metaphilosophy, Vol. 36, No. 4 (July), pp. 449–459.
- Dennett D. C. 2005b: *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, Mass.: MIT Press.
- Dennett D. C. 2006a: *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Viking Penguin.
- Dennett D. C. 2006b: *From Typo to Thinko: When Evolution Graduated to Semantic Norms* // Levinson S., Jaisson P. (eds.). *Evolution and Culture*. Cambridge, Mass.: MIT Press, pp. 133–145.
- Dennett D. C. 2006c: *The Higher-Order Truths about Chmess* // Topoi, pp. 39–41.
- Dennett D. C. 2007a: *The God Delusion by Richard Dawkins* // Free Inquiry, Vol. 27, No. 1 (December/January 2007).
- Dennett D. C. 2007b: *Heterophenomenology Reconsidered* // Phenomenology and Cognitive Science, Vol. 6, Nos. 1 and 2, pp. 247–270.
- Dennett D. C. 2007c: *Varieties of Content* // Presentation at Concepts: Content and Constitution, A Symposium, University of Copenhagen, Amager, Denmark, May 12.
- Dennett D. C. 2007d: *What RoboMary Knows* // Alter T., Walter S. (eds.). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press, pp. 15–31.
- Dennett D. C. 2008: *Competition in the Brain* // World Question Center. Edge.org, December. (Также опубликовано в сборнике: Brockman 2009.)
- Dennett D. C. 2009a: *Darwin's 'Strange Inversion of Reasoning'* // Proceedings of the National Academy of the Sciences of the United States of America, Vol. 106, Suppl. 1, pp. 10061–10065.
- Dennett D. C. 2009b: *Heterophenomenology* // Bayne T., Cleeremans A., and Wilken P. (eds.). *The Oxford Companion to Consciousness*. Oxford: Oxford University Press, pp. 345–346.
- Dennett D. C. 2009c: *Intentional Systems Theory* // McLaughlin B., Beckermann A., and Walter S. (eds.). *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press, pp. 339–350.

Dennett D. C. 2010: *Homunculi Rule: Reflections on Darwinian Populations and Natural Selection* by Peter Godfrey-Smith, Oxford University Press, 2009 // Biology and Philosophy. Доступно в сети: <http://ase.tufts.edu/cogstud/papers/homunculi.pdf>.

Dennett D. C. 2011a: *Cycles* // World Question Center. Edge.org. (Перепечатано: Brockman 2012, pp. 81–88.)

Dennett D. C. 2011b: *Shall We Tango? No, but Thanks for Asking* // Journal of Consciousness Studies, Vol. 18, Nos. 5 and 6, pp. 23–34.

Dennett D. C. 2012: *Sakes and Dints* // Times Literary Supplement, March 2, pp. 12–14.

Dennett D. C. 2013a: *The Evolution of Reasons* // Bashour B., Muller H. D. (eds.). *Contemporary Philosophical Naturalism and Its Implications*. New York: Routledge, pp. 13–47.

Dennett D. C. 2013b: *Expecting Ourselves to Expect: the Bayesian brain as a projector* // Behavioral and Brain Sciences, June 36 (3), pp. 209–210.

Dennett D. C. 2013c: *Kinds of Things – Towards a Bestiary of the Manifest Image* // Ross D., Ladyman J., and Kincaid H. (eds.). *Scientific Metaphysics*. Oxford: Oxford University Press, pp. 96–107.

Dennett D. C. 2013e: *Turing's 'Strange Inversion of Reasoning'* // Cooper S. B., Van Leeuwen J. (eds.). *Alan Turing – His Work and Impact*. Elsevier, pp. 569–573.

Dennett D. C., Akins K. 2008: *The Multiple Drafts Model* // Scholarpedia, Vol. 3, No. 4, 4321: [http://www.scholarpedia.org/wiki/index.php?title=Multiple\\_drafts\\_model](http://www.scholarpedia.org/wiki/index.php?title=Multiple_drafts_model).

Dennett D. C., Plantinga A. 2011: *Science and Religion: Are They Compatible?* Oxford: Oxford University Press.

Dennett D. C., Westbury C. F. 2000: *Mining the Past to Construct the Future: Memory and Belief as Forms of Knowledge* // Schacter D., Scarry E. (eds.). *Memory, Brain, and Belief*. Cambridge, Mass.: Harvard University Press, pp. 11–32.

Denton M. 1985: *Evolution: A Theory in Crisis*. London: Burnett Books.

Dewdney A. K. 1984: *The Planiverse: Computer Contact with a Two-Dimensional World*. New York: Poseidon Press.

Eigen M. 1992: *Steps towards Life*. Oxford: Oxford University Press.

Ellis H., Young A. 1990: *Accounting for Delusional Misidentifications* // British Journal of Psychiatry, Vol. 157, pp. 239–248.

Feynman R. 1985: *“Surely You’re Joking, Mr. Feynman!”: Adventures of a Curious Character*. New York: W. W. Norton.

Fitch T. 2008: *Nano-Intentionality: A Defense of Intrinsic Intentionality* // Biology and Philosophy, Vol. 23, pp. 157–177.

Fitzgibbon C. D., Fanshawe J. H. 1988: *Stotting in Thomson’s Gazelles: An Honest Signal of Condition* // Behavioral Ecology and Sociobiology, Vol. 23, No. 2 (August), pp. 69–74.

Fodor J. 1975: *The Language of Thought*. Hassocks, Sussex: Harvester Press.

Fodor J. 2003: *Why Would Mother Nature Bother?* // London Review of Books, Vol. 25, No. 5, pp. 17–18.

Fodor J. 2008: *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Fodor J., Lepore E. 1992: *Holism: A Shopper’s Guide*. Oxford: Blackwell.

Fodor J., Piatelli-Palmerini M. 2010: *What Darwin Got Wrong*. New York: Farrar, Straus & Giroux.

Frankfurt H. 1969: *Alternate Possibilities and Moral Responsibility* // Journal of Philosophy, Vol. 65, pp. 829–833.

Gell-Mann M. 1995: *The Quark and the Jaguar: Adventures in the Simple and the Complex*. New York: St. Martin’s.

Gibson J. J. 1979: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Godfrey-Smith P. 2009: *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.

Godfrey-Smith P. 2011: *Agents and Acacias: Replies to Dennett, Sterelny, and Queller* // Biology and Philosophy, Vol. 26, pp. 501–515.

Goffman E. 1959: *The Presentation of Self in Everyday Life*. Edinburgh: University of Edinburgh Social Sciences Research Centre.

Gould S. J. 1977: *Ever since Darwin*. New York: W. W. Norton.

Gould S. J. 1989a: *Tires to Sandals* // Natural History, April, pp. 8–15.

Gould S. J. 1989b: *Wonderful Life: The Burgess Shale and the Nature of History*. New York: W. W. Norton.

Gould S. J. 1992a: *The Confusion over Evolution* // New York Review of Books, November 19.

Gould S. J. 1992b: *Life in a Punctuation* // Natural History, October, pp. 10–21.

- Gould S. J. 1993: *Confusion over Evolution: An Exchange* // New York Review of Books, January 14, pp. 43–44.
- Graham G., Horgan T. 2000: *Mary Mary Quite Contrary* // Philosophical Studies, Vol. 99, pp. 59–87.
- Greene J., Cohen J. 2004: *For the Law, Neuroscience Changes Everything and Nothing* // Philosophical Transactions of the Royal Society, Vol. 359, pp. 1775–1785.
- Harris S. 2012: *Free Will*. New York: Free Press.
- Haugeland J. 1981: *Mind Design*. Cambridge, Mass.: MIT Press / Bradford Book.
- Haugeland J. 1985: *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press.
- Hawking S. W. 1988: *A Brief History of Time*. New York: Bantam. [Хокинг С. *Краткая история времени: от Большого взрыва до чёрных дыр* / Пер. с англ. Н. Я. Смородинской. СПб.: Амфора, 2001.]
- Hayes P. 1978: *The Naïve Physics Manifesto* // Michie D. (ed.). *Expert Systems in the Microelectronic Age*. Edinburgh: Edinburgh University Press.
- Heider F., Simmel M. 1944: *An Experimental Study of Apparent Behavior* // American Journal of Psychology, Vol. 57, No. 2, pp. 243–259.
- Hillard J. N. 1938: *Card Magic*. Minneapolis: Carl W. Jones.
- Hofstadter D. 1979: *Gödel Escher Bach*. New York: Basic Books. [Хофштадтер Д. *Гёдель, Эшер, Бах: эта бесконечная гирлянда* / Перевод М. Эскиной. Самара: Бахрах-М, 2001.]
- Hofstadter D. 1982: *Metamagical Themas: Can Inspiration Be Mechanized?* // Scientific American, September, pp. 18–34. (Перепечатано под названием *On the Seeming Paradox of Mechanizing Creativity* в Hofstadter 1985, pp. 526–546.)
- Hofstadter D. 1985: *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York: Basic Books.
- Hofstadter D. 1997: *Le Ton Beau de Marot: In Praise of the Music of Language*. New York: Basic Books.
- Hofstadter D. 2007: *I Am a Strange Loop*. New York: Basic Books.
- Hofstadter D., Dennett D. C. (eds.). 1981: *The Mind's I*. New York: Basic Books. [Хофштадтер Д., Деннет Д. *Глаз разума* / Перевод М. Эскиной. Самара: Бахрах-М, 2003.]
- Holland J. 1975: *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Hume D. (1739) 1964: *A Treatise of Human Nature*. Oxford: Clarendon. [Юм Д. *Трактат о человеческой природе, или попытка применить основанный на опыте метод рассуждения к моральным предметам* / Перевод С. И. Церетели // Юм Д. *Сочинения в 2 томах*. Т. I. 2<sup>е</sup> изд. М.: Мысль.]
- Humphrey N. 1987: *Scientific Shakespeare* // The Guardian (London), August 26.
- Humphrey N., Dennett D. C. 1989: *Speaking for Our Selves: An Assessment of Multiple Personality Disorder* // Raritan: A Quarterly Review, Vol. 9 (Summer), pp. 68–98. (Перепечатано с примечаниями: Occasional Paper 8, Center on Violence and Human Survival, John Jay College of Criminal Justice, City University of New York, 1991.)
- Hurley M., Dennett D. C., Adams R. B., Jr. 2011: *Inside Jokes: Using Humor to Reverse-Engineer the Mind*. Cambridge, Mass.: MIT Press.
- Jackendorff R. 1987: *Consciousness and the Computational Mind*. Cambridge, Mass.: MIT Press/Bradford Book.
- Jackendorff R. 1993: *Patterns in the Mind: Language and Human Nature*. Harlow, Essex: Harvester Wheatsheaf; New York: Basic Books, 1994.
- Jackson F. 1982: *Epiphenomenal Qualia* // Philosophical Quarterly, Vol. 32, pp. 127–136.
- Kane R. 1996: *The Significance of Free Will*. Oxford: Oxford University Press.
- Knobe J., Nichols S. (eds.). 2008: *Experimental Philosophy*. Oxford: Oxford University Press.
- Leibniz G. (1714) 1898: *Monadology*. In *The Monadology and Other Philosophical Writings* / Translated by Robert Latta. Oxford: Oxford University Press. [Лейбниц Г.-В. *Сочинения в четырех томах*: Т. I. М.: Мысль, 1982. С. 413–429.]
- Lenat D. B., Guha R. V. 1990: *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, Mass.: Addison-Wesley.
- Lettvin J. Y., Maturana U., McCulloch W., and Pitts W. 1959: *What the Frog's Eye Tells the Frog's Brain* // Proceedings of the Institute of Radio Engineers, Vol. 47, pp. 1940–1951.
- Levine J. 1994: *Out of the Closet: A Qualophile Confronts Qualophobia* // Philosophical Topics, Vol. 22, pp. 107–126.
- Lewis D. 1978: *Truth in Fiction* // American Philosophical Quarterly, Vol. 15, pp. 37–46.
- Lloyd M., Dybas H. S. 1966: *The Periodical Cicada Problem* // Evolution, Vol. 20, pp. 132–149.
- Lloyd Morgan C. 1894: *An Introduction to Comparative Psychology*. London: W. Scott.

- Ludlow P., Nagasawa Yu., Stoljar D. (eds.) 2004: *There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. Cambridge, Mass.: MIT Press / Bradford Books.
- Mackenzie R. B. 1868: *The Darwinian Theory of the Transmutation of Species Examined*. London: Nisbet.
- Maynard Smith J. 1978: *The Evolution of Sex*. Cambridge: Cambridge University Press.
- McClelland J., Rumelhart D., and The PDP Research Group 1986: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 2. Cambridge, Mass.: MIT Press.
- Menabrea L. F. 1842: *Sketch of the Analytic Engine Invented by Charles Babbage* // Bibliothèque Universelle de Genève, no. 82 (October). Translated by Augusta Ada King, Countess of Lovelace, 1843, with notes, in Scientific Memoirs, Vol. 3, pp. 666–731.
- Meyer S. C. 2009: *Signature in the Cell: DNA and the Evidence for Intelligent Design*. New York: HarperOne.
- Millikan R. 1984: *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.
- Millikan R. 1993: *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: MIT Press.
- Nagel Th. 1974: *What Is It Like to Be a Bat?* // Philosophical Review, Vol. 83, pp. 435–450. [Нагель Т. Каково быть летучей мышью? / Перевод с англ. М. А. Эскиной. Самара: Бахрах-М, 2003.]
- Nagel Th. 2009: *Recommendation for Book of the Year* // Times Literary Supplement, November 27.
- Nagel Th. 2010: *Letter to the editor* // Times Literary Supplement, January 1.
- Neugebauer O. 1989: *A Babylonian Lunar Ephemeris from Roman Egypt* // E. Leichty, M. de J. Ellis, and P. Gerardi (eds.). *A Scientific Humanist: Studies in Honor of Abraham Sachs*. Philadelphia: Occasional Publications of the Samuel Noah Kramer Fund no. 9, pp. 301–304.
- Pinker S. 2002: *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking. [Пинкер С. Чистый лист. Природа человека. Кто и почему отказывается признавать ее сегодня. М.: Альпина Нон-фикшн, 2018.]
- Popek G. J., Goldberg R. P. 1974: Formal Requirements for Virtualizable Third Generation Architectures // Communications of the ACM, Vol. 17, No. 7, pp. 412–421. Доступно в сети: <http://doi.acm.org/10.1145/361011.361073>.
- Poundstone W. 1985: *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge*. New York: William Morrow.
- Putnam H. 1975: *The Meaning of 'Meaning'* // Gunderson K. (ed.). *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*. Vol. 7. Minneapolis: University of Minnesota Press. (Перепечатано: Putnam 1975: *Mind, Language and Reality* (Philosophical Papers, Vol. 2). Cambridge: Cambridge University Press.)
- Quine W. V. O. 1960: *Word and Object*. Cambridge, Mass.: MIT Press. [Куайн У. В. О. Слово и объект / Перевод с англ. А. З. Черняка и Т. А. Дмитриева. М.: Праксис; Логос, 2000.]
- Quine W. V. O. 1987: *Universal Library* // Quiddities: An Intermittently Philosophical Dictionary. Cambridge, Mass.: Harvard University Press.
- Rapoport A. 1960: *Fights, Games, and Debates*. Ann Arbor: University of Michigan Press.
- Rapoport A. 1961: Three Modes of Conflict // Management Science, Vol. 3, p. 210.
- Ridley M. 1993: *The Red Queen: Sex and the Evolution of Human Nature*. New York: Macmillan.
- Ridley M. 2004: *Nature via Nurture. London: Fourth Estate*. (Также опубликовано под названием *The Agile Gene: How Nature Turns on Nurture*. New York: HarperCollins.)
- Ross A. 2013: *Inconceivable Minds*. (Philosophy PhD dissertation.) University of North Carolina at Chapel Hill.
- Ruina A. 2011: *Cornell Ranger, 2011, 4-Legged Bipedal Robot* // [ruina.tam.cornell.edu/research/topics/locomotion\\_and\\_robotics/ranger/Ranger2011/](http://ruina.tam.cornell.edu/research/topics/locomotion_and_robotics/ranger/Ranger2011/).
- Rumelhart D. E., McClelland J. L., and The PDP Research Group 1986: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1. Cambridge, Mass.: MIT Press.
- Ryder D., Kingsbury J., Williford K. (eds.) 2013: *Millikan and Her Critics*. Oxford: Wiley-Blackwell.
- Sanford D. 1975: *Infinity and Vagueness* // Philosophical Review, Vol. 84, pp. 520–535.
- SchÖnborn Ch. 2005: *Finding Design in Nature* // New York Times, July 7.
- Searle J. 1980: *Minds, Brains and Programs* // Behavioral and Brain Sciences, Vol. 3, pp. 417–458. [Сёрл Дж. Р. Сознание, мозг и программы / Перевод А. Л. Блинова // Грязнов А. Ф. (Ред.). Аналитическая философия: становление и развитие. М.: Дом интеллектуальной книги; Прогресс-Традиция, 1998. С. 376–400.]
- Searle J. 1982: *The Myth of the Computer* // New York Review of Books, Vol. 29 (April 29).

- Searle J. 1988: *Turing the Chinese Room* // Singh T. (ed.). *Synthesis of Science and Religion, Critical Essays and Dialogues*. San Francisco: Bhaktivedanta Institute.
- Sellars W. 1962: *Philosophy and the Scientific Image of Man* // *Science, Perception and Reality*. London: Routledge & Kegan Paul.
- Seung S. 2007: *The Once and Future Science of Neural Networks*. Presented at the Society for Neuroscience meeting, San Diego, November 4.
- Siegel L. 1991: *Net of Magic: Wonders and Deceptions in India*. Chicago: University of Chicago Press.
- Sims K. 1994: *Evolved Virtual Creatures* // <http://www.karlsims.com/evolved-virtualcreatures.html>.
- Strawson G. 2003: *Evolution Explains It All for You* // New York Times, March 2.
- Strawson G. 2010: *Your Move: The Maze of Free Will* // The Stone, New York Times online, July 22, 2010: [http://www.scribd.com/doc/86763712/Week-2-](http://www.scribd.com/doc/86763712/Week-2-Strawson-The-Maze-of-Free-Will)
- Thompson E. 2007: *Mind in Life*. Cambridge, Mass.: Belknap Press, Harvard University Press.
- Vohs K. D., Schooler J. W. 2008: *The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating* // *Psychological Science*, pp. 49–54.
- Von Neumann J., Burks A. (ed.). *Theory of Self-Reproducing Automata*. Champaign-Urbana: University of Illinois Press.
- Von Neumann J., Morgenstern O. 1944: *Theory of Games and Economic Behavior*. Princeton, N. J.: Princeton University Press.
- Von Uexküll J. (1934) 1957: *A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds* // Schiller C. H. (ed. and trans.). *Instinctive Behavior: The Development of a Modern Concept*. New York: International Universities Press.
- Wang H. 1957: *A Variant to Turing's Theory of Computing Machines* // *Journal of the Association for Computing Machinery*, pp. 63–92.
- Wiggins D. 1973: *Towards a Reasonable Libertarianism* // Honderich T. (ed.). *Essays on Freedom of Action*. London: Routledge & Kegan Paul, pp. 31–63.
- Wimsatt W. C. 1980: *Randomness and Perceived Randomness in Evolutionary Biology* // *Synthese*, vol. 43, pp. 287–290.
- Wolfe T. 2000: *Sorry, But Your Soul Just Died* // *Hooking Up*. New York: Farrar, Straus & Giroux.
- Wooldridge D. 1963: *The Machinery of the Brain*. New York: McGraw-Hill.
- Wright R. 2000: *Nonzero: The Logic of Human Destiny*. New York: Pantheon.
- Zahavi A. 1987: *The Theory of Signal Selection and Some of Its Implications* // Delfino V. P. (ed.). Bari, 9–14 April 1985. Bari: Adriatici Editrici, pp. 305–327.



## Цитаты и иллюстрации, защищенные авторским правом

Цитата на странице 37 взята из книги “Вы, конечно, шутите, мистер Фейнман!” Ричарда Фейнмана в изложении Ральфа Лейтона. © *Richard P. Feynman, Ralph Leighton*, 1985. Использовано с разрешения *W. W. Norton & Company, Inc.*

Цитата на страницах 252–253 взята из книги “Рассказ предка” Ричарда Докинза. © *Richard Dawkins*, 2004. Использовано с разрешения *Houghton Mifflin Harcourt Publishing Company* и *The Orion Publishing Group*. Все права защищены.

Древо жизни со страниц 254–255 создано Леонардом Эйзенбергом. © *Leonard Eisenberg*. Все права защищены. <http://www.evogeneao.com>.

Карикатура на странице 263 нарисована Сидни Харрисом: <http://www.sciencecartoonsplus.com>.

Фотографию термитника со страницы 276 сделала Фиона Стюарт.

Фотография храма Святого Семейства со страницы 276 принадлежит *Diario di Viaggio*.

---

notes

## Сноски

Полужирным шрифтом выделены названия инструментов мышления, более подробно описанные в других главах этой книги. Ищите их в алфавитном указателе, поскольку не каждому из инструментов посвящена отдельная глава. (*Reductio ad absurdum*: лат. сведение к абсурду. – Прим. ред.)

Эзоп, как и Гомер, едва ли не столь же легендарен, как его басни, которые веками передавались из уст в уста, пока их впервые не записали за несколько сотен лет до эпохи Платона и Сократа. Возможно, Эзоп не был греком. Существуют косвенные свидетельства его эфиопского происхождения.

Философ У. В. О. Куайн (1960) назвал это семантическим восхождением, подъемом на следующую ступень от разговоров об электронах, справедливости, лошадях или о чем угодно еще, чтобы говорить о разговорах об электронах, справедливости, лошадях или о чем угодно еще. Порой люди возражают против этого перехода философов на следующий уровень (“Для вас, ребята, здесь все семантика!”), а порой переход действительно не несет пользы или вовсе сбивает с толку, однако когда он все же необходим – когда люди не слышат друг друга или обманываются негласными представлениями о значении собственных слов, семантического восхождения или перехода на метауровень, – он становится ключом к ясности.

Многие пассажи этой книги взяты из книг и статей, которые я публиковал ранее, и переработаны, чтобы сделать их практичнее и универсальнее и применять не только в оригинальном контексте, что отличает большинство хороших инструментов. К примеру, история о фон Неймане уже появлялась в моей книге “Опасная идея Дарвина”, опубликованной в 1995 г., а эти рассуждения о прикладных инструментах Хофштадтера я позаимствовал из своей статьи “«Странная логика» Дарвина”, опубликованной в 2009 г. в журнале *PNAS*. Вместо того чтобы снабжать каждый подобный момент постраничной сноской, я привожу список источников в конце книги.

Вот два из лучших изречений: “Философия для науки – что голуби для статуй” и “Философия для науки – что порнография для секса: дешевле, проще и некоторым больше по душе”. (Не буду называть авторов этих высказываний, но оставляю им возможность самим заявить о своем авторстве.)

Это не позволяет штурманам определить их точное местоположение, то есть указать конкретную точку на карте, но позволяет провести *линию*. Они находятся где-то на этой *линии положения*. Нужно несколько часов подождать, пока переместится солнце. Затем выбрать на линии положения произвольную точку и рассчитать, на какой высоте было бы солнце, если бы эта точка была вашим фактическим местоположением. Сделав наблюдение, необходимо сравнить результаты, внести правки и получить новую линию положения. Точка пересечения двух линий положения показывает ваше положение на карте. За прошедшие часы изменится не только высота, но и компасный азимут солнца, поэтому линии будут пересекаться под значительным углом. На практике в эти часы вы обычно находитесь в движении, поэтому необходимо переместить первую линию положения в направлении вашего движения, рассчитав скорость и проведя новую линию параллельно первой. Реальность всегда вносит свои коррективы, поэтому лучше провести три разных линии положения. Если они пересекутся в одной точке, вы либо невероятно хороши в своем деле, либо столкнулись с невероятной удачей, но гораздо чаще линии формируют небольшой треугольник, называемый *треугольником погрешности*. Центр треугольника погрешности принимается за новое расчетное положение вашего судна.



Таков идеал, но мы не всегда живем по этим принципам, поскольку человеческая природа берет свое. Одна из известных, но нерешенных проблем в текущей научной практике заключается в том, что отрицательные результаты – эксперименты, которые не показали того, что должны были показать, – публикуются недостаточно часто. Этот недостаток системы описан и осужден в знаменитой фейнмановской “Лекции о самолетопоклонниках” – торжественной речи, которую Фейнман прочитал в Калтехе в 1974 г., а затем опубликовал в 1985-м.

Эта конференция запомнилась мне еще купанием Поппера в Гранд-канале. Он поскользнулся, выходя из моторной лодки на пристань острова Сан-Джорджо, полетел ногами вперед в канал и успел погрузиться по колено, прежде чем его поймали и вытащили на причал двое ловких лодочников. Организаторы конференции не на шутку перепугались и собрались было бежать обратно в гостиницу, чтобы принести девяностолетнему сэру Карлу сухие брюки, но оказалось, что других брюк он с собой не привез, а до его выступления на открытии конференции оставалось менее получаса! На помощь пришла итальянская смекалка – минут через пять мне довелось наблюдать незабываемое зрелище: сэр Карл царственно восседал на маленьком стуле в самом центре зала со сводчатым потолком и мраморным полом (построенного по проекту Палладио), а вокруг него на коленях стояло не меньше полдюжины одетых в мини-юбки девушек, сушивших его штанины своими фенами. Из центра зала к стенам тянулись удлинители, придававшие композиции сходство с разноцветной маргариткой, в серединке которой сидел сэр Карл, серьезный и невозмутимый. Через пятнадцать минут он был совсем сухим и потрясал кулаком со сцены, отстаивая правильность своей дуалистической позиции.

Турнир Аксельрода (Axelrod, Hamilton 1981; Axelrod 1984) открыл дорогу в благодатное поле теоретических исследований эволюции альтруизма. Я описываю его в общих чертах в “Опасной идее Дарвина” (Dennett 1995, pp. 479–480), а в последнее время в лабораториях по всему миру наблюдается бум различных вариаций в форме симуляций и экспериментов. Потрясающе простая интерпретация Рапопорта – “я не трону тебя, если ты не тронешь меня” – стала зерном, из которого выросли все последующие исследования и модели.

Здесь я сформулировал правила Рапопорта сам, по памяти из писем, которыми мы с Рапопортом обменивались много лет назад и которые сейчас, очевидно, утеряны. Сэмюэл Рут недавно сказал мне, что впервые правила Рапопорта появились в его книге “Споры, игры и дебаты” (1960) и его статье “Три вида конфликтов” (1961), где сформулировано первое правило, авторство которого приписано Карлу Роджерсу, а также вариации на тему остальных. Моя версия кажется несколько более емкой и универсальной.

Его превзошел философ Джерри Фодор (2008), который взял в привычку выделять свои “безусловно” курсивом и повторять их (например, на с. 38), словно говоря: *“Получайте, скептики! Получайте, скептики!”*

Недавно Мириам увидела в интернете, что я использую ее термин, и связалась со мной. Ее версия истории немного отличается, но суть остается неизменной: “В нашей семье термин носит пренебрежительный характер – это идея, которая маскируется под Истину, чтобы превознести говорящего”. Мириам любезно согласилась, чтобы я использовал ту версию истории, которой поделился со мной ее отец (как я ее помню), и дал несколько более узкое определение ее блестящему неологизму.

Они умеют говорить, скажут энтузиасты, просто мы еще не открыли и не научились переводить их языки. Такая мысль бесконечно привлекательна, однако проведенные в последние годы подробнейшие исследования коммуникации обезьян, птиц и дельфинов показывают, что способность животных “делиться мыслями” крайне ограничена. Если это язык, то термитов можно считать инженерами, богомолы исповедуют религию, а волки заседают в парламентах.

Джерри Фодор (1975) очень эффективно использовал этот риторический вопрос в своей новаторской книге “Язык мысли” и подчеркнул подразумеваемый ответ, процитировав вызывающее замечание Линдона Джонсона: “Другого президента у вас нет” (р. 27).



Этот вывод часто называют *холизмом* ментального (или интенционального); против этого холизма решительно выступал Джерри Фодор, который утверждал, что без труда может представить существо, имеющее ровно одно убеждение (Fodor, Lepore 1992).

Важные детали этого примера мы разберем в главе 27.

Приношу свои извинения тем из вас, кто не прочь представить, каково быть маргариткой, и в утешение предлагаю концепцию *Umwelt*: она позволяет вам отдать должное реальной исключительности маргаритки и не хватить при этом лишку.

Это высказывание приписывается многим, включая философа Людвиг Витгенштейна, художника Пауля Клее и критика Виктора Шкловского.

Исключения лишь подтверждают правило. Затруднения с народной психологией – один из наиболее явных симптомов аутизма как в легкой, так и в тяжелой форме. Аутисты, которым вообще под силу интерпретировать поведение людей, например Тэмпл Грандин, судя по всему, *действительно* полагаются на выведенную ценой больших усилий теорию. В то время как мы без труда и даже без отчета понимаем значение жестов, улыбок и обстоятельств, ей приходится делать интерпретации на основании того, что она наблюдает. Она неумоимо собирает свидетельства, чтобы отличать дружеское приветствие от угрозы и понимать, когда люди приходят к компромиссу. Она научилась узнавать обещания и покупки, научилась шутить и врать. Шелдон Купер из “Теории Большого взрыва” – известный персонаж с синдромом Аспергера, легкой формой аутизма, и мы видим, как его блестящий научный ум вскипает всякий раз, когда ему приходится анализировать детали будничных ситуаций так, словно он анализирует влияние катализаторов на химическую реакцию в пробирке. Тэмпл Грандин и вымышленный Шелдон Купер явно пользуются теорией сознания. Если ею пользуемся и мы, наш вариант теории сильно отличается от их вариантов.

Я назвал это “интенциональной” установкой, а не установкой “рационального агента” и не установкой “убеждение-желание”, потому что в то время (в 1971 г.) большое количество философских трудов по *интенциональности* использовало логику “интенциональных идиом”, которые ссылались на *пропозициональные установки* убеждения, желания, ожидания. Заключение в этих идиомах очемность не только приводила к возникновению логических проблем, но и указывала на их решения. В убеждении или предложении может говориться о несуществующем объекте, например о Санта-Клаусе. Несуществующее может быть и объектом желания, например фонтан вечной молодости или просто стакан холодной воды посреди пустыни. *Очемность* интенциональной установки заключается в том, *о чем* думают агенты. При нормальном ходе событий мы не принимаем интенциональную установку *намеренно* – обычно это происходит произвольно и редко осознается.

В моей статье “Интенциональные системы в когнитивной этологии: в защиту парадигмы Панглосса”, опубликованной в 1983 г. в журнале *Behavioral and Brain Sciences*, приводятся подробности, сопровождаемые целым хором возражений и опровержений. Более свежие рассуждения об интенциональной установке и опровержение типичных возражений можно найти в моем эссе “Теория интенциональных систем” (2009с), которое легло в основу этой главы.

Это главная мысль моей книги “Объясненное сознание” (1991а). Идея о существовании такого места – я называю его картезианским театром, – очевидно, невероятно привлекательна, поскольку многие мыслители все равно не понимают, что его на самом деле нет, несмотря на все доводы, приводимые мною и другими учеными.



Нечасто можно встретить ремарку, что в любимом мысленном эксперименте философов “мозг в колбе” мозг должен какое-то время прожить в теле, чтобы произвести настройку аппаратного обеспечения в соответствии с программным, представляющим собой годы опыта. Если вы “просто” сконструируете мозг с нуля, позволив элементам установить случайные связи (иными словами, подключите новый мозг произвольным образом), вероятность того, что этот мозг будет иметь какую-либо компетенцию и генерировать логичные идеи, намерения и проекты, обладая при этом какими-либо (очевидными) воспоминаниями, будет **Бесконечно мала** (см. главу 35). См. также главу 32.

*Экспериментальный образец* – это полезный, но опасный элемент инженерного жаргона: когда вас просят сконструировать очень сложное, требующее тонкой настройки устройство, вы создаете очень простое устройство – решая тем самым так называемую *модельную задачу*, – а когда оно справляется с упрощенной задачей, вы объявляете, что создали “экспериментальный образец” и теперь нужно просто *масштабировать* его до полностью работающего устройства, которое требовалось изначально, и для этого необходимы лишь время и деньги, поскольку самая сложная, “концептуальная” задача уже решена. Иногда все так и получается.

Не всем по вкусу эта мысль! Она высмеивается в недавно вышедшей книге “Философские основы нейронауки” (2003), написанной нейробиологом Максом Беннетом и философом Питером Хакером. Подробное опровержение (и их опровержение моего опровержения) можно найти в работе Bennett et al. 2009.

Обзор моделей см. в моей статье “Логическая география вычислительных методов. Взгляд с Восточного полюса” (1986). Первую полную антологию можно найти в работе McClelland, Rumelhart, and PDP Research Group (1986).

Да, рисунки возбуждения на затылочной (зрительной) коре (и в других зонах коры головного мозга) действительно принимают (примерную) форму объекта, находящегося у вас перед глазами, но *смотрит ли* хоть что-нибудь в вашем мозге на эти формы? Возможно, что-то *вроде как смотрит* на эти формы. И все же к этой мысли нужно относиться с осторожностью.

Благодарю своего коллегу Джорджа Смита, который познакомил меня с регистровыми машинами, когда в середине 1980-х гг. мы вместе читали вводный курс лекций о компьютерах в Университете Тафтса. Он разглядел огромный педагогический потенциал регистровых машин и нашел способ объяснить принцип их работы, которым я воспользуюсь и здесь, обращаясь к несколько другой аудитории. Мастерская учебных программ, которую мы с Джорджем основали в Университете Тафтса, выросла именно из этого курса лекций.

Решения задач и упражнений ищите в приложении в конце книги.

Это означает, что остальные упражнения обязательны! Я не шучу. Если вы хотите с пользой применять этот инструмент мышления, вам нужно практиковаться как можно больше, пока вы не овладеете им в совершенстве. Работа над простыми обязательными упражнениями, возможно, займет у вас лишний час или два, но оно того стоит.



Ада Лавлейс, дочь поэта лорда Байрона, была выдающимся математиком – и не только. В 1843 г. она опубликовала свой перевод итальянских комментариев к аналитической машине Бэббиджа и дополнила его своими заметками, которые оказались длиннее и подробнее того, что она перевела: *Menabrea* (1842). В ее заметки вошла и тщательно проработанная инструкция по использованию машины Бэббиджа для вычисления чисел Бернулли. Именно поэтому Лавлейс часто называют первым программистом в истории.

Попытки определить эти уровни совершаются в рамках исследований когнитивной нейробиологии и вычислительной нейробиологии. Две этих области различаются в основном уклоном: специалисты по вычислительной нейробиологии сосредоточены на создании работающих (компьютерных) моделей своих идей, а специалисты по когнитивной нейробиологии часто довольствуются набросками высокоуровневых шаблонов способностей и взаимодействий, которые должны внедряться в любую модель более низкого уровня. В компьютерной науке подобная война идет годами: на одном полюсе стоят ИИ-визионеры, которые не пишут работающие программы, а демонстрируют факты о *спецификациях* любой успешной программы для выполнения той или иной задачи, а на другом полюсе – прожженные инженеры, которых сложно впечатлить, пока они не увидят реальный работающий код. Остальное, по их мнению, нельзя назвать программным обеспечением – это лишь “иллюзорное обеспечение” (англ. *vaporware*). Но их нетерпимость к самым склонным к теоретизированию коллегам не идет ни в какое сравнение с враждебностью, наблюдаемой между специалистами разного уровня по нейробиологии. Один заведующий лабораторией (который специализировался на изучении кальциевых каналов в аксонах нейронов) однажды сказал мне: “В нашей лаборатории говорят: работая над одним нейроном, ты занимаешься нейробиологией. Работая над двумя нейронами – психологией”. И это был не комплимент! Поскольку когнитивная нейробиология забирает львиную долю внимания прессы (все в восторге от новых открытий в сфере зрительных иллюзий, памяти, сознания, речи и понимания, но не каждый интересуется сотнями разных нейромодуляторов и их рецепторов, нейрон-астроглиальными взаимодействиями... или кальциевыми каналами), специалисты по вычислительной нейробиологии нередко страдают от профессиональной ревности по отношению к специалистам по когнитивной нейробиологии.

Мой коллега Рэй Джекендофф (1993) называет это парадоксом изучения языка: дети каким-то образом без труда усваивают и используют правила грамматики, которые до сих пор не могут сформулировать профессиональные лингвисты.

На самом деле великое воображение Декарта позволило ему хотя бы приблизиться к нескольким прекрасным идеям. Он представил, что провода могут открывать крошечные клапаны или поры и выпускать потоки сдерживаемых “животных духов” (спинномозговой жидкости), которые затем выполняют некоторую гидравлическую работу, – чем не предвестник электронного усилителя? Более того, он заметил, что по крайней мере часть разумных (уместных, соответствующих) действий такое устройство может совершать механически, и описал тем самым автоматический *рефлекс*. Примером может служить то, как человек отдергивает руку от пламени, когда жар тянет за соответствующий проводок: все это лишь механика мозга, разум здесь не нужен!

Отступление для специалистов: СYC (Lenat, Guha 1990), бесспорно, самый впечатляющий проект по использованию в искусственном интеллекте (ИИ) подобной языку мысли системы – энциклопедическая (англ. *enCYClopedic*), по большей части вручную спроектированная база данных, управляемая собственным механизмом логического вывода. СYC разрабатывается более четверти века с привлечением множества рук, но ее сила объясняется категорически небиологическим и непсихологическим дизайном. (См. великолепную статью о СYC в “Википедии”.) Пунктик Майка насчет рыжих, вероятнее всего, нельзя считать аксиоматизированной микротеорией о рыжих, написанной на языке мысли и внедренной в большую базу данных наподобие СYC. Мы еще не знаем, на что способны пунктики подобного рода, поскольку еще не изучали их непосредственно – но изучали в очень простых моделях (таких, как архитектуры поглощения, разработанные с учетом анатомии насекомых Родни Бруксом и его коллегами – см. Brooks 1987). В теоретическом отношении проект Брукса *Cog* (по созданию человекоподобного робота – см. Dennett 1994b) был интересен прежде всего тем, что он втолкнул эти совершенно непропозициональные модели содержательных структур в сферу человеческой психологии.

Прекрасный дверной упор получается из тяжелого утюга вроде того, каким гладила одежду ваша бабушка. Если вы найдете такой утюг в антикварном магазине или на сайте “коллекционеров”, убедитесь, что покупаете настоящую вещь, а не копию, потому что на продажу сегодня выставляется много современных дверных упоров, отлитых в форме старинных тяжелых утюгов. Лет через сто кто-нибудь, возможно, займется производством оригинальных дверных упоров, как две капли воды похожих на четвертачки – те самые устройства, которые ваши прадеды использовали тогда, когда деньгами служили монеты.

Философ Хилари Патнэм (1975) изобрел Землю-Двойник много лет назад, и мой насос интуиции воспроизводит все важные детали его концепции. Мой рассказ о четвертачнике, по сути, представляет собой перенастройку рычагов предложенного Патнэмом насоса интуиции. На протяжении последних тридцати пяти лет философы обсуждали десятки и даже сотни других его вариаций. В оригинальном варианте Патнэма речь шла о воде на Земле и эрзац-воде (с составом XYZ, а не H<sub>2</sub>O) на Земле-Двойнике, но в таком случае возникает ряд сложностей, не имеющих отношения к нашему применению этого насоса интуиции, поэтому пускай шмошади в нашем примере станут аналогами четвертаков бальбоа.

Перевод А. Сергеева. – *Прим. пер.*



В знаковой работе Текумсе Фитча “Наноинтенциональность. В защиту внутренней интенциональности” (2008) делается предположение, что клетки-эукариоты – но не их предки-прокариоты – были первыми сущностями, обладающими “внутренней” интенциональностью, поскольку их способность к самозащите была существенно более развитой, чем подобная способность их предков. Текумсе Фитч подчеркивает автономию, то есть субъектность, отдельных клеток, что и сподвигло меня на пересмотр гомункулярного функционализма (см. главу 20), однако я не могу согласиться с его попыткой остановить конечный регресс на эукариотах. Прокариоты, как и эукариоты, не могут быть “заменены машиной”, а следовательно, субъектность доходит до белков и достигает нижней точки на эгоистичных генах.

Категоричное и страстное отрицание всего, что написано в этом абзаце, см. в книге Фодора и Пьятелли-Пальмарини “В чем ошибался Дарвин” (2010). Либо в корне неправы они, либо автор книги, которую вы читаете. Более подробная аргументация в поддержку изложенной в настоящей книге позиции приводится в части VI, а также в моей книге “Эволюция причин” (готовится к публикации).

Рут Милликен развивает эту мысль лучше меня в серии блестящих книг, первая из которых вышла в 1984 г. О современном состоянии вопроса см. в новой работе *Millikan and Her Critics* (Ryder et al. 2013).

Как сказал философ Джон Хогланд (1981, р. 23), первый принцип ИИ гласит: “Если позаботиться о синтаксисе, *семантика позаботится о себе сама*”. Можно по-разному трактовать эту фразу, но в своем первом, чересчур многообещающем варианте, она спровоцировала работу по созданию гигантской базы данных, аксиоматизированной формализации мирового знания, которую сможет обслуживать и использовать (исключительно синтаксическая) машина вывода. (Лучший пример этого – СУС.) Большинство экспертов полагает, что на практике это неисполнимо, но фраза живет, поскольку прекрасно выражает идею о том, что мозг представляет собой своеобразный компьютер (а следовательно, синтаксический движок), который, благодаря своей конструкции, в общих чертах выполняет работу семантического движка.

Компьютер *IBM Watson* появился как раз вовремя, чтобы моя научная фантастика фактически стала научным фактом. *Watson* можно считать ящиком А, в то время как в ящике Б содержится шведский аналог *Watson*, независимо разработанный Бо. Когда я впервые опубликовал описание этого мысленного эксперимента (Dennett 1995a), я мог лишь сказать: “Реальным примером подобного проекта можно считать масштабный проект Дуласа Лената СҮС (от англ. *encyclopedia*), разрабатываемый компанией MCC (Lenat, Guha 1990)”. См. также сноску о СҮС в главе 28. В 1995 г. невозможно было представить, что ИИ станет таким, какой мы сейчас наблюдаем в *Watson*, но его развитие шло большими шагами. В отличие от проекта СҮС, код которого написан в основном вручную, *Watson* может самостоятельно штудировать интернет в поисках истин, а также эффективно использовать статистические характеристики доступных данных. И *Watson*, и СҮС по-своему вроде как понимают данные в своих базах – им доступно гораздо, гораздо более глубокое понимание, чем другим компьютерам с большими базами данных.

Когда я впервые описал этот насос интуиции, формат ASCII (*American Standard Code for Information Interchange*; “американская стандартная кодировка для информационного обмена”) был стандартом почти для всей текстовой обработки, электронной почты и интернета. Позднее ей на смену пришла расширенная, обратно совместимая кодировка UTF-8 (*Universal [Character-set] Transformation Format – 8 bit*; “универсальный [побуквенный] формат преобразования – 8 бит”). Таким образом, ASCII по-прежнему остается частью UTF-8.

Только для философов: кое-кто утверждал, что мое представление о паттернах, изложенное в статье “Реальные паттерны” (1991b), представляет собой *эпифеноменализм* о содержимом. Это мой ответ.

Выдвинув эту гипотезу, марсиане могут заняться вариацией на тему **радикального перевода** Куайна, но их задача будет вдвое сложнее, поскольку они не смогут выяснять у своих “информаторов”, А и Б, названия предметов, показывая их и спрашивая: “Мяч? Карандаш?”



Такую же мораль можно вывести из интерпретации фактов эволюционной истории. Даже если вы в состоянии в мельчайших подробностях описать каждый каузальный факт в истории каждого жирафа, который когда-либо жил на земле, вы ни за что не сумеете объяснить наблюдаемые закономерности – к примеру, появление у жирафов длинной шеи, – если только не подниметесь на уровень-другой и не спросите “Почему?”, ища *мотивы* Матери-Природы. (Подробнее об этом в части VI.)

У меня свои представления о том, какой должна быть адекватная научная теория смыслов: на мой взгляд, на место ригидных, булевых (логических) конструкций GOFAI (“старого доброго искусственного интеллекта”) должны прийти более гибкие, байесовские (статистические, вероятностные) сети поиска паттернов, однако сейчас не время и не место для разработки этих идей.

Некоторым показалось, что мое сравнение идеи Дарвина с универсальной кислотой вдохновлено ДНК, ведь она есть не что иное, как дезоксирибонуклеиновая кислота, но я имел в виду нечто более универсальное: ДНК – не единственный агент эволюции на нашей планете, и как знать, какие еще примеры эволюции таит наша вселенная?

Борхес называет немного другие цифры: в каждой книге 410 страниц, на каждой странице 40 строк, а в каждой строке 80 знаков. Общее число знаков в книге достаточно близко к моему (1 312 000 против 1 000 000), чтобы разница не имела значения. Я взял более круглые цифры, чтобы облегчить подсчеты. Борхес ограничился набором из 25 символов, которых достаточно для прописных букв испанского алфавита (а также пробела, запятой и точки в качестве единственных знаков препинания), но недостаточно для английского алфавита. Я взял набор из 100 символов, чтобы в нем наверняка нашлось место для прописных и строчных букв, а также знаков препинания всех языков с латинским алфавитом.



Вавилонская библиотека конечна, но весьма любопытно, что в ее стенах содержатся все грамматически верные фразы английского языка. Однако их количество бесконечно, а библиотека конечна! И все же любую английскую фразу, какой бы длинной она ни была, можно разбить на фрагменты длиной в пятьсот страниц, каждый из которых хранится где-то в библиотеке! Как это возможно? Некоторые книги могут использоваться более одного раза. Проще всего понять самый вопиющий пример: поскольку некоторые книги содержат только один символ на множестве пустых страниц, многократное использование этих ста книг позволит создать любой текст любой длины. Как Куайн (1987) заметил в своем информативном и любопытном эссе “Универсальная библиотека”, если вы прибегнете к стратегии повторного использования книг и переведете все содержимое в кодировку ASCII, используемую вашим текстовым процессором, вы сможете уместить всю Вавилонскую библиотеку в двух чрезвычайно тонких томах, в одном из которых будет напечатан 0, а в другом – 1! (Куайн также указывает, что идею об универсальной библиотеке задолго до Борхеса предложил психолог Теодор Фехнер.)

Куайн (1987) с той же целью вводит термин “гиперастрономический”.

“Долина смеха” (De Vries 1953). Далее следует: “Нисколько не стесняйтесь. Зовите меня в любое время дня и ночи...” Возможно, Де Врис также изобрел игру, чтобы продемонстрировать, насколько сильное воздействие (хоть пагубное, хоть нет) может оказать одно лишь типографское изменение. Один из лучших примеров: *Whose woods are these, I think I know; his house is in the Village though* (игра слов: в оригинале – “Я знаю, чьи это леса, их хозяин живет в деревне”; в измененном варианте – “Я знаю, чьи это косяки, их хозяин живет в Виллидж”). Игру быстро подхватили: в естественном состоянии, говорит нам мутант-Гоббс, *the wife of man, solitary, poore, nasty, brutish, and short* (игра слов: в оригинале – “жизнь человека одинока, бедна, беспросветна, тупа и кратковременна”; в измененном варианте – “жена человека одинока, бедна, ворчлива, тупа и низкоросла”). Или вот вам вопрос: *Am I my brothel's keeper?* (игра слов: в оригинале – “Разве я сторож брату моему?”; в измененном варианте – “Разве я сторож борделю моему?”).



Не совсем невозможно. Геостационарные спутники, которые движутся по орбите в унисон с вращением Земли, представляют собой своего рода реальные, а не волшебные небесные крюки. Их ценность – и финансовая устойчивость инвестиций в них – объясняется тем, что нам часто очень хочется прикрепить что-либо (например, антенну, камеру или телескоп) к определенной точке в небе. Увы, спутники бесполезны для *подъема* грузов, поскольку они находятся слишком высоко. Тем не менее эта идея рассматривалась. Как выяснилось, чтобы веревка из самого прочного искусственного волокна, когда-либо создававшегося на земле, выдержала хотя бы собственный вес, не говоря уже о грузе, наверху ее диаметр должен составлять более ста метров, пускай даже она будет истончаться до едва заметной лески. Даже если вам удастся протянуть такой кабель, вряд ли вам захочется, чтобы он случайно упал с орбиты на какой-нибудь город!

Я обрадовался, когда изобрели *небесный кран*, который должен был помочь опустить марсоход “Кьюриосити” на поверхность Марса 6 августа 2012 г., и особенно меня обрадовал тот факт, что его назвали небесным *краном*, а не небесным *крюком*, поскольку он представляет собой продукт блестящей инженерии, а не чудо.

Обратите внимание на серьезное различие в наследии Митохондриальной Евы и Y-хромосомного Адама: у всех нас, мужчин и женщин, в клетках есть митохондрии, которые мы унаследовали от матерей; если вы мужчина, то у вас есть Y-хромосома, которую вы унаследовали от отца, в то время как почти все – но не совсем все – женщины не имеют Y-хромосомы вовсе.

Отбросив логические возможности в сторону и посмотрев на фактический ход истории, мы увидим, что некоторые мужчины пытались монополизировать оплодотворение в своих популяциях и, судя по всему, иногда поразительно преуспевали в этом: могущественные короли и воины оплодотворяли сотни женщин. Плохо ли, хорошо ли это, но их вклад в человеческий генофонд, вероятно, непропорционально велик.

Дэйл Петерсон обратил мое внимание на тот факт, что отбойники шимпанзе постепенно оттачиваются в процессе непрерывного использования на протяжении долгого времени. Это дает нам прекрасное “недостающее звено” в серии, которая началась с того, что случайный камень был использован, чтобы расколоть орех, после чего для этой цели был выбран оказавшийся поблизости камень самой удачной формы, после чего были организованы поиски более удобного камня, после чего было замечено (невооруженным глазом, с применением зрительной памяти), что любимый камень становится тем лучше, чем дольше он используется, и наконец был достигнут тот уровень развития, когда камень стал специально обтесываться для конкретной задачи.

От англ. *futile* – бесполезный, тщетный. – Прим. пер.

Роль абиотических циклов в изменении вероятности появления репродуктивных клеток более подробно обсуждается в моей книге “Эволюция причин” (готовится к публикации).

Вот тот же самый фрагмент на немецком, напечатанный более читабельным шрифтом: *Freunde, Römer, Mitbürger, gebt mir Gehör! Ich komme, Cäsars Leiche zu bestatten, nicht, ihn zu loben.*



Этим прекрасным словом программисты называют ошибку кодирования семантического, а не синтаксического уровня. Если вы забыли поставить скобку – это опечатка; если забыли задать внутреннюю переменную – одумка. *Одумки* возможны в любой человеческой деятельности, интерпретируемой семантически или интенционально и имеющей четкие каноны правильности или эстетики. Назвать даму блаженной, а не благородной – значит, допустить одумку, а не опечатку. См. Dennett 2006b.

Грубо говоря, формат MIDI в музыке сродни кодировке ASCII в обработке текста – это лингва франка для общения компьютерных программ с внешним миром.

Один из читателей раннего черновика этой книги решил, что я все это выдумал. Если вас терзает такое же подозрение, ознакомьтесь с обстоятельной и чрезвычайно серьезной статьей о философских зомби в *Stanford Encyclopedia of Philosophy*, доступной онлайн: <http://plato.stanford.edu/entries/zombies/>.

Здесь философы снова не преминут обвинить меня в подмене метафизического вопроса эпистемологическим: “Не спрашивайте, *откуда мы узнаем*, что кто-то зомби! Спрашивайте, *что значит* быть зомби!” Такой отрыв метафизики от эпистемологии возможен разве что в воображении.

Перевод Р. Райт-Ковалевой. – *Прим. пер.*

Перевод Е. Суриц. – *Прим. пер.*

Даг прицепился к словосочетанию “клочки бумаги” из эссе Сёрла и показал, как оно подталкивало людей существенно недооценивать масштабы и сложность используемого программного обеспечения. Он включил это замечание в свой комментарий к эссе Сёрла, опубликованный в нашей книге, и Сёрл (1982) резко ответил ему на страницах *New York Review of Books*, поскольку, хотя мы правильно перепечатали его статью, в своем комментарии Даг ошибся и написал “несколько листов”, когда у Сёрла были “клочки”, и это, по утверждению Сёрла, полностью исказило смысл его довода! Если Сёрл прав – если это крошечное неумышленное искажение действительно изменило механику опыта, – в некотором роде это доказывает нашу точку зрения: если столь незначительное изменение может свести мысленный эксперимент на нет, то на это стоит обратить внимание всех тех, кто собирается воспользоваться таким насосом интуиции!

Ознакомьтесь с виртуозной демонстрацией Хофштадтера в книге “Глаз разума” (Hofstadter, Dennett 1981). Показывая, что насос интуиции “Китайская комната” обладает (как минимум) пятью регуляторами, Даг по очереди поворачивает каждый из них, создавая целое семейство мысленных экспериментов, подталкивающих к разным выводам. В заключении я заметил:

Каждый вариант настройки нашего насоса интуиции рождает немного иной нарратив, в котором на задний план отходят иные проблемы и из которого извлекаются иные уроки. Чтобы понять, какой версии – или каким версиям – доверять, необходимо внимательно изучить каждую из них и определить, какие особенности нарратива выполняют основную работу. Если чрезмерное упрощение *пугает* интуитивные озарения, а не просто помогает избавиться от не относящихся к делу сложностей, нам не следует доверять напрашивающимся выводам. [p. 460]



Как часто замечается, эта убежденность вторит убежденности Декарта, который еще в 1637 г. в своем “Рассуждении о методе” сказал, что легче всего отличить машину от человека с нематериальной душой, если вступить с ней в беседу.

Можно представить себе, что машина сделана так, что произносит слова – и даже слова, связанные с физическим воздействием или объектами, которые провоцируют изменения в ее органах, как, например, если тронуть ее в каком-нибудь месте, и она спросит, что от нее хотят, тронуть в другом – закричит, что ей больно, и так далее. Но никогда она не сумеет построить фразы таким образом, чтобы ответить на сказанное в ее присутствии, хотя на это способны даже самые глупые люди.

Неизвестно, вдохновлялся ли Тьюринг насосом интуиции Декарта при создании собственного.

Спасибо лингвисту, китаеведу и ученому-энциклопедисту Дэвиду Мозеру, который предоставил мне этот пример, появившийся в Китае как раз в тот момент, когда у меня возникла в нем необходимость.

Перевод С. Церетели, В. Васильева, В. Швырева. – *Прим. пер.*

Например: как вы решаете, кто из вас хозяин, а кто – альтеры? как вы объясняете себе тот факт, что ваши воспоминания о недавних событиях состоят из кратких эпизодов присутствия, перемежающихся часами и целыми днями забытья? вам это не кажется подозрительным?

В стихотворной пьесе Т. С. Элиота “Убийство в соборе” Бекета спрашивают о событии из его прошлого и он отвечает:

Те времена прошли. Я вспоминаю:

Забвения не стоят. (*Перевод В. Топорова*)

Ученые не сходятся во мнении, можно ли считать гетерофеноменологию предпочтительной методологией для научного изучения сознания. Одни исследователи считают, что она вносит ясность, однозначно определяя условия, в которых наука может изучать сознание, другие полагают, что она просто повторяет очевидное, а третьи упорно отказываются ее принимать. Ряд лучших опубликованных работ на эту тему перечислен в разделе “Источники”.

Если вы сомневаетесь, что внушение подобного убеждения может привести к таким последствиям, ознакомьтесь с работой Vohs, Schooler (2008) и более поздней литературой, в которой приводятся экспериментальные данные.

Подробнее о теоретических следствиях этого компромисса в пространстве и времени см. в работе Dennett 1987, chapter 9.



Совершенно другой взгляд на двумерную физику и инженерию излагается в книге А. Дьюдни “Планиверсум” (1984), в которой автор значительно углубляет идеи, высказанные во “Флатландии” Эдвина Эббота (1884), создавая при этом мощный инструмент мышления.

Чудесный пример безудержного высокоуровневого прогнозирования содержится в фильме “Принцесса-невеста” в сцене, где Уоллес Шоун в роли Виззини пытается перехитрить Кэри Элвеса в роли Уэсли, когда тот предлагает ему угадать, в каком из кубков яд, но в итоге переигрывает самого себя. См. [http://www.dailymotion.com/video/xhr71a\\_never-go-in-against-a-sicilian-when-death-is-on-the-line\\_shortfilms](http://www.dailymotion.com/video/xhr71a_never-go-in-against-a-sicilian-when-death-is-on-the-line_shortfilms).

На официальном сайте королевской семьи говорится: “Сегодня Корона сохраняет за собой право собственности на всех немеченых лебедей-шипун, обитающих в открытой воде, но королеве принадлежат только лебеди, обитающие на определенных отрезках Темзы и ее притоков. Настоящее право собственности делится с Почтенным обществом виноделов и Почтенным обществом красильщиков, которые получили право собственности по королевскому указу в пятнадцатом веке. Само собой, сегодня лебедей не едят”.

Работа Хейса вдохновила и других. См. антологию под редакцией Боброу (1985). Авторы одного из эссе замечают: “Термин «наивная физика» неоднозначен. Что это – плохая физика? Психология? Искусственный интеллект? Физика?” (Bobrow 1985, p. 13). На мой взгляд, наивная физика – это попытка ввести физику в нашу манифестную картину мира, вполне допускающую автоматическое дедуктивное мышление.

Что касается другой задачи из этой книги, ответы таковы: *slop, love, even, (Laura) Dern*, а также *smut, hope, iron, (Sean) Penn*.

Первого апреля 2012 г. в интернете появился прекрасный шахматный розыгрыш (*Rajlich: Busting the King's Gambit* // ChessBase News, April 2, 2012: <http://chessbase.com/newsdetail.asp?newsid=8047>). Утверждалось, что гигантский консорциум компьютеров, круглосуточно работавший на протяжении нескольких месяцев, доказал, что королевский гамбит (дискредитированный, но не “опровергнутый” дебют) неизменно приводит к поражению, если только не соблюдается одно маловероятное условие. Я попался на удочку – и мне особенно неловко из-за этого, поскольку ранее я писал о другом розыгрыше Мартина Гарднера (во введении к работе Hurley, Dennett, and Adams 2011). Розыгрыш объясняется в статье *The ChessBase April Fools Prank* // ChessBase News, April 4, 2012: <http://chessbase.com/newsdetail.asp?newsid=8051>.