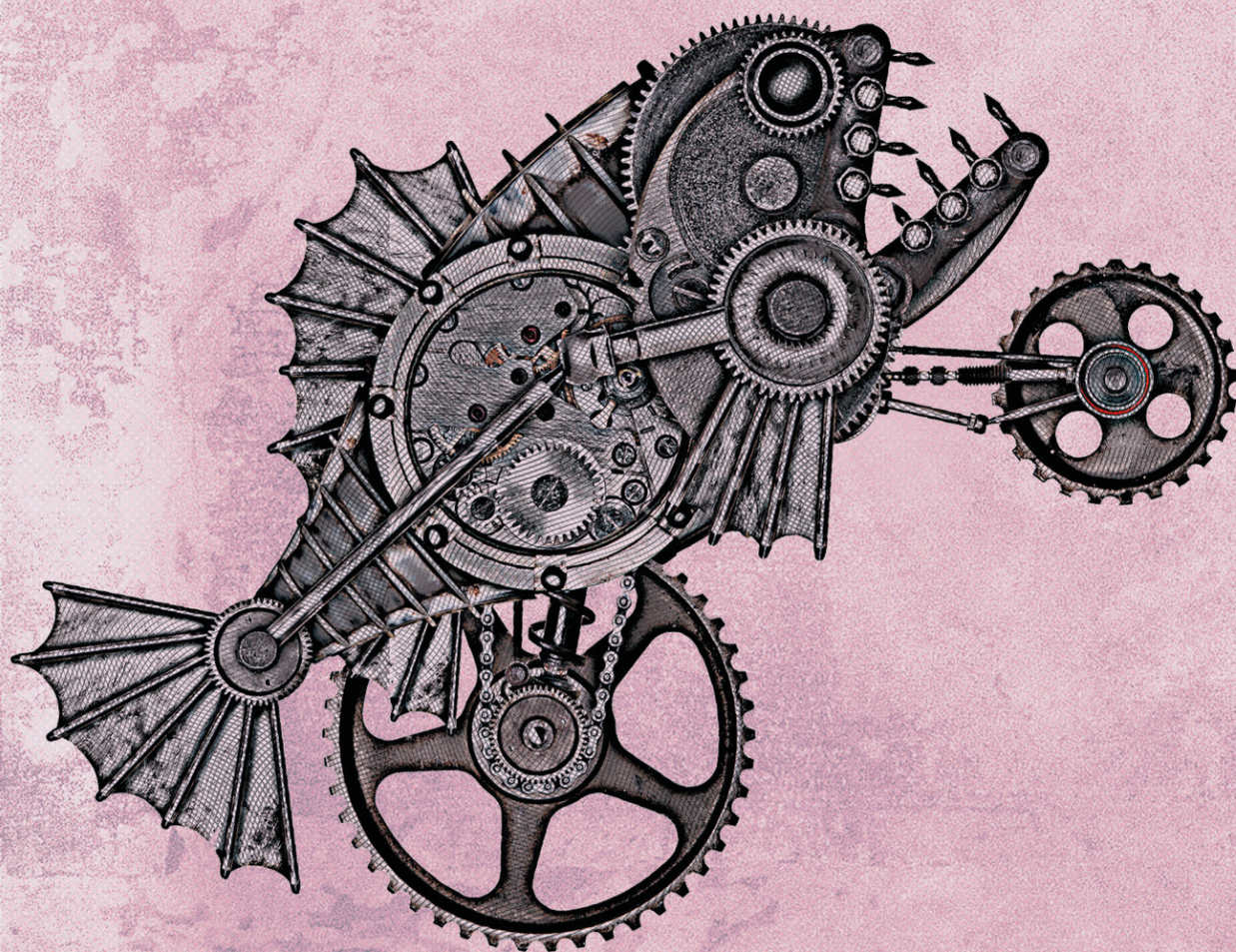




МАРВИН
МИНСКИЙ



МАШИНА ЭМОЦИЙ

Шляпа Оливера Сакса

Марвин Минский

Машина эмоций

«Издательство АСТ»

2006

УДК 159.95
ББК 88.3

Минский М.

Машина эмоций / М. Минский — «Издательство АСТ»,
2006 — (Шляпа Оливера Сакса)

ISBN 978-5-17-114660-3

Марвин Минский – американский ученый, один из основоположников в области теории искусственного интеллекта, сооснователь лаборатории информатики и искусственного интеллекта в Массачусетском технологическом институте, лауреат премии Тьюринга за 1969 год, медали «Пионер компьютерной техники» (1995 год) и еще целого списка престижных международных и национальных наград. Что такое человеческий мозг? Машина, – утверждает Марвин Минский, – сложный механизм, который, так же, как и любой другой механизм, состоит из набора деталей и работает в заданном алгоритме. Но если человеческий мозг – механизм, то что представляют собой человеческие эмоции? Какие процессы отвечают за растерянность или уверенность в себе, за сомнения или прозрения? За ревность и любовь, наконец? Минский полагает, что эмоции – это всего лишь еще один способ мышления, дополняющий основной мыслительный аппарат новыми возможностями. В формате PDF A4 сохранен издательский макет.

УДК 159.95

ББК 88.3

ISBN 978-5-17-114660-3

© Минский М., 2006

© Издательство АСТ, 2006

Содержание

| | |
|-----------------------------------|----|
| Введение | 6 |
| Глава 1. Влюбленность | 12 |
| 1.1. Увлечение | 12 |
| Конец ознакомительного фрагмента. | 14 |

Марвин Минский Машина эмоций

Marvin Minsky
The Emotion Machine

© Marvin Minsky, 2006
© Перевод. В. Дегтярева, 2018
© Перевод. А. Курьшева, 2018
© Издание на русском языке AST Publishers, 2020

* * *

Посвящается Глории, Маргарет, Генри и Джулиане

Прошлые соавторы

Пуш Сингх
Джон Маккарти
Сеймур Пейперт
Оливер Селфридж
Р. Дж. Соломонофф

Импраймеры

Джордж А. Миллер
Дж. К. Р. Ликлайдер
Эндрю М. Глисон
Соломон Лефшец
Уоррен С. Мак-Каллох
Клод Э. Шеннон

Поддержка

Джеффри Эпштейн
Кацухико Ниси
Николас Негропonte
Гарвардское общество научных сотрудников
Управление военно-морских исследований
Корпорация *Toshiba*

Введение

Нора Джойс своему мужу Джеймсу: Почему ты не пишешь книг, которые можно читать?

Я надеюсь, что эта книга будет полезна всем, кто хочет понять, как работает человеческий разум, ищет совета о том, как мыслить более продуктивно, или стремится создать более умные машины. Читатели, которые интересуются темой искусственного интеллекта, наверняка найдут в ней кое-что полезное. Также она, как мне кажется, представляет интерес для психологов, неврологов, философов и ученых в сфере теории вычислительных машин, поскольку в ней выдвигается много новых теорий по самым сложным вопросам, с которыми сталкиваются эти специалисты.

Мы все восхищаемся крупными достижениями в области искусства, а также естественных и гуманитарных наук, но редко отдаем должное тому, чего достигаем сами в ходе своей повседневной жизни. Мы узнаем вещи, которые видим, понимаем слова, которые слышим, и запоминаем то, что пережили, чтобы позже использовать эти знания в применении к новым проблемам и возможностям.

А еще мы делаем кое-что замечательное, на что не способны никакие другие существа: всякий раз, когда нам не удается достигнуть цели, думая привычным способом, *мы начинаем думать о самих своих мыслях* – и если такое «рефлексивное мышление» показывает, где мы поступили не так, это позволяет нам сформировать новый и более продуктивный способ думать. Однако мы пока очень мало знаем о том, как мозгу удастся все это делать. Как работает воображение? Чем порождается сознание? Что такое эмоции, чувства и мысли? Как мы вообще умудряемся думать?

Сравните это с прогрессом, который наблюдается в изучении материальных феноменов. Что такое твердые вещества, жидкости и газы? Что такое цвета, звуки и температуры? Что такое силы, напряжения и натяжения? Какова природа энергии? В наше время почти все подобные тайны уже объяснены с помощью очень небольшого числа простых законов, таких как уравнения, открытые физиками вроде Ньютона, Максвелла, Эйнштейна и Шредингера.

Естественно, психологи, стремясь подражать физикам, пустились на поиски некоего компактного свода законов, который мог бы объяснить, что происходит внутри нашего мозга. Однако в этой книге вы увидите аргументы в пользу того, что это начинание обречено на неудачу, потому что такого простого набора законов не существует, ведь мозг состоит из сотен участков, каждый из которых сформировался для выполнения определенных видов задач; одни распознают ситуации, другие приказывают мышцам выполнять действия, третьи формулируют цели и планы, а четвертые накапливают и используют огромное количество знаний. И хотя нам еще мало известно о том, как работают эти сотни мозговых центров, мы знаем, что их строение основано на информации, которая содержится в десятках тысяч унаследованных генов, – так что каждый участок мозга функционирует по-своему, опираясь на несколько иной набор законов.

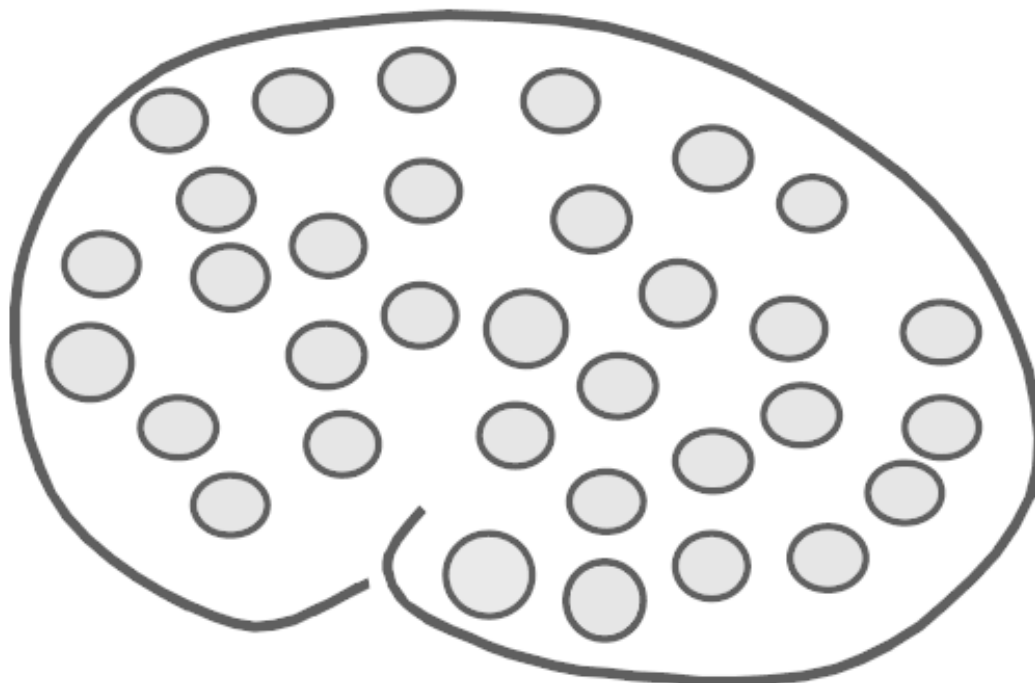
Вместе с осознанием того факта, что в нашем мозге действуют столь сложные механизмы, приходит понимание, что мы должны двигаться ровно в противоположном направлении от физиков: вместо поиска простых объяснений нам нужно искать более сложные способы объяснить самые привычные психические феномены.

Например, значения таких слов, как «чувства», «эмоции» или «сознание», кажутся настолько естественными, ясными и прямыми, что мы даже не видим, с какой стороны начинать их обдумывать. Однако в данной книге утверждается, что каждое из этих слов пытается описать действие обширной сети процессов, происходящих внутри нашего мозга. Например,

в главе четвертой будет показано, что термин «сознание» отсылает нас более чем к двадцати различным процессам такого типа!

Казалось бы, превращая то, что поначалу считалось простым и ясным, в проблемы, которые теперь представляются более сложными, мы только ухудшаем ситуацию. Однако если взглянуть на вещи масштабнее, это повышение сложности на самом деле упростит нам задачу. Ведь, разделив старую загадку на части, мы получим вместо большой задачи несколько новых и более мелких, каждая из которых по-прежнему может быть сложной, но уже не будет казаться неразрешимой. Кроме того, глава девятая разъяснит нам, что отношение к себе как к сложному механизму не должно умалить в нас чувства собственного достоинства, а только усилить чувство ответственности за свои действия.

Процесс деления старых больших вопросов на более мелкие мы в этой книге начнем с того, что изобразим типичный мозг как состоящий из огромного множества частей, которые будем называть «ресурсы».



Это изображение пригодится всякий раз, как нам захочется объяснить какую-либо психическую деятельность (например, гнев, любовь или смущение), для того чтобы продемонстрировать, как данное состояние ума становится результатом функционирования определенного набора умственных ресурсов. Например, состояние, называемое «гнев», судя по всему, пробуждает ресурсы, помогающие нам реагировать с необычайной скоростью и силой, одновременно подавляя ресурсы, которые, наоборот, заставляют планировать поведение и действовать более осмотрительно; иными словами, гнев заменяет осторожность агрессивностью, а сочувствие – враждебностью. Аналогичным образом чувство, называемое «страх», оперирует ресурсами так, чтобы заставить вас отступить.

Читатель: *Иногда у меня бывает такое состояние, когда все кажется радостным и ярким. А бывает иначе (хотя ничего не изменилось): все окружающее начинает казаться мрачным и темным, и друзья называют*

меня грустным или подавленным. Почему я становлюсь жертвой таких состояний – или настроений, чувств, предрасположенностей – и чем вызывается все их странное влияние?

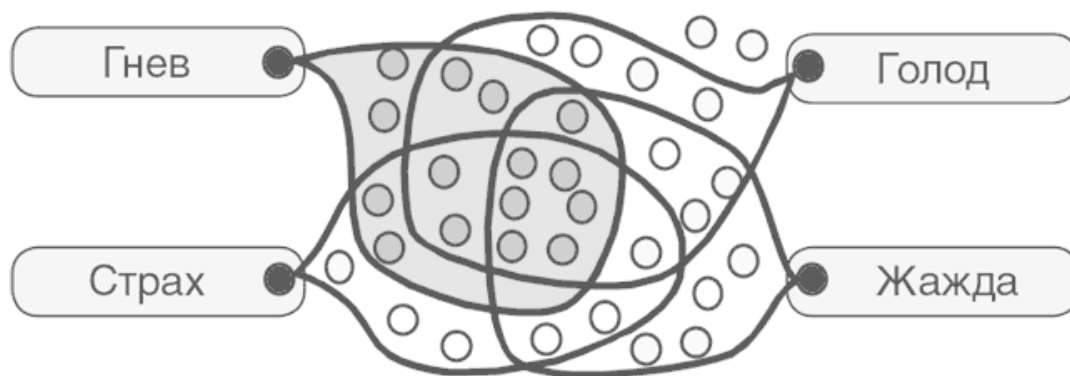
На это существует несколько распространенных ответов: «Такие изменения вызываются действием химических веществ в мозге», или «Это результат избытка стресса», или «Они появляются, если поддаваться депрессивным мыслям». Однако в подобных заявлениях почти ничего не говорится о том, как именно функционируют эти процессы, тогда как идея выбора определенных ресурсов способна более конкретно показать, каким образом меняется способ думать. Например, глава первая начнется с размышления об очень знакомом феномене:

Когда кто-то, кого вы знаете, влюбляется, возникает ощущение, как будто перед вами почти что другой человек: он думает иначе, у него появляются иные цели и приоритеты. Как будто где-то щелкнули выключателем и запустили новую программу.

Что такого может происходить внутри мозга, что объясняло бы подобные перемены в мыслительных процессах? Вот как подходит к вопросу эта книга:

Каждое из наших основных «эмоциональных состояний» вызывается включением определенных ресурсов и одновременным выключением других, в результате чего изменяется то, как ведет себя наш мозг.

Но что именно активирует такие наборы ресурсов? В последующих главах мы постараемся доказать, что нашему мозгу необходим и еще один тип ресурсов, которые мы будем называть «Критики». Каждый из них специализируется на распознавании определенного состояния, а затем активирует конкретный набор других ресурсов. Некоторые из наших Критиков заложены в нас с рождения и являются источником определенных «инстинктивных» реакций – таковы, например, гнев, голод, страх и жажда, которые развились, помогая выжить нашим предкам. Если точнее, гнев и страх развивались для защиты и обеспечения безопасности, в то время как голод и жажда – для добычи пропитания.



Однако по мере обучения и роста мы также развиваем умение активировать другие, новые наборы используемых ресурсов – они приводят к возникновению психических состояний того типа, который мы считаем скорее «интеллектуальным», чем «эмоциональным». Например, всякий раз, когда что-то кажется вам трудным, ваш ум начинает переключаться между различными способами мышления, выбирая разные наборы ресурсов, которые помогли бы разделить проблему на более мелкие части, найти подходящие аналогии или выудить решение из воспоминаний – или даже обратиться за помощью к другому человеку. Иными словами:

Каждый из наших основных способов мышления вызывается включением определенных ресурсов и одновременным выключением других – в результате чего изменяется то, как себя ведет наш мозг.

Далее в книге будут изложены аргументы в пользу того, что именно этим, возможно, объясняется уникальная находчивость, присущая человеческому виду. Например, в первых главах мы попытаемся показать, как этим можно объяснить такие ментальные состояния, как Любовь, Привязанность, Горе и Депрессия с точки зрения того, как они используют наши ресурсы. Последующие главы расскажут о том же, но в приложении к более «интеллектуальному» мышлению.

Читатель: *Кажется странным, что вы дали одинаковое описание как эмоциям, так и обычному мышлению. Но мышление по сути своей рационально – это что-то сухое, отстраненное и логичное, – в то время как эмоции оживляют его, добавляя иррациональные чувства и предубеждения.*

Существует традиционное представление о том, что эмоции *привносят* дополнительные параметры в простые, одномерные мысли так же, как художники используют цвета, чтобы придать сложность черно-белым рисункам. Тем не менее в этой книге утверждается иное: многие из наших эмоциональных состояний, напротив, проистекают из того, что некие определенные способы думать начинают *подавлять* функционирование определенных ресурсов! Например, в главе первой будет рассказано о «романтическом увлечении» как о состоянии, в котором мы подавляем некоторые из ресурсов, с помощью которых могли бы заметить в объекте нашего интереса какие-либо недостатки. Кроме того, я считаю, что существование чисто логического рационального мышления – это миф, поскольку на разум человека всегда влияют его предположения, ценности и цели.

Читатель: *Мне все еще кажется, что ваш взгляд на эмоции слишком многое игнорирует. Например, такие эмоциональные состояния, как страх и отвращение, затрагивают не только мозг, но и тело – например, когда мы чувствуем дискомфорт в груди или в животе, частое сердцебиение, слабость, когда нас бросает в дрожь или пот.*

Я согласен с тем, что такая позиция может показаться слишком радикальной, но иногда, чтобы изучить новую идею, необходимо отодвинуть старые в сторону – по крайней мере временно. Например, по самому распространенному мнению, эмоции глубоко связаны с состояниями тела. Однако в главе седьмой будет высказано противоположное мнение: в ней части тела рассматриваются как ресурсы, которые мозг может использовать для изменения (или поддержания) своих психических состояний! Например, иногда вы можете заставить себя не отступить от запланированного действия, просто сохранив определенное выражение лица.

Итак, хотя эта книга и называется «Машина эмоций», в ней доказывается, что эмоциональные состояния не особенно отличаются от процессов, которые мы называем «мышление». Эмоции – это определенные способы думать, которые мы используем, чтобы подстегнуть свою находчивость, – за исключением случаев, когда наши страсти разрастаются до того, что парализуют нас. И этот диапазон способов мыслить, как представляется, играет настолько значительную роль в том, что мы называем «интеллект», что стоит, пожалуй, использовать «находчивость» как термин. И это касается не только эмоциональных состояний, но и всех наших умственных процессов:

Если вы «понимаете» что-либо только одним способом, это все равно что почти не понимать – потому что, оказавшись в тупике, вы не сможете из него выбраться. Но если вы способны посмотреть на феномен с разных сторон, тогда, если нечто приведет вас в смятение, вы сможете

переключаться между различными точками зрения, пока не найдете ту, которая вам подойдет!

Соответственно, разрабатывая механизмы, имитирующие наш разум, – иными словами, создавая искусственный интеллект, – мы должны следить, чтобы они также обладали в этом плане достаточным разнообразием:

Если программа имеет только один алгоритм работы, то стоит ему оказаться непригодным, и она будет загнана в тупик. Но программа, имеющая несколько вариантов последующих действий, может в таком случае переключиться на другой подход или найти подходящую замену.

Эта идея и является центральной темой настоящей книги, и она категорически расходится с популярным мнением о том, что у каждого человека есть центральное ядро – какой-то невидимый дух или «самость», из которых проистекают все его умственные способности. Кажется уничижительным предполагать, что все наши добродетели вторичны или что мы не заслуживаем признания за свои достижения, поскольку они достаются нам в качестве подарков из какого-то иного источника. Напротив, в моем понимании, чувство собственного достоинства должно порождаться тем, что каждый из нас создал из самого себя: из огромного набора всевозможных способов справляться с различными ситуациями и трудностями. Именно это разнообразие отличает нас от большинства других животных – а также от всех машин, которые мы создали в прошлом, – и каждая глава этой книги рассказывает о некоторых источниках уникальной гибкости человеческого мышления.

Глава 1. Мы рождаемся, уже обладая множеством ментальных ресурсов.

Глава 2. Мы узнаем новое из взаимодействия с другими людьми.

Глава 3. Эмоции – это разные способы думать.

Глава 4. Мы учимся думать о своих недавних мыслях.

Глава 5. Мы учимся думать на нескольких уровнях.

Глава 6. Мы накапливаем множество знаний, основанных на здравом смысле.

Глава 7. Мы переключаемся между способами думать.

Глава 8. Мы находим разные способы репрезентации феноменов.

Глава 9. Мы строим разные модели самих себя.

На протяжении веков психологи искали способ объяснить повседневные процессы, происходящие в нашем мозге, но многие мыслители и сегодня рассматривают природу разума как загадку. В самом деле, по-прежнему широко распространено мнение, что разум соткан из компонентов, которые могут существовать только в живых организмах, что ни одна машина не способна чувствовать или думать, беспокоиться о том, что может с ней произойти, и даже просто осознавать факт собственного существования – или что машина никогда не сумела бы породить идею, способную привести к созданию великого живописного полотна или симфонии.

Книга, лежащая перед вами, преследует все эти цели сразу: выработать теорию о том, как работает человеческий мозг, и разработать механизм, способный чувствовать и думать. Затем полученные идеи можно попытаться использовать как для изучения самих себя, так и для разработки искусственного интеллекта.

Примечание. В этой книге термин «ресурс» используется для описания того, что в моей предыдущей книге *The Society of Mind*¹ называлось «агентом». Я внес это изменение, потому что слишком многие читатели предполагали, что «агент» – это нечто персонифицированное

¹ Марвин Минский. Сообщество разума. М.: АСТ, 2018.

(как, например, туристический агент), нечто способное действовать независимо, и что ментальные агенты могут сотрудничать почти так же, как это делают люди. Однако все обстоит совсем иначе: чаще всего ресурсы заточены под то, чтобы выполнять определенные функции для некоторых других, и неспособны напрямую взаимодействовать с большинством остальных ресурсов человека. Подробнее о том, как соотносятся эти две книги, см. в статье 2003 года за авторством Пуша Сингха (Сингх, 2003), который помог развить многие из идей, собранных в этой книге.

Глава 1. Влюбленность

1.1. Увлечение

*Мои глаза в тебя не влюблены, –
Они твои пороки видят ясно.
А сердце ни одной твоей вины
Не видит и с глазами не согласно².*

Шекспир

Многие люди считают абсурдом думать о человеке как о машине, поэтому мы часто слышим подобные высказывания:

Читатель: *Конечно, от машин есть польза. Мы можем заставить их складывать бесконечные столбцы цифр или собирать другие машины на заводах. Но ничто механическое никогда не сумеет испытывать такие искренние чувства, как любовь.*

В наши дни никто не удивляется, что машины действуют логически, ведь логика основана на простых и ясных правилах – из тех, что легко использовать компьютерам. Но *Любовь*, по самой своей натуре, как скажут некоторые, нельзя объяснить с помощью механики, и мы никогда не сможем создать машины, обладающие такими человеческими свойствами, как чувства, эмоции и сознание.

Что такое *Любовь* и как она работает? Хотим ли мы узнать о ней больше, или она из тех вещей, что мы на самом деле не хотим понимать? Давайте посмотрим, как наш друг Чарльз пытается описать свое недавнее увлечение:

Я только что влюбился в изумительную девушку. Ни о чем другом не могу думать. Моя милая безгранично идеальна – красота ее неопишима, характер безупречен, а ум невероятен. Нет ничего, что я бы ради нее не сделал.

На поверхности такие утверждения кажутся вполне позитивными, они состоят из прилагательных в превосходной степени. Но обратите внимание на одну странную особенность: сколько в этом отрывке частиц «не» и приставок «без». Очевидно, что на самом деле это отрицательные утверждения, которые описывают говорящего их человека.

*Изумительную. Неопишима.
(Не могу понять, что именно привлекает меня в ней.)
Ни о чем другом не могу думать.
(Мой мозг практически перестал работать.)
Безгранично идеальна. Невероятен.
(Ни один разумный человек в подобное не поверит.)
Характер безупречен.
(Я утратил способность к критическому мышлению.)
Нет ничего, что я бы ради нее не сделал.
(Я забросил большинство своих обычных дел.)*

² Сонет 141, перевод С. Я. Маршака.

Наш друг все эти утверждения воспринимает как положительные. Это чувство приносит ему счастье и стимулирует к деятельности, к тому же избавляет от уныния и одиночества. Но что, если большинство из этих приятных эффектов – следствие того, что подавляет мысли о том, что на самом деле говорит ему любимая:

*О, Чарльз, женщине нужны определенные вещи. Ей нужно, чтобы ее любили, хотели, ценили, баловали, осыпали лестью, холили и лелеяли. Ей нужны сочувствие, нежность, преданность, понимание, влюбленность, лесть и поклонение – я ведь не слишком многого прошу, правда, Чарльз?*³

Таким образом любовь может заставить нас не обращать внимания на большинство дефектов и недостатков и относиться к изъянам как к достоинствам, пусть даже, как сказал Шекспир, мы все же можем отчасти их осознавать:

*Когда клянешься мне, что вся ты сплошь
Служить достойна правды образцом,
Я верю, хоть и вижу, как ты лжешь*⁴.

Мы одинаково умело обманываем себя – не только в личной жизни, но и когда имеем дело с абстрактными идеями. Здесь мы тоже часто закрываем глаза, когда речь заходит о противоречиях между нашими убеждениями. Посмотрите, что пишет Ричард Фейнман:

Это было самое начало, и идея казалась мне такой очевидной, что я влюбился в нее без памяти. И, как и с влюбленностью в женщину, это возможно только тогда, когда не слишком много о ней знаешь и поэтому не видишь ее недостатков. Недостатки проявляются позже, но к этому времени любовь уже достаточно сильна, чтобы удержать тебя рядом. Так я и придерживался этой теории, несмотря на все сложности, благодаря юношескому энтузиазму

³ Адаптированная цитата из комедийной программы Барри Тука и Марти Фельдмана: см. Barry Took, Marty Feldman, *Round the Horne*, BBC Radio, 1966.

⁴ Перевод С. Я. Маршака.

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.