



# МЕТА- МЫШЛЕНИЕ

КАК НЕЙРОНАУКИ  
ПОМОГАЮТ НАМ  
ПОНЯТЬ СЕБЯ

СТИВЕН М. ФЛЕМИНГ

**Стивен М. Флеминг**  
**Метамышление. Как**  
**нейронауки помогают**  
**нам понять себя**

*[http://www.litres.ru/pages/biblio\\_book/?art=68808756](http://www.litres.ru/pages/biblio_book/?art=68808756)*

*Метамышление. Как нейронауки помогают нам понять себя:*

*ISBN 978-5-6048294-1-7*

### **Аннотация**

Где бы мы ни оказались – на экзамене или приеме у врача, в супермаркете или за рулем машины, – нам постоянно приходится принимать решения. Стоит ли доверять поставленному диагнозу или рекламе «полезных» сладостей? Точно ли мы помним ответ на вопрос в билете? Уверены ли, что успеем проскочить на желтый? И вообще, откуда мы знаем, что знаем или, что не менее важно, не знаем?

В этой книге когнитивный нейробиолог Стивен М. Флеминг подробно и понятно рассказывает о метапознании – способности обдумывать свои мысли. словно дирижер, направляющий музыкантов, оно незаметно влияет на все наши действия. И тренируя этот навык, мы лучше понимаем себя, других людей и окружающий нас мир. Почему нам порой легче поверить незнакомцу, чем своим детским воспоминаниям? Что делать,

когда не получается вспомнить слово, которое вертится на языке? Как работает эскапизм и почему мы забываем обо всем, погружаясь в сюжет захватывающего фильма или игры? Нейробиологи нашли новый ответ на античный призыв «познать самого себя», и он обещает не только философскую, но и практическую пользу.

В формате PDF A4 сохранен издательский макет книги.

# Содержание

Предисловие	6
Часть I	26
Глава 1	26
Обратные задачи и способы их решения	28
Построение моделей мира	39
Используем неопределенность, чтобы сомневаться	47
Конец ознакомительного фрагмента.	53

**Стивен М. Флеминг**  
**Метамышление**  
**Как нейронауки помогают**  
**нам понять себя**

© 2021 by Stephen M. Fleming

© Bakhtiar Alfani / Shutterstock.com, фото на обложке

© 2021 by Hachette Book Group, Inc, дизайн обложки

© Д. Виноградов, перевод, 2023

© ООО «Индивидуум Принт», 2023

*Хелен и Финни*

# Предисловие

Представьте, что вы пришли на прием к врачу, чтобы обсудить беспокоящую вас боль в груди. Вы сдаете анализы крови, проходите обследование и через неделю возвращаетесь в клинику, где врач разбирает с вами результаты. Проблема выглядит серьезной, и доктор поспешно рекомендует операцию по шунтированию сердца. Когда вы спрашиваете, почему она уверена в необходимости этой операции, врач рассказывает вам о ходе своих размышлений, в том числе о том, что она, возможно, ошибается, и о том, к чему может привести эта ошибка, после чего повторяет свой совет сделать операцию. Как бы вы поступили?

Теперь представьте, что после того, как вы прошли обследование, результаты поступают в компьютерную программу с искусственным интеллектом, которая с уверенностью заключает, что проблема выглядит серьезной, и рекомендует провести операцию по шунтированию сердца. Когда вы спрашиваете своего врача, действительно ли операция необходима, она не может вам ответить. Она не знает, почему была дана такая рекомендация. Все, что она может сказать, – в прошлом искусственный интеллект был очень точен в своих рекомендациях, когда получал весь перечень результатов, поэтому было бы разумно довериться ему и согласиться на операцию. Как бы вы поступили на сей раз?

Пожалуй, в первом случае ответ кажется очевидным: если врач уверен в себе и может объяснить причины этой уверенности, вы чувствуете, что стоит довериться ее совету. Во втором случае, однако, все не так однозначно. Многие из нас интуитивно ощущают, что, если человек или машина будут принимать ответственные решения, которые касаются нас, мы должны иметь возможность попросить их объяснить, *почему* они пришли к тому или иному выводу. Многие из наших правовых устоев, которые возлагают ответственность и вину за ошибки, основаны на возможности обосновать и защитить наши действия и мотивы. Без этих объяснений нам пришлось бы слепо доверять друг другу или нашим машинам. Как ни странно, некоторые из наиболее эффективных алгоритмов машинного обучения хуже всего поддаются объяснению. В отличие от них, люди охотно объясняют, что и почему они делают, – и все благодаря нашей способности размышлять, задумываться и знать самих себя, то есть отдавать себе отчет в том, как мы познаем, решаем, помним, мыслим и чувствуем.

У психологов есть особое название для этого вида самосознания: метапознание – от греческого «мета», что переводится как «после» или «за пределами». Буквально это означает способность мыслить о собственном мышлении. Метапознание – это изящная, красивая и, откровенно говоря, причудливая особенность человеческого разума, которая увлекала ученых и философов на протяжении веков. В

знаменитой книге «Система природы» (1735) Карл Линней детально описал физические особенности сотен видов. Добравшись до нашего рода *Ното*, он настолько очаровался способностью людей к метапознанию, что просто приписал к их определению одну строчку на латыни: «*nosce te ipsum*» – «те, кто познал себя»<sup>1</sup>.

Самосознание – определяющая черта человеческого опыта. Возьмем студентку Джейн, которая готовится к экзамену по физике. Что происходит у нее в голове? Несомненно, она мысленно перебирает множество фактов и формул, которые ей необходимо понять и усвоить. А еще, возможно не осознавая этого, она принимает решение, как, когда и что ей учить. Какая обстановка ей больше подойдет: оживленное кафе или тихая библиотека? Полезнее ли будет перечитать записи или порешать задачи? Не стоит ли переключиться на другую тему? А может, лучше вообще бросить все это и пойти погулять с друзьями?

Правильные решения, несомненно, повысят шансы Джейн на успех. Ей не хотелось бы попасть впросак, посчитав, что она хорошо усвоила материал, если это не так, или выбрав сомнительную стратегию подготовки. Но никто не примет решения за нее. Ей придется положиться на свое са-

---

<sup>1</sup> Линней К. Система природы. Санкт-Петербург: При Императорской Академии наук, 1804–1805; Flavell J. H. Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry // American Psychologist. 1979. № 34; Nelson T. O. и др. Cognition and Metacognition at Extreme Altitudes on Mount Everest // Journal of Experimental Psychology: General. 1990.

мосознание.

Наша способность к саморефлексии продолжает играть важную роль за пределами классных комнат и экзаменационных аудиторий. Обратимся к опыту Джеймса Нестора, писателя и фридайвера. В своей книге «Глубина» («Deep») Нестор рассказывает, как он ездил в прибрежные районы Греции и Багамских островов, чтобы сделать репортаж о соревнованиях по фридайвингу. На таких турнирах у участников только одна цель: задержать дыхание и погрузиться глубже, чем все остальные. В качестве доказательства, что они достигли определенной глубины, дайверы приносят на поверхность бирку с выбитым на ней номером. Если после всплытия они теряют сознание, погружение не засчитывается. Чтобы достичь успеха, профессиональные фридайверы должны четко осознавать свою способность погружаться на глубину, избегая травм или даже гибели. Легкая неуверенность в себе приведет к посредственным результатам, в то время как даже небольшая *сверхуверенность* может привести к летальному исходу. Показательно, что основная часть подготовки фридайверов проходит на суше и заключается в психологическом исследовании их возможностей и ограничений под водой<sup>2</sup>.

А как насчет случая Джудит Кеппел, одной из первых участниц британского шоу «Кто хочет стать миллионером»?

---

<sup>2</sup> Nestor J. Deep: Freediving, Renegade Science, and What the Ocean Tells Us About Ourselves. Boston: Eamon Dolan, 2014.

В каждом раунде участников спрашивают, уверены ли они в том, что знают правильный ответ, и точно ли хотят рискнуть своим выигрышем ради шанса на более высокий приз, вместо того чтобы просто забрать деньги. Ставки высоки. Ошибиться – значит потерять все, что вы заработали. Джудит приняла решение продолжать, когда на кону стояло 500 000 фунтов стерлингов.

Вопрос на миллион звучал следующим образом: «Какой король был женат на Алиеноре Аквитанской?» После короткой дискуссии с ведущим Крисом Таррантом Джудит остановилась на ответе «Генрих II». Затем Таррант задал свой убийственный, самый мучительный для участников вопрос: «Это ваш окончательный ответ?» И вновь успех зависит от самосознания. Вам нужно понять, насколько вы близки к правильному ответу, прежде чем соглашаться на авантюру. Джудит осталась при своем мнении и стала первым победителем в истории шоу.

Истории Джейн, Джеймса и Джудит объединяет то, насколько сильно их успех или неудача зависели от точности самосознания. Чтобы прочувствовать силу метапознания, представим, как бы сложились их истории, если бы оно сработало плохо. Джейн могла бы опрометчиво посчитать, что задачи по механике жидкости даются ей легко и необязательно задерживаться на этой теме. Она бы думала, что держит все под контролем, даже если бы это было не так. Подобная метакогнитивная ошибка может привести к про-

валу на экзамене, несмотря на неплохие способности Джейн и ее усердную учебу. В случае с Джудит мы можем выделить два типа возможных метакогнитивных ошибок: она знала бы правильный ответ, но не осознавала этого и поэтому упустила бы возможность стать миллионером. Или же она проявила бы излишнюю самоуверенность, рискнув поставить все на неправильный ответ. Для Джеймса подобная самоуверенность могла бы стать вопросом жизни и смерти. Если бы он подумал, что способен совладать с большей глубиной, чем это было в действительности, то, подобно подводному Икару, спустился бы чересчур далеко и осознал свою ошибку, только когда было бы уже слишком поздно.

Мы часто упускаем из виду, насколько сильно метапознание влияет на нашу жизнь как в позитивном, так и в негативном ключе. В отличие от способности решать уравнения, совершать спортивные подвиги или запоминать исторические факты, самосознание представляется менее важным качеством. Для большинства из нас метапознание – словно дирижер оркестра: время от времени он незаметно корректирует и направляет музыкантов в верном (или неверном) направлении, при этом оставаясь не замеченным и недооцененным публикой. Даже если дирижера не будет на месте, оркестр продолжит играть – равно как Джейн, Джеймс и Джудит продолжили бы готовиться к экзамену, погружаться под воду и отвечать на вопросы игрового шоу, даже если бы их самосознание временно отключилось. Однако именно хороший

дирижер способен стать тем фактором, что отделит заурядную репетицию от концерта мирового уровня, аналогично тому, как малозаметное влияние метапознания может отделить успех от неудачи, а жизнь – от смерти.

Исторически сложилось, что самосознание трудно измерить, определить и изучить, его роль зачастую игнорируется в том числе и по этой причине. Но сейчас ситуация меняется. Завесу над саморефлексией человеческого разума приоткрывает новая научная отрасль – метакогнитивная нейронаука. Сочетая инновационные лабораторные эксперименты с новейшими технологиями нейровизуализации, мы получаем все более детальное представление о работе самосознания как когнитивного и как биологического процесса. Дальше мы с вами увидим, наука о метапознании может продвигать нас как никогда далеко в познании самих себя<sup>3</sup>.

## **Создание науки о самосознании**

Загадка самосознания привлекла меня еще в подростковые годы, когда я заинтересовался книгами о мозге и разуме. Помню, как во время летних каникул, лежа у бассейна, я оторвал взгляд от страницы одной из этих книг и задумал-

---

<sup>3</sup> Shimamura A. C. Toward a Cognitive Neuroscience of Metacognition // *Consciousness and Cognition*. 2000. Т. 9. № 2. С. 313–323; Fleming S. M. и др. Domain-Specific Impairment in Metacognitive Accuracy Following Anterior Prefrontal Lesions // *Brain*. 2014. № 10. С. 2811–2822.

ся: почему простая активность мозговых клеток в моей голове должна приводить к *этому* уникальному переживанию – восприятию света, мерцающего на поверхности воды? И что еще важнее: каким образом тот же самый мозг, в котором происходит этот опыт, вообще позволяет мне размышлять о таких вещах? Одно дело – быть сознательным, другое – знать, что я сознателен, и думать о своей сознательности. Голова пошла кругом. Я попался на крючок.

Сегодня я руковожу нейролабораторией по исследованию самосознания в Университетском колледже Лондона. Моя группа – одна из нескольких, работающих в Центре нейровизуализации человека Wellcome, расположенном в элегантном таунхаусе на Квин-сквер в Лондоне<sup>4</sup>. На цокольном этаже нашего здания размещены массивные аппараты для сканирования мозга, которые каждая из групп Центра использует для изучения различных аспектов работы сознания и мозга – как мы видим, слышим, запоминаем, говорим и так далее. В моей лаборатории студенты и постдоки изучают способность мозга к самосознанию. Тот факт, что какой-то уникальный элемент биологического устройства человека наделил нас способностью обращать мысли на самих себя, кажется мне удивительным.

Однако еще совсем недавно все это представлялось бессмыслицей. Как отмечал французский философ девятнадцатого века Огюст Конт, «мыслящий индивид не может разде-

---

<sup>4</sup> The MetaLab, <https://metacoglab.org>.

лить себя надвое – одна половина рассуждает, а другая наблюдает. Поскольку если наблюдаемый и наблюдающий орган будут идентичны, как может быть сделано какое-либо наблюдение?»<sup>5</sup>. Иными словами, как может один и тот же мозг обращать свои мысли на самого себя?

Аргумент Конта был созвучен научному мышлению того времени. С наступлением эпохи Просвещения в Европе все более популярным становился взгляд на самосознание как на нечто особенное и непостижимое для науки. Саморефлексию же западные философы использовали в качестве философского инструмента, подобно тому как математики задействуют алгебру для поиска новых математических истин. Опираясь на самоанализ, Рене Декарт пришел к своему знаменитому заключению «Я мыслю, следовательно, я существую», отметив при этом, что «ничто не может быть воспринято мною с большей легкостью и очевидностью, нежели мой ум». Декарт предполагал, что душа является вместилищем мысли и разума и что она приказывает нашему телу действовать. Душа не может быть разделена надвое – она просто *есть*. По этой причине самосознание казалось загадочным и неопределимым явлением, недоступным для науки<sup>6</sup>.

Сегодня мы уже знаем, что у Конта не было оснований для беспокойства. Человеческий мозг не является единым, неделимым органом. Напротив, он состоит из миллиардов кро-

---

<sup>5</sup> *Конт О.* Общий обзор позитивизма. Ленанд, 2019.

<sup>6</sup> *Декарт Р.* Рассуждения о методе. М.: АСТ, 2014.

шечных нейронов, потрескивающих от электрической активности и сплетенных в схему умопомрачительной сложности. Из взаимодействия между этими клетками и складывается вся наша психическая жизнь: мысли и чувства, надежды и мечты, которые то вспыхивают, то угасают.

Эта электрическая схема отнюдь не спутанный клубок связей без видимой структуры, напротив, она характеризуется протяженной архитектурой, подразделяющей мозг на отдельные области, каждая из которых занимается специальными вычислениями. Не обязательно отмечать на карте города все дома, чтобы ею можно было пользоваться. Схожим образом, мы можем получить приблизительное представление о работе различных областей человеческого мозга в масштабе целых участков, а не отдельных клеток. Некоторые области коры головного мозга находятся ближе к местам поступления данных (например, глазам), а другие – дальше по цепочке обработки информации. Например, какие-то области в основном занимаются зрением (зрительная кора, расположенная в задней части мозга), другие – обработкой звуков (слуховая кора), а третьи – хранением и извлечением воспоминаний (например, гиппокамп).

Отвечая Конту в 1865 году, британский философ Джон Стюарт Милль предвосхитил мысль о том, что самосознание может зависеть от взаимодействия процессов, протекающих в пределах одного мозга, и, таким образом, является легитимным объектом научного исследования. Благодаря

ря появлению мощных технологий нейровизуализации, таких как функциональная магнитно-резонансная томография (фМРТ), сегодня нам известно, что процесс саморефлексии действительно активизирует определенные нейронные сети, а повреждения этих сетей (в том числе вследствие заболевания) могут привести к критическим нарушениям самосознания<sup>7</sup>.

## Познай себя лучше

Я часто думаю, что если бы мы так сильно не привыкли к нашей способности к самосознанию, то были бы потрясены тем, что мозг способен проделывать этот удивительный фокус. Представьте на мгновение, что вы – исследователь, отправившийся изучать новые формы жизни на далекую планету. Биологам на Земле не терпится узнать, как они устроены и что ими движет. Но никто и не подумает просто спросить их об этом! А вот марсианин, приземлившийся на Землю и подучивший английский, испанский или французский, может поступить именно так. Он будет шокирован, узнав, что мы способны рассказать ему кое-что о том, каково это – помнить, мечтать, смеяться, плакать, испытывать радость или сожаление, и все это благодаря тому, что мы осознаем

---

<sup>7</sup> *Mill J. S. Auguste Comte and Positivism: Reprinted from the Westminster Review. London: N. Trübner, 1865.*

себя<sup>8</sup>.

Но самосознание появилось у нас не только для того, чтобы мы могли рассказывать друг другу (и возможным гостям с Марса) о своих мыслях и чувствах. Самосознание играет центральную роль в нашем восприятии мира. Мы не просто воспринимаем окружающее нас пространство, а можем размышлять о красоте заката, сомневаться в ясности нашего зрения и раздумывать, не обманывают ли наши чувства иллюзии или трюки фокусника. Мы не просто принимаем решения о том, какую работу предпочесть или за кого выходить замуж, а можем размышлять о том, хороший или плохой выбор мы сделали. Мы не просто воскрешаем в памяти наши детские воспоминания, но можем сомневаться в их достоверности.

Кроме того, самосознание позволяет нам понять, что у других людей тоже есть сознание, подобное нашему. Осознавая себя, я получаю возможность спросить: «Как я это вижу?» – и, что не менее важно, – «Как это видят другие?» Художественная литература потеряла бы смысл, если бы мы утратили способность думать о сознании других людей и сравнивать их опыт с нашим собственным. Без самоанализа не было бы и системного образования. Мы бы не знали, кому нужно обучаться и способны ли мы обучать других. Мысль о том, что самосознание является катализатором челове-

---

<sup>8</sup> Dennett D. C. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books, 1996.

ского процветания, изящно выразил Владимир Набоков:

[Нас отличает от животных] то, что мы понимаем, что разумеем что-то о бытии. Другими словами, если я осознаю не только то, что я есмь, но еще осознаю, что осознаю это, значит, я отношусь к роду человеческому. Все прочее лишь вытекает из этого: блеск мысли, поэзия, мироощущение. В этом плане разрыв между обезьяной и человеком неизмеримо больше, чем между амебой и обезьяной<sup>9</sup>.

В свете этих бесчисленных преимуществ неудивительно, что развитие самосознания издавна считалось разумной и достойной целью. В диалоге Платона «Хармид» Сократ возвращается с Пелопоннесской войны. По дороге домой он спрашивает местного мальчика, Хармида, понял ли тот, что значит *софросюне* – греческое слово, означающее сдержанность или умеренность, а также сущность правильно прожитой жизни. После долгих споров двоюродный брат мальчика Критий говорит, что ключ к *софросюне* прост – самосознание. Сократ в итоге формулирует тезис: «Следовательно, один только рассудительный человек может познать самого себя и выявить, что именно он знает и что – нет... Никто другой всего этого не может»<sup>10</sup>.

Древних греков побуждала «познать самих себя» и знаме-

---

<sup>9</sup> Из интервью Владимира Набокова Джеймсу Моссмону, опубликованному в сборнике «Строгие суждения».

<sup>10</sup> *Hamilton E. и др. The Collected Dialogues of Plato. Princeton, NJ: Princeton University Press, 1961*

нитая надпись, высеченная на стене храма в Дельфах. Для них самосознание было постоянной деятельностью; к нему нужно было стремиться. Эти взгляды перекочевали и в средневековую религиозную традицию: например, итальянский священник и философ Фома Аквинский высказывал предположение, что в отличие от Бога, который знает себя по умолчанию, нам необходимо прикладывать усилия, чтобы познать собственный разум. Фома Аквинский и его монахи проводили в молчаливом созерцании долгие часы. Они верили, что, лишь участвуя в организованной саморефлексии, могут уподобиться образу Божьему<sup>11</sup>.

Схожее представление о стремлении к самосознанию можно наблюдать в восточных традициях, например в буддизме. Духовная цель просветления заключается в растворении эго, что позволяет здесь и сейчас обрести более прозрачное и достоверное знание о нашем сознании. Основатель китайского даосизма Лао-Цзы выразил идею о том, что обретение самосознания является одним из высших стремлений, когда писал: «Знать, а казаться незнающим – вот совершенство. Не знать, а думать, что знаешь, – это болезнь»<sup>12</sup>.

Сегодня существует множество сайтов, блогов и книг по самосовершенствованию, которые призывают нас «найти се-

---

<sup>11</sup> Self-Knowledge: A History / Под. ред. U. Renz. Oxford: Oxford University Press, 2017.

<sup>12</sup> *Baggini J.* How the World Thinks: A Global History of Philosophy. 2018. Reprint, London: Granta, 2019.

бя» и развивать свое самосознание. Все это из лучших побуждений. Однако, хотя нас постоянно побуждают к развитию самосознания, мало кто уделяет внимание тому, как самосознание действительно работает. Я нахожу это необъяснимым. Было бы странно призывать людей чинить машины, если они не знают, как работает двигатель, или ходить в спортзал, если они не знают, какие мышцы следует нагружать. Данная книга призвана заполнить этот пробел. Она не будет сборником ценных советов или источником цитат. Вместо этого я хочу, опираясь на новейшие исследования в области психологии, компьютерных и нейронаук, предложить читателю руководство по строительным блокам, из которых собирается самосознание. Разобравшись, как работает самосознание, мы сможем использовать его с большей пользой и таким образом попытаемся откликнуться на афинский призыв.

Кроме того, я стремлюсь помочь нам лучше использовать машины, как существующие в настоящий момент, так и те, что, вероятно, появятся в ближайшем будущем. Уже сегодня мы вынуждены иметь дело со сложными системами, принимающими непонятные нам решения, – вспомните воображаемый визит к компьютерному доктору и его не поддающийся объяснениям совет о проведении операции. Нас окружают разумные, но бессознательные алгоритмы – от моделей прогнозирования климата до автоматизированных рыночных трейдеров, – и подобные инструменты готовы вторг-

нуться во все сферы нашей жизни. Во многих случаях эти алгоритмы делают нашу жизнь проще и продуктивнее, они даже могут помочь справиться с невиданными ранее проблемами вроде глобального потепления. Но существует опасность, что подчинение сверхумным «черным ящикам» ограничит человеческую автономию: исключив из уравнения метапознание, мы перестанем понимать, почему и как были приняты те или иные решения, и вместо этого будем вынуждены слепо следовать советам машинных алгоритмов. Философ Дэниел Деннет на этот счет пишет: «Реальная опасность, как мне кажется, заключается не в том, что превосходящие нас в интеллекте машины узурпируют нашу роль вершителей судеб, а в том, что мы переоценим возможности наших новейших инструментов мышления, преждевременно уступив им полномочия, выходящие далеко за пределы их компетенции»<sup>13</sup>. Мы узнаем, что наука о самосознании открывает нам альтернативный вариант будущего, гарантирующий, что осознание компетенции останется на вершине списка приоритетов как для нас самих, так и для наших машин.

## Взгляд в будущее

Давайте посмотрим в будущее. Основной посыл этой книги в том, что человеческий мозг является вместилищем осо-

---

<sup>13</sup> *Dennett D. C. From Bacteria to Bach and Back: The Evolution of Minds. London: Penguin, 2018.*

бых алгоритмов самосознания. Разбором их работы мы и займемся в первой части книги. Мы увидим, что нейронные схемы, поддерживающие работу метапознания, не свалились на нас с неба. Они укоренены в функциях человеческого мозга, сформированного эволюцией. Это означает, что многие строительные кирпичики метапознания закладываются еще на ранних этапах развития человека, а также обнаруживаются и у других видов животных. Мы рассмотрим как бессознательные процессы, формирующие составные элементы самомониторинга, так и сознательные – позволяющие вам осознавать свои текущие переживания. Как станет ясно, говоря о самосознании, мы имеем в виду целую совокупность способностей, благодаря которым человек познаёт себя – например, способность признавать свои ошибки или описывать свой опыт<sup>14</sup>.

К концу первой части мы разберемся, из каких важнейших компонентов складывается полноценная способность к

---

<sup>14</sup> Терминология, используемая учеными и философами, изучающими самосознание и метакогнитивные процессы, может вызвать путаницу. В этой книге я использую понятия метапознания и самоконтроля, чтобы ссылаться на любые процессы, который контролирует другой когнитивный процесс, например осознание, что мы сделали ошибку в решении математической проблемы. Самоконтроль и метапознание могут иногда происходить неосознанно. Тогда как самосознание – это способность осознанно рефлексировать о нас, нашем поведении и нашей ментальной жизни. Некоторые психологи ограничивают использование понятия самосознание, чтобы оно означало только телесный самоанализ или анализ местоположения и внешнего вида тела. Однако в этой книге, я заинтересован в понимании ментального состояния.

самосознанию. Мы также поймем, как и почему этот процесс может разладиться, как в результате заболеваний вроде шизофрении и деменции самосознание может сбоить. Во второй части мы переключим внимание на то, как мы используем самосознание для обучения, принятия решений и сотрудничества с другими людьми во многих сферах нашей жизни. Разобравшись, как и почему самосознание может искажаться, и осознав как его силу, так и хрупкость, мы научимся избегать ситуаций, в которых оно может дать осечку. Чтобы понять, почему осознание себя и того, как себя осознают другие, имеет решающее значение для построения лучшего и более справедливого общества, мы углубимся в несколько важных сфер человеческой деятельности, в том числе обсудим роль метапознания в допросах свидетелей, а также в политике и науке. Мы изучим, как самосознание помогает нам отделять реальность от выдумки и как с его помощью можно даже научиться управлять нашими сновидениями. Мы увидим, что иногда наше самосознание бездействует, а значит, во множестве ситуаций люди ничем не отличаются от черных ящиков, будучи неспособными объяснить, что и почему они сделали.

Кроме всего прочего, нам станет понятно, что, несмотря на все ограничения, способность людей к самосознанию и самообъяснению лежит в основе наших понятий об автономии и ответственности. Мы изучим роль самосознания в обучении и преподавании. Разберемся, почему самосозна-

ние скорее мешает добиваться результатов в спорте, но при этом помогает тренировать других спортсменов. Мы увидим, как цифровые технологии меняют наше осознание себя и других в целом ряде важнейших аспектов. Я покажу, что в условиях возрастающей политической поляризации и дезинформации особую важность получает развитие способностей саморефлектировать и подвергать сомнению наши убеждения и мнения. Мы подумаем, почему даже самые мощные современные компьютеры не обладают метапознанием, а также увидим, как стремительное распространение машинного обучения проводит границу между алгоритмами разума и алгоритмами самосознания. Мы обсудим, что это может означать для общества и как мы можем это контролировать – стоит ли нам пытаться внедрить самосознание в наши компьютеры или удостовериться, что мы сможем понимать и использовать создаваемые нами машины. Чем бы ни завершились эти усилия, они могут стать ключом к решению некоторых острейших социальных проблем.

Надеюсь, к концу этой книги нам станет понятно, почему культивирование самосознания всегда и везде было необходимым условием процветания и успеха – от древних Афин до зала заседаний директоров Amazon. Но мы забегаем вперед. Чтобы раскрыть тайны самосознания, нам предстоит начать с самых простых строительных кирпичиков: с двух особенностей работы нашего сознания – отслеживания неопределенности и мониторинга своих действий. Хотя они и мо-

гут показаться простыми, эти особенности составляют фундамент самосознающего мозга.

# Часть I

## Как разум осознает себя

### Глава 1

#### Как быть неуверенным

*Другой источник идей есть внутреннее восприятие действий нашего ума, <...> и хотя этот источник не есть чувство, поскольку не имеет никакого дела с внешними предметами, тем не менее он очень сходен с ним и может быть довольно точно назван внутренним чувством.*

– Джон Локк. «Опыт о человеческом разумении». Кн. 2

Есть там что-нибудь – или нет? Такой вопрос встал перед Станиславом Петровым одним ранним сентябрьским утром 1983 года. Петров был подполковником ПВО советских войск и отвечал за мониторинг системы спутников раннего предупреждения. Холодная война между США и Советским Союзом находилась на пике, и угроза запуска ядерных ракет большой дальности одной из сторон была вполне реальна. В то роковое утро в командном центре Петрова сработала сигнализация, оповестившая, что пять американ-

ских ракет направляются в сторону Советского Союза. Согласно доктрине взаимного гарантированного уничтожения, задачей Петрова было немедленно сообщить об атаке своему руководству, чтобы то могло нанести контрудар. Время поджимало – через двадцать пять минут ракеты разорвались бы на советской земле<sup>15</sup>.

Но Петров посчитал, что сигнал тревоги вряд ли вызвала настоящая ракета, и доложил о сбое системы. Он больше верил в ненадежность спутников и помехи на экране радара, чем во внезапный ракетный удар Соединенных Штатов, который несомненно бы привел к началу ядерной войны. Несколько минут нервного ожидания подтвердили его правоту. Ложная тревога стала результатом ошибки спутников, принявших солнечные блики на поверхности облаков за ракеты, несущиеся в верхних слоях атмосферы.

Петров воспринимал мир в оттенках серого и был готов усомниться в том, что говорили ему системы спутников и органы чувств. Его готовность осознать двусмысленность положения и усомниться спасла мир от катастрофы. В этой главе мы увидим, что представление о неопределенности служит ключевым ингредиентом в рецепте наших систем самосознания. Человеческий мозг представляет собой изощренный механизм для отслеживания неопределенности, которая играет в работе мозга огромную роль, а не только участвует

---

<sup>15</sup> *Steele J.* Stanislav Petrov Obituary // The Guardian. 2017. 11 октября. URL: [theguardian.com/world/2017/oct/11/stanislav-petrov-obituary](http://theguardian.com/world/2017/oct/11/stanislav-petrov-obituary).

в принятии ответственных решений, как в истории Петрова. Без способности оценивать неопределенность мы вряд ли вообще смогли бы воспринимать мир. Замечательный, хотя и побочный, эффект этой способности – неопределенностью можно пользоваться, чтобы сомневаться в самих себе.

## **Обратные задачи и способы их решения**

Сложность задачи Петрова заключалась в том, что ему нужно было отличить сигнал радара от помех. Один и тот же след на экране мог быть как настоящей ракетой, так и сбоем системы. Понять лишь по изображению, с чем имеешь дело, невозможно. Такая задача называется обратной – потому что для ее решения необходимо развернуть причинно-следственную цепочку задом наперед и сделать предположение о причинах получаемой нами информации. Схожим образом наш мозг постоянно решает обратные задачи, не зная, что в действительности происходит в мире.

Причина в том, что мозг заперт внутри темной черепушки и контактирует с внешним миром лишь посредством ограниченной низкокачественной информации, предоставляемой органами чувств. Возьмем, к примеру, простую, на первый взгляд, задачу – определить, была ли только что вспышка света в затемненной комнате. Если сделать вспышку достаточно тусклой, то иногда может показаться, что свет горит, даже если в действительности это не так. Поскольку глаза и

мозг подвержены помехам, при каждом повторении стимула нейроны в зрительной коре головного мозга возбуждаются не совсем одинаково. Иногда, даже когда свет не горит, случайные помехи приводят к высокой частоте возбуждения нейронов, подобно тому как след на экране радара Петрова стал результатом атмосферных помех. Поскольку мозг не знает, вызвана ли эта высокая частота стимулом или помехами, то, когда нейроны зрительной коры активно возбуждаются, будет казаться, что свет зажжен<sup>16</sup>.

Так как каждое из наших чувств – осязание, обоняние, вкус, зрение и слух – имеет доступ лишь к небольшому, искаженному кусочку реальности, им необходимо объединить ресурсы, чтобы выдвинуть наилучшее предположение о том, что же на самом деле происходит. Наши чувства похожи на слепцов из древней индийской притчи. Тот, кто держит слона за ногу, говорит, что, должно быть, слон – это столб; тот, кто ощупывает хвост, говорит, что слон похож на веревку; тот, кто трогает хобот, говорит, что он похож на ветку дерева; тот, кто прикасается к уху, считает, что он напоминает веер; тот, кто ощупывает живот слона, – что он вроде стены; а тот, кто изучает бивень, – что он похож на массивную трубу. В конце концов проходящий мимо незнакомец сообщает им, что на самом деле они все правы – у слона есть все описанные черты. Лучше бы им объединить свои наблюдения, говорит он, а не спорить.

---

<sup>16</sup> *Green D. M. Signal Detection Theory and Psychophysics. New York: Wiley, 1966.*

Математическая формула, известная как теорема Байеса, – мощный инструмент для осмысления такого рода проблем.

Чтобы увидеть, как она помогает нам решать обратные задачи, можем сыграть в следующую игру. У меня есть три игральные кости, две из которых – обычные кубики с числами от одного до шести, а третья – либо с тройками, либо с нулями на всех гранях. Втайне от вас я одновременно брошу все три кости и назову общий итог. Каждый раз я могу использовать один кубик с подвохом – тот, на котором всегда выпадает ноль, или тот, на котором всегда выпадает тройка. Например, при первом броске я могу выбросить два, четыре и ноль (на третьем кубике), что в сумме составит шесть. Ваша задача: основываясь только на общем результате, сделать наилучшее предположение о том, какой кубик с подвохом я использовал – со всеми тройками или со всеми нулями<sup>17</sup>.

В этой игре ноль или тройка на кубике с подвохом соответствуют «скрытым» состояниям мира: ракетам в дилемме Петрова и включенному свету в случае с нейронами зрительной коры. Тем или иным образом нам нужно использо-

---

<sup>17</sup> Основы правила Байеса были впервые выявлены арабским математиком XI века Ибн аль-Хайтамом, развиты английским священником и математиком Томасом Байесом в 1763 году и применены к целому ряду научных проблем французским математиком XVIII века Пьером-Симоном Лапласом. *McGrayne S. B. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy.* New Haven, CT: Yale University Press, 2012.

вать полученную искаженную информацию – сумму значений всех трех игральных костей, – чтобы определить скрытое условие.

Иногда это легко. Если я скажу, что общая сумма четыре или меньше, то вы поймете, что, раз значение такое маленькое, на третьем кубике должен был выпасть ноль. Если общий результат превышает 12 (две шестерки плюс число больше нуля), то вы точно знаете, что на третьем кубике выпала тройка. Но что насчет значений между этими крайностями? Если общий результат равен шести или восьми? Это сложнее.

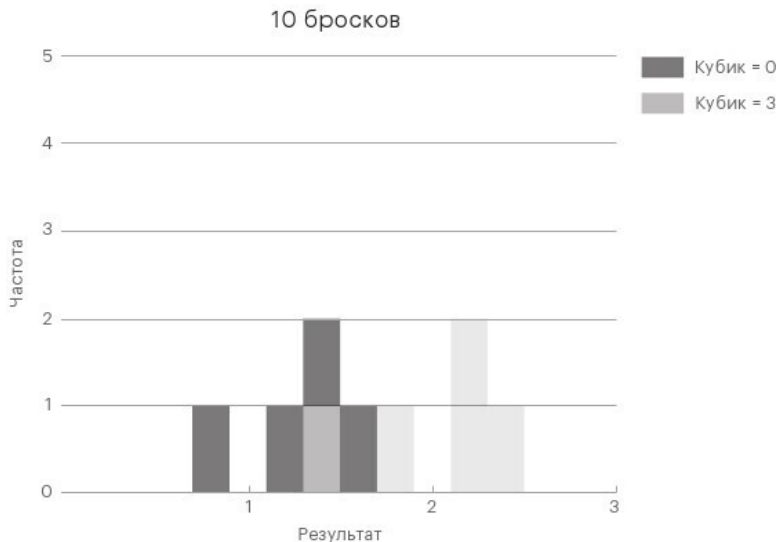
Один из способов решения этой задачи – метод проб и ошибок. Мы можем много раз бросить кости, записать общий результат и понаблюдать за истинным положением вещей: что на самом деле выпадает на третьем кубике при каждом броске. Первые несколько бросков в игре могут выглядеть следующим образом:

Бросок	Кубик 1	Кубик 2	Кубик с подвохом	Результат
1	2	4	0	6
2	5	1	3	9
3	5	6	3	14

И так далее, еще много бросков. Более простой способ:

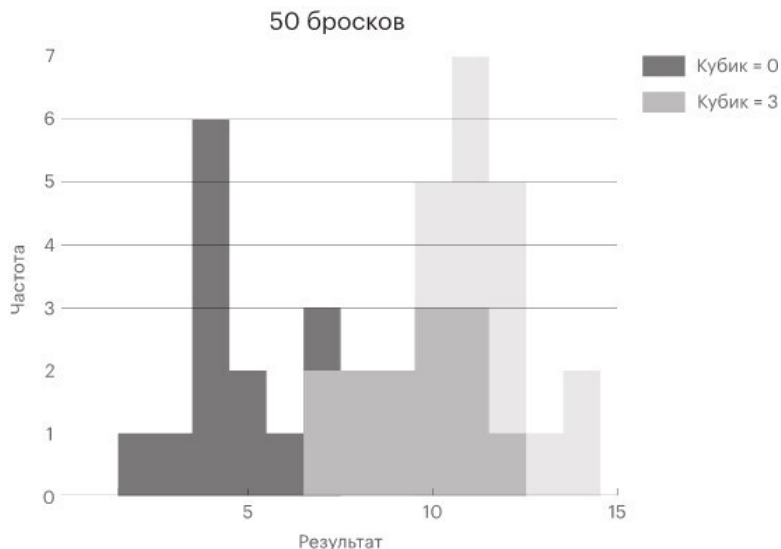
представить эти данные в виде диаграммы, отображающей, сколько раз наблюдается определенное общее значение (скажем, шесть) и какой в этот момент использовался кубик (ноль или три). Кубики с подвохом можно обозначить определенными цветами: здесь я выбрал серый для кубика с нулями и белый для кубика с тройками.

После десяти бросков график может выглядеть следующим образом.



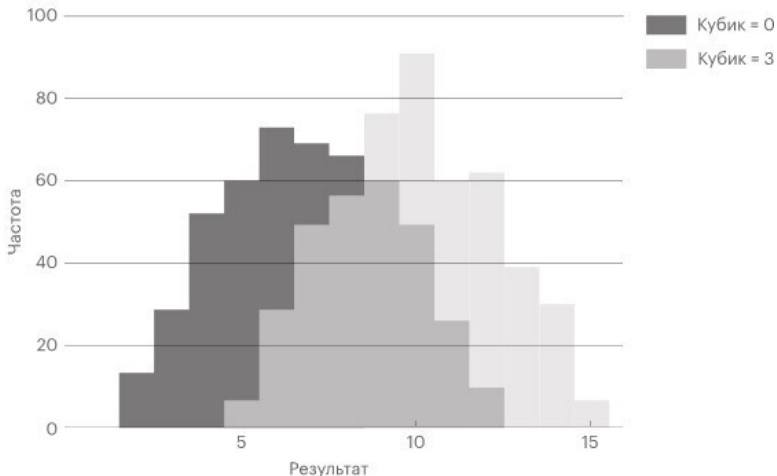
Не слишком информативно, и виден лишь разброс различных результатов, как и в нашей таблице. Но после пяти-

десяти бросков начинает проявляться закономерность.



А после тысячи картина становится весьма ясной.

## 1000 бросков



Результаты, полученные в ходе нашего эксперимента, формируют две отчетливые вершины. Большинство бросков попадает в средний диапазон, а пики приходятся на семерку и десятку. Это логично. В среднем два обычных кубика дают в сумме около семи, и поэтому, добавив к этому числу ноль или три от кубика с подвохом, мы, как правило, получим семь или десять. То, что ранее говорила нам интуиция, подтвердилось: вы видите, что значение четыре или меньше получается только при броске кубика с нулем на всех гранях, а значение 13 и больше – только при использовании кубика с тройкой.

Теперь, вооружившись этими данными, давайте вернемся

к нашей игре. Если я назову определенное общее значение, например десять, и попрошу вас угадать число на кубике с подвохом, что вы должны ответить? Согласно приведенному выше графику, вероятнее всего, в этом случае на кубике выпала тройка. Благодаря правилу Байеса мы знаем, что относительная высота белых и серых столбиков (при условии, что мы провели эксперимент достаточное число раз) отражает, насколько выше вероятность выпадения тройки по сравнению с нулем – в данном случае примерно в два раза. Согласно Байесу, оптимальное решение для этой игры – всегда называть наиболее вероятное значение кубика, то есть три, если общий результат девять или больше, и ноль – если восемь или меньше.

Только что мы набросали алгоритм для принятия решений на основе искаженной информации. Кубик с подвохом всегда маячит на заднем плане и вносит свой вклад в общий результат. Но его истинное значение заглушено помехами, создаваемыми двумя обычными кубиками. Точно так же Петров не мог определить наличие ракеты лишь по искаженному помехами сигналу радара. Наша игра является примером общего типа задач, связанных с принятием решений в условиях неопределенности, – их можно решить, применив правило Байеса.

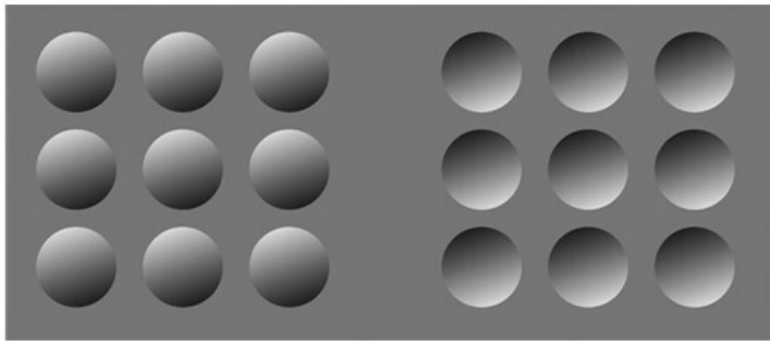
В случае с судьбоносным решением Петрова набор потенциальных объяснений ситуации ограничен: либо это настоящая ракета, либо ложная тревога. Аналогичным образом, в

нашей игре в кости есть только два объяснения: это кубик либо с тройками на гранях, либо с нулями. Но в большинстве ситуаций не только наша сенсорная система подвержена помехам, а еще существует целый диапазон потенциальных объяснений для поступающей информации. Представьте себе нарисованный круг около 20 сантиметров диаметром, который находится на расстоянии одного метра от вашего глаза. Отраженный от круга свет движется по прямой линии, проходит через хрусталик глаза и создает маленькое изображение (круг) на сетчатке. Поскольку изображение на сетчатке двухмерное, мозг может интерпретировать его как вызванное любым бесконечным числом кругов разного размера, расположенных на соответствующих расстояниях. Примерно такое же изображение на сетчатке вызвал бы круг диаметром 40 сантиметров, расположенный на расстоянии двух метров, или круг диаметром восемь метров на расстоянии 40 метров. Во многих случаях нам просто не хватает информации, чтобы определить, что мы видим.

В таких более сложных обратных задачах догадаться, в чем состоит лучшее объяснение, можно на основе дополнительной информации из других источников. К примеру, чтобы оценить фактический диаметр круга, мы можем использовать такие подсказки, как различия в изображении, получаемом разными глазами, разницу в текстуре, положении и оттенке близлежащих объектов и так далее.

Чтобы в реальном времени понять, как это происходит,

Взгляните на эти две картинки.

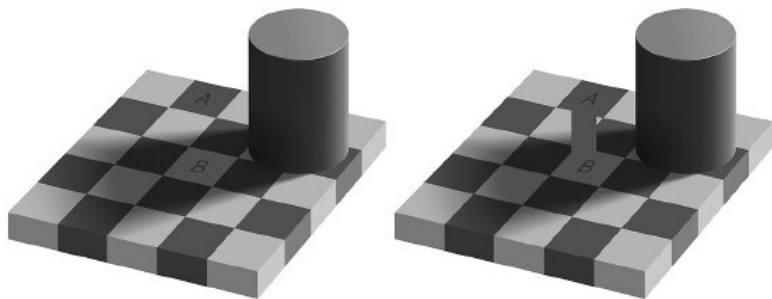


На изображении слева большинство людей видят группу выпуклых бугорков, приподнятых над поверхностью страницы. Изображение справа, напротив, выглядит как совокупность маленьких ямок или впадин. В чем же разница?

Иллюзия возникает из-за того, что ваш мозг решает обратную задачу. На самом деле левая и правая картинки – одно и то же изображение, развернутое на 180 градусов (можете перевернуть книгу и проверить!). Разными они кажутся, поскольку наша зрительная система привыкла, что свет падает сверху – источник света, заливающего окружающее пространство, обычно располагается где-то у нас над головой. В свою очередь, освещение снизу вверх – например, свет костра на склоне скалы или лучи прожекторов, направленные на верхушку собора, – статистически встречается реже. Когда

мы смотрим на эти две картинки, наш мозг интерпретирует светлые части левого изображения как свет, падающий на выпуклые бугорки, а темные части правого изображения – как тени, создаваемые ямками, несмотря на то что обе картинki составлены из одного и того же исходного материала.

Другая поразительная иллюзия – изображение, созданное ученым Эдвардом Адельсоном.



### Шахматная доска Адельсона

Клетки, обозначенные на левом рисунке буквами А и В, на самом деле окрашены в идентичный оттенок серого; они имеют одинаковую яркость. Клетка В кажется светлее, поскольку ваш мозг «знает», что она расположена в тени: чтобы воспроизвести для глаза тот же уровень освещенности, что и у клетки А, которая целиком находится на свету, она должна быть светлее изначально. Эквивалентность клеток А и В легко можно оценить, соединив их – как на правом рисунке.

Подсказка, которую дает этот дополнительный мостик, перекрывает фактор тени в интерпретации мозгом изображения (чтобы убедиться, что левое и правое изображения одинаковы, попробуйте закрыть их нижнюю половину листом бумаги).

Получается, что на самом деле эти удивительные иллюзии – вовсе не иллюзии. Одну интерпретацию изображения дает научная аппаратура – показатели, выдаваемые измерителями освещенности и компьютерными мониторами. Другую – наши зрительные системы, настроенные на обнаружение закономерностей, таких как тени или свет, падающий сверху вниз. Эти закономерности помогают нашим системам выстраивать действенные модели мира. В реальном мире, где есть свет, тени и полутени, эти модели обычно оказываются верными. Многие оптические иллюзии хитроумным способом воздействуют на работу системы, искусно настроенной на получение перцептивных выводов<sup>18</sup>. Кроме того, как мы узнаем из следующего раздела, некоторые принципы устройства мозга согласуются с тем, как эта система в массовых масштабах решает обратные задачи.

## **Построение моделей мира**

Зрительная система – одна из наиболее изученных частей мозга человека и обезьян. Различные области задней части

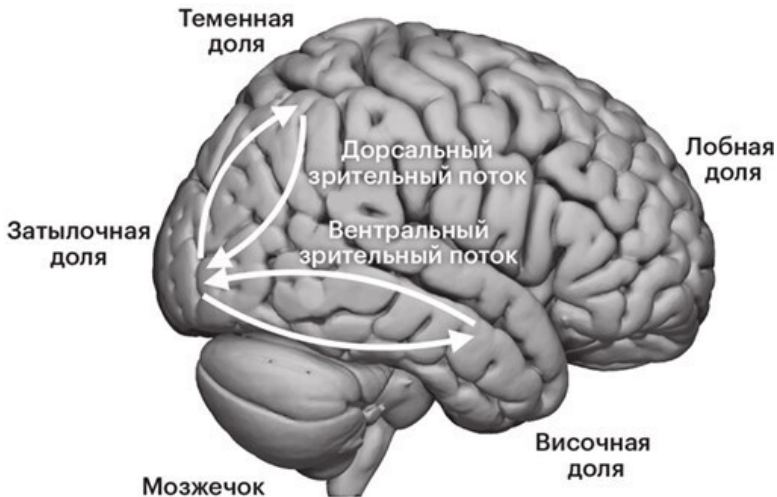
---

<sup>18</sup> Выводы, получаемые в результате чувственного восприятия. – *Прим. пер.*

мозга обрабатывают разные аспекты визуального сигнала. Чем выше их цифровое обозначение, тем более продвинутой стадии обработки изображения они соответствуют. Области V1 и V2 извлекают информацию о направлении линий и формах, V4 – о цвете, а V5 – о движении объектов. На выходе из областей V мы попадаем в русло вентрального зрительного потока, задача которого – собрать все части информации воедино и идентифицировать цельные объекты, такие как лица, тела, столы и стулья. Параллельно области мозга, входящие в дорсальный зрительный поток, отслеживают, где располагаются и как перемещаются объекты<sup>19</sup>.

---

<sup>19</sup> *Felleman D. J. u др* Distributed Hierarchical Processing in the Primate Cerebral Cortex // *Cerebral Cortex*. 1991. Т. 1. № 1. С. 1–47.; *Zeki, S. u др*. The Autonomy of the Visual Systems and the Modularity of Conscious Vision // *Philosophical Transactions of the Royal Society B: Biological Sciences* 353. 1998. № 1377. С. 1911–1914.



Правое полушарие мозга человека. Отмечено расположение четырех долей головного мозга, мозжечка и ключевых зрительных путей

В начале вентрального зрительного потока отдельные клетки мозга кодируют лишь небольшую часть внешнего мира – например, участок в нижнем левом углу нашего поля зрения. Но по мере продвижения вверх по иерархической структуре клетки, подобно отдаляющейся камере, начинают расширять свой фокус. В конце концов достигается высшая точка иерархии, в которой уже не столь важно, где отображается стимул. Гораздо большую роль играет, что он воссо-

здает: лицо, дом, кошку, собаку и так далее. Объектив максимально отдален, и информация об идентичности объекта представлена независимо от его расположения.

Однако крайне важно, что информация в зрительной системе не течет в одном направлении. Долгое время преобладало мнение, согласно которому обработка информации в мозге является системой прямого регулирования: она получает информацию из внешнего мира, обрабатывает ее тайными, хитроумными способами, а затем выдает команды, заставляющие нас ходить и говорить. Сейчас такая модель («вход – выход») вытеснена множеством данных, которые в нее не вписываются. В зрительной системе, например, существует столько же, если не больше, связей, направленных в противоположную сторону. Они так и называются – обратная связь, или «сверху вниз». Информация распространяется как вперед, так и назад; постоянные циклы нейронной активности поставляют данные как с нижних уровней иерархии на верхние, так и в обратном направлении. Такой способ рассмотрения механизмов сознания известен как предиктивная обработка. Это радикально иное понимание работы мозга, хотя оно имеет уже продолжительную интеллектуальную историю, о которой свидетельствует обширная библиография в примечаниях<sup>20</sup>.

---

<sup>20</sup> Oxford University Press, 2016; *Clark A.* Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science // *Behavioral and Brain Sciences* 36. 2013. № 3. С. 181–204.; *Craik K.* The Nature of Explanation. Cambridge: Cambridge University Press, 1963; *Friston K.* The Free-Energy Principle: A Unified

Архитектура предиктивной обработки особенно хорошо подходит для решения обратных задач. Вместо пассивного получения информации мозг может использовать связи «сверху вниз», для того чтобы активно строить наше восприятие внешнего мира и придавать форму тому, что мы видим, слышим, думаем и чувствуем. Более высокие уровни иерархии предоставляют информацию о том, с чем мы можем столкнуться в той или иной ситуации, а также о диапазоне гипотез, которые мы способны принять. Например, вы знаете, что у вашего друга есть лабрадор, и поэтому ожидаете увидеть собаку, когда входите в его дом, но не знаете, где именно в вашем зрительном поле она появится. Эта предварительная высокоуровневая информация – пространственно-инвариантное понятие «собака» – обеспечивает соответствующим контекстом более низкие уровни зрительной системы, помогая им легко интерпретировать размытое пятно в форме собаки, устремляющееся к вам, когда вы открываете дверь.

Степень, до которой наши системы восприятия должны полагаться на такие закономерности, в свою очередь, зависит от того, насколько мы сомневаемся в информации, поступающей от наших органов чувств. Вспомните дилемму Петрова. Если бы он был уверен в безупречности и безошибоч-

---

Brain Theory? // Nature Reviews Neuroscience 11. 2010. № 2. С. 127–138; *Helmholtz H. L. F. Treatise on Physiological Optics*. London: Thoemmes Continuum, 1856; *Gregory R. The Intelligent Eye*. New York: McGraw-Hill, 1970; *Hohwy J. The Predictive Mind*. Oxford: Oxford University Press, 2013.

ности технологии обнаружения ракет, то в меньшей степени был бы готов усомниться в том, что говорила ему система. Стоит ли нам корректировать свои убеждения при получении новых данных, зависит от того, насколько надежной мы считаем эту информацию.

В действительности байесовские версии прогностической обработки говорят нам о том, что стоит комбинировать различные источники информации (наши предварительные убеждения и данные, поступающие через органы чувств) обратно пропорционально нашей неуверенности в них. Можно представить этот процесс как помещение теста для пирога в гибкую форму для выпечки. Форма – это наши предварительные предположения о мире. Тесто же представляет собой сенсорную информацию – световые и звуковые волны, улавливаемые глазами и ушами. Если поступающие данные точны или информативны, то тесто будет густым или почти твердым и на него почти не повлияет форма для выпечки. Если же, напротив, данные менее точны, то тесто будет более жидким и конечный продукт примет соответствующие очертания.

К примеру, глаза предоставляют более точную информацию о местонахождении объектов, нежели слух. Это означает, что зрение может изолировать предполагаемый источник звука, исказив наше восприятие его местоположения. Этим умело пользуются чревовещатели, способные «передать» свой голос марионетке, которую они держат на расстоянии

вытянутой руки. Истинное мастерство чревовещания заключается в умении говорить, не шевеля губами. Если добиться этого, мозг зрителей сделает все остальное, соотнеся звук с его следующим наиболее вероятным источником – говорящей куклой<sup>21</sup>.

Таким образом, вполне логично, что отслеживание неопределенности – неотъемлемая часть того, как мозг обрабатывает сенсорную информацию. Наблюдения за клетками зрительной коры головного мозга показывают, как это может происходить. Хорошо известно, что движущиеся объекты, такие как машущая рука или прыгающий мяч, активируют нейроны в области мозга обезьян, известной как МТ (аналог человеческой V5). Но клетки в МТ активируются не при любом направлении движения. Некоторые из клеток больше реагируют на объекты, движущиеся влево, другие – вверх, вниз и во всех других направлениях. Когда частота возбуждения клеток МТ фиксируется в результате многократных воспроизведений различных направлений движения, формируется распределение, подобное тому, что мы наблюдали в игре в кости. В каждый отдельно взятый момент времени

---

<sup>21</sup> *Kersten D. и др.* Object Perception as Bayesian Inference // *Annual Review of Psychology*. 2004. № 55. С. 271–304; *Ernst M. O.* Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion // *Nature* 415. 2009. № 6870. С. 429–433; *Pick H. L. и др.* Sensory Conflict in Judgments of Spatial Direction // *Perception & Psychophysics*. 1969. Т. 6. № 4. С. 203–205; *Bertelson C.* Ventriloquism: A Case of Crossmodal Perceptual Grouping // *Advances in Psychology*. 1999. № 129. С. 347–362; *McGurk H.* Hearing Lips and Seeing Voices // *Nature* 264. 1978. № 5588. С. 746–748.

эти популяции клеток МТ можно считать сигнализирующими о неопределенности в отношении конкретного направления движения, аналогично тому, как искаженное общее значение игральные кости сигнализирует о вероятности выпадения ноля или тройки<sup>22</sup>.

Неопределенность крайне важна и для оценки состояния нашего собственного тела. Информация о том, где в пространстве располагаются конечности, как быстро бьется сердце или какова интенсивность болевого стимула, поставляется в череп сенсорными нейронами. С точки зрения мозга разница между электрическими импульсами, проходящими по зрительному нерву, и нейронными сигналами, поступающими из кишечника, сердца, мышц или суставов, весьма незначительна. Все это – сигналы, сообщающие о том, что может происходить за пределами черепа, и искаженные иллюзиями вроде описанных выше оптических. В одном известном эксперименте поглаживание резиновой руки синхронно с настоящей (скрытой) рукой испытуемого убеждало его, что резиновая рука – его собственная.

В свою очередь, иллюзия обладания новой резиновой рукой приводила к ослаблению нейронных сигналов, посылаемых мозгом настоящей руке. Подобно тому как кукла перехватывает голос чрево вещателя, синхронность наблюдений

---

<sup>22</sup> *Born R. T. u др.* Structure and Function of Visual Area MT // Annual Review of Neuroscience. 2005. № 28. С. 157–189; *Ma W. J. u др.* Bayesian Inference with Probabilistic Population Codes // Nature Neuroscience 9. 2006. № 11. С. 1432–1438.

за резиновой рукой и ощущений при поглаживании уменьшает чувство обладания настоящей рукой<sup>23</sup>.

## **Используем неопределенность, чтобы сомневаться**

Конечно, никто не говорит, что каждый раз, познавая мир вокруг, мы специально прибегаем к уравнениям Байеса. Напротив, механизмы, используемые мозгом для решения обратных задач, срабатывают сами по себе – немецкий физик Герман фон Гельмгольц назвал это процессом «бессознательных умозаключений». Мозг быстро, буквально мгновенно оценивает влияние света и тени на впадины, выпуклости и шахматные доски, изображения которых мы видели на предыдущих страницах.

Аналогичным образом мы воссоздаем лицо близкого друга, вкус хорошего вина и запах свежее испеченного хлеба, комбинируя предварительные предположения и информацию от органов чувств; тщательно взвешивая их с учетом соответ-

---

<sup>23</sup> *Apps M. A. J. u др.* The Free-Energy Self: A Predictive Coding Account of Self-Recognition // *Neuroscience & Biobehavioral Reviews*. 2014. № 41. С. 85–97; *Blanke O. u др.* Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness // *Neuron* 88. 2015. № 1. С. 145–166; *Botvinick M. M. u др.* Rubber Hands ‘Feel’ Touch That Eyes See // *Nature* 391. 1998. № 6669. С. 756; *Della G. u др.* Decreased Motor Cortex Excitability Mirrors Own Hand Disembodiment During the Rubber Hand Illusion // *eLife*. 2016. № 5. e14972; *Seth A. K.* Interoceptive Inference, Emotion, and the Embodied Self // *Trends in Cognitive Sciences* 17. 2013. № 11. С. 565–573.

ствующим неопределенностям. Нейробиолог Анил Сет называет наше восприятие мира «контролируемой галлюцинацией» – наилучшим предположением о том, что на самом деле есть.

Очевидно, что оценка неопределенности, характеризующей те или иные источники информации, – основа нашего восприятия мира. Но изобретательные решения обратных задач дают замечательный побочный эффект. Оценивая неопределенность для того, чтобы воспринимать мир, мы обретаем способность сомневаться в том, что воспринимаем. Чтобы увидеть, как неопределенность с легкостью превращается в сомнение, давайте снова обратимся к игре в кости. Чем ближе общее значение к 15 или к нулю, тем больше мы уверены, что на кубике с подвохом выпала соответственно тройка или ноль. Но в средней части графика, где серые и белые столбики равны по высоте (общие значения равняются семи и восьми), доказательств недостаточно для любого из вариантов. Если я спрошу вас, насколько вы уверены в своем ответе, будет разумно, если вы усомнитесь, когда речь пойдет о значениях семь и восемь, но будете более уверены в случае меньших или больших результатов. Другими словами, мы знаем, что, скорее всего, знаем ответ, если неопределенность низкая, и знаем, что, скорее всего, *не знаем* ответ, когда неопределенность высокая.

Правило Байеса дает математическую основу для размышлений об этих оценках неопределенности, которые на-

зывают еще решениями второго типа, поскольку они касаются точности других решений – в отличие от решений первого типа, которые касаются окружающего мира. Согласно теореме Байеса, нам следует больше сомневаться, когда дело касается ответов, приходящихся на центр графика, поскольку именно они чаще всего приводят к ошибкам и с наименьшей вероятностью оказываются правильными. И напротив, по мере приближения к краям распределения вероятность правильного ответа возрастает. Используя неопределенность, присущую решению обратных задач, мы в качестве бонуса достигаем рудиментарной формы метапознания – и никаких дополнительных механизмов для этого не требуется<sup>24</sup>.

Поскольку отслеживание неопределенности играет основополагающую роль в том, как мозг воспринимает мир, неудивительно, что эта форма метапознания доступна множеству животных. Один из первых – и наиболее изобретательных – экспериментов по изучению метапознания у животных был проведен психологом Дэвидом Смитом, работавшим с бутылконосым дельфином по имени Натиа. Смит

---

<sup>24</sup> *Kiani R. u др.* Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex // *Science*. 2009. № 5928. С. 759–764; *Carruthers P.* How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition // *Behavioral and Brain Sciences* 32. 2009. № 2. С. 121–138; *Insabato A. u др.* Neural Correlates of Metacognition: A Critical Perspective on Current Tasks // *Neuroscience & Biobehavioral Reviews*. 2016. № 71. С. 167–175; *Meyniel F. u др.* The Sense of Confidence During Probabilistic Learning: A Normative Account // *PLOS Computational Biology* 11. 2015. № 6. e1004305.

обучил Натуга нажимать на разные рычаги в аквариуме в зависимости от частоты слышимого им звука. Низкочастотный звук варьировался от очень низкого до относительно высокого, почти высокочастотного. Таким образом, как и в нашей игре в кости, создавалась зона неопределенности, когда сложно было понять, какой же ответ правильный<sup>25</sup>.

Когда Натуга научился справляться с этим заданием, в аквариум добавили третий рычаг, нажав на который можно было пропустить текущий звук и сразу перейти к следующему – дельфиний аналог пропуска вопроса в тесте. Смит рассудил, что если Натуга, будучи не уверен в ответе, откажется принимать решение, вместо того чтобы угадывать, то сможет добиться более высокой общей точности. Именно это Смит и обнаружил. Результаты показали, что чаще всего Натуга нажимал на третий рычаг, когда звук был пограничным. Как пишет Смит, «в случае неуверенности дельфин явно сомневался и колебался между двумя возможными ответами, но когда был уверен, то так устремлялся к выбранному ответу, что разгонял волну и заливал аппаратуру исследователей»<sup>26</sup>.

---

<sup>25</sup> *Smith J.* и др. *The Uncertain Response in the Bottlenosed Dolphin (Tursiops Truncatus)* // *Journal of Experimental Psychology: General* 124. 1995. № 4. С. 391–408.

<sup>26</sup> Вероятно, что альтернативные теории, которые не требуют измерений неопределенности, могут объяснить поведение животных в этих экспериментах. Например, когда третий уровень представлен Натуга, мы наблюдаем три возможных ответа: низкая тональность, высокая тональность и «не знаю» (отказ как ответ). Через некоторое время Натуга может понять, что нажатие низких и высоких нот, когда тональность находится посередине, приводит к наказанию и от-

Макаки – обезьяны, которые встречаются по всей Азии (и любят воровать еду у туристов в храмах и святилищах), – тоже легко обучаются отслеживать свою неуверенность в похожих ситуациях. В одном эксперименте макак обучали определять самую большую фигуру на экране компьютера. Затем им нужно было выбрать между двумя иконками. Первая иконка означала рискованную ставку (три кусочка еды в случае правильного ответа, за ошибку еду убирали), в то время как другой, безопасный вариант гарантировал один кусочек еды – обезьяний вариант игры «Кто хочет стать миллионером?». Обезьяны чаще выбирали рискованный вариант, когда отвечали правильно, что красноречиво свидетельствует о метапознании. Что впечатляет еще больше, они сразу же, без дополнительного обучения, справились с оценкой уверенности в своих ответах в другом тесте на память, исключив предположение, что они просто учатся ассоциировать определенные стимулы с ответами разной степени уверенности. С помощью похожего задания исследователи из лаборатории Адама Кепекса, расположенной в Колд-Спринг-Харбор в Нью-Йорке, продемонстрировали, что крысы тоже могут оценивать свою правоту насчет того, какой из двух запа-

---

существованию рыбы. Отказ отвечать – менее рискованный вариант, который позволяет ему перейти быстро на следующий уровень, во время которого он сможет получить рыбу. Возможно, он просто следует простому правилу, что-то в духе, «когда я слышу средний звуковой сигнал, я нажимаю рычаг отказа», так что он не ощущает неуверенности, ответил ли он правильно или нет. Carruthers C. Meta-Cognition in Animals: A Skeptical Look // *Mind & Language*. 2008. № 1. С. 58–89.

хов преобладает в смешанном аромате. Есть даже некоторые свидетельства, что птицы, как и обезьяны, способны перенести свои наработанные метакогнитивные навыки из одного испытания в другое<sup>27</sup>.

Если чувствительность к неопределенности – это фундаментальное свойство работы мозга, то вполне логично, что у людей этот первый кирпичик метапознания можно обнаружить в самом раннем возрасте. Вдохновившись тестами Смита, Луиза Гупиль и Сид Куидер из Высшей нормальной школы в Париже решили исследовать, как неопределенность своих решений отслеживают 18-месячные младенцы. Малышам, сидящим на коленях у матерей, демонстрировали заманчивую игрушку и давали поиграть с ней, возбуждая интерес. Затем им показывали, как игрушку прячут в одну из двух коробок. Наконец после небольшой паузы малышам разрешалось заглянуть в одну из коробок, чтобы достать игрушку.

---

<sup>27</sup> Kornell N. *и др.* Transfer of Metacognitive Skills and Hint Seeking in Monkeys // *Psychological Science* 18. 2007. № 1. С. 64–71; Shields W. E. *и др.* Uncertain Responses by Humans and Rhesus Monkeys (*Macaca mulatta*) in a Psychophysical Same-Different Task // *Journal of Experimental Psychology: General* 126. 1997. № 2. С. 147; Kepecs A. *и др.* Neural Correlates, Computation and Behavioural Impact of Decision Confidence // *Nature* 455. 2008. № 7210. С. 227–231; Fujita K. *и др.* Are Birds Metacognitive? // *Foundations of Metacognition*. Oxford: Oxford University Press, 2012. С. 50–61. Шесть голубей и две из трех бентамок были более склонны к рисковому варианту, когда правильно выполняли задание на визуальный поиск. Два голубя также продемонстрировали стабильное обобщение этой метакогнитивной способности на другие цветовые наборы.

# Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

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