STEVEN PINKER

Rationality

WHAT IT IS

↓

WHY IT SEEMS SCARCE

↓

WHY IT MATTERS

Author of the *Sunday Times* bestseller
ENLIGHTENMENT NOW

allen lane

Contents

About the Author

Steven Pinker is an experimental cognitive scientist. Currently Johnstone Professor of Psychology at Harvard, he has also taught at Stanford and MIT. He has won many prizes for his research, teaching and his eleven books, including *The Language Instinct*, *How the Mind Works, The Blank Slate, The Better Angels of Our Nature,* and *The Sense of Style.* He is a member of the National Academy of Sciences, a two-time Pulitzer Prize finalist, a Humanist of the Year, a recipient of nine honorary doctorates, and one of *Foreign Policy*'s 'World's Top 100 Public Intellectuals' and *Time*'s '100 Most Influential People in the World Today'.

ALSO BY STEVEN PINKER

Language Learnability and Language Development

Learnability and Cognition

The Language Instinct

How the Mind Works

Words and Rules

The Blank Slate

The Stuff of Thought

The Better Angels of Our Nature

Language, Cognition, and Human Nature: Selected Articles

The Sense of Style

Enlightenment Now

EDITED BY STEVEN PINKER

Visual Cognition

Connections and Symbols (with Jacques Mehler)

Lexical and Conceptual Semantics (with Beth Levin)

The Best American Science and Nature Writing 2004

To Roslyn Wiesenfeld Pinker

What is a man, If his chief good and market of his time
Be but to sleep and feed? A beast, no more.
Sure he that made us with such large discourse,
Looking before and after, gave us not
That capability and godlike reason To fust in us unus'd.

—*HAMLET*

Preface

Rationality ought to be the lodestar for everything we think and do. (If you disagree, are your objections rational?) Yet in an era blessed with unprecedented resources for reasoning, the public sphere is infested with fake news, quack cures, conspiracy theories, and "post-truth" rhetoric.

How can we make sense of making sense—and its opposite? The question is urgent. In the third decade of the third millennium, we face deadly threats to our health, our democracy, and the livability of our planet. Though the problems are daunting, solutions exist, and our species has the intellectual wherewithal to find them. Yet among our fiercest problems today is convincing people to accept the solutions when we do find them.

Commentaries by the thousands have lamented our shortfall of reason, and it's become conventional wisdom that people are simply irrational. In social science and the media, the human being is portrayed as a caveman out of time, poised to react to a lion in the grass with a suite of biases, blind spots, fallacies, and illusions. (The *Wikipedia* entry for cognitive biases lists almost two hundred.)

Yet as a cognitive scientist I cannot accept the cynical view that the human brain is a basket of delusions. Hunter-gatherers—our ancestors and contemporaries—are not nervous rabbits but cerebral problem solvers. A list of the ways in which we're stupid can't explain why we're so smart: smart enough to have discovered the laws of nature, transformed the planet, lengthened and enriched our lives, and, not least, articulated the rules of rationality that we so often flout.

To be sure, I am among the first to insist that we can understand human nature only by considering the mismatch between the environment in which we evolved and the environment we find ourselves in today. But the world to which our minds are adapted is not just the Pleistocene savannah. It's any nonacademic, nontechnocratic milieu—which is to say, most of human experience—in which the modern instruments of rationality like statistical formulas and datasets are unavailable or inapplicable. As we shall see, when people are given problems that are closer to their lived reality and framed in the ways in which they naturally encounter the world, they are not as witless as they appear. Not that this gets us off the hook. Today we do have refined instruments of reason, and we are best off, as individuals and as a society, when we understand and apply them.

This book grew out of a course I taught at Harvard which explored the nature of rationality and the puzzle of why it seems to be so scarce. Like many psychologists, I love to teach the arresting, Nobel Prize–winning discoveries of the infirmities that afflict human reason, and consider them to be among the deepest gifts to knowledge that our science has contributed. And like many, I believe that the benchmarks of rationality that people so often fail to measure up to should be a goal of education and popular science. Just as citizens should grasp the basics of history, science, and the written word, they should command the intellectual tools of sound reasoning.

These include logic, critical thinking, probability, correlation and causation, the optimal ways to adjust our beliefs and commit to decisions with uncertain evidence, and the yardsticks for making rational choices alone and with others. These tools of reasoning are indispensable in avoiding folly in our personal lives and public policies. They help us calibrate risky choices, evaluate dubious claims, understand baffling paradoxes, and gain insight into life's vicissitudes and tragedies. But I knew of no book that tried to explain them all.

The other inspiration for this book was my realization that for all its fascination, the cognitive psychology curriculum left me ill equipped to answer the questions I was most frequently asked when I told people I was teaching a course on rationality. Why do people believe that Hillary Clinton ran a child sex ring out of a pizzeria, or that jet contrails are really mind-altering drugs dispersed by a secret government program? My standard lecture bullet points like "the gambler's fallacy" and "base-rate neglect" offered little insight into just the enigmas that are making human irrationality so pressing an issue today. Those enigmas drew me into new territories, including the nature of rumor, folk wisdom, and conspiratorial thinking; the contrast between rationality within an individual and in a community; and the distinction between two modes of believing: the reality mindset and the mythology mindset.

Finally, though it may seem paradoxical to lay out rational arguments for rationality itself, it's a timely assignment. Some people pursue the opposite paradox, citing reasons (presumably rational ones, or why should we listen?) that rationality is overrated, such as that logical personalities are joyless and repressed, analytical thinking must be subordinated to social justice, and a good heart and reliable gut are surer routes to well-being than tough-minded logic and argument. Many act as if rationality is obsolete—as if the point of argumentation is to discredit one's adversaries rather than collectively reason our way to the most defensible beliefs. In an era in which rationality seems both more threatened and more essential than ever, *Rationality* is, above all, an affirmation of rationality.

A MAJOR THEME of this book is that none of us, thinking alone, is rational enough to consistently come to sound conclusions: rationality emerges from a community of reasoners who spot each other's fallacies. In that spirit I thank the reasoners who made this book more rational. Ken Binmore, Rebecca Newberger Goldstein, Gary King, Jason Nemirow, Roslyn Pinker, Keith Stanovich, and Martina Wiese incisively commented on the first draft. Charleen Adams, Robert Aumann, Joshua Hartshorne, Louis Liebenberg, Colin McGinn, Barbara Mellers, Hugo Mercier, Judea Pearl, David Ropeik, Michael Shermer, Susanna Siegel, Barbara Spellman, Lawrence Summers, Philip Tetlock, and Juliani Vidal reviewed chapters in their areas of expertise. Many questions arose as I planned and wrote the book, and they were answered by Daniel Dennett, Emily-Rose Eastop, Baruch Fischhoff, Reid Hastie, Nathan Kuncel, Ellen Langer, Jennifer Lerner, Beau Lotto, Daniel Loxton, Gary Marcus, Philip Maymin, Don Moore, David Myers, Robert Proctor, Fred Shapiro, Mattie Toma, Jeffrey Watumull, Jeremy Wolfe, and Steven Zipperstein. I counted on the expert transcription, fact-checking, and reference hunting by Mila Bertolo, Martina Wiese, and Kai Sandbrink, and on original data analyses by Bertolo, Toma, and Julian De Freitas. Also appreciated were the questions and suggestions from the students and teaching staff of General Education 1066: Rationality, especially Mattie Toma and Jason Nemirow.

Special thanks go to my wise and supportive editor, Wendy Wolf, for working with me on this book, our sixth; to Katya Rice, for copy-editing our ninth; and to my literary agent, John Brockman, for his encouragement and advice on our ninth. I appreciate as well the support over many years from Thomas Penn, Pen Vogler, and Stefan McGrath of Penguin UK. Ilavenil Subbiah once again designed the graphics, and I thank her for her work and her encouragement.

Rebecca Newberger Goldstein played a special role in the conception of this book, because it is she who impressed on me that realism and reason are ideals that must be singled out and defended. Love and gratitude go as well to the other members of my family: Yael and Solly; Danielle; Rob, Jack, and David; Susan, Martin, Eva, Carl, and Eric; and my mother, Roslyn, to whom this book is dedicated.

1

How Rational an Animal?

> Man is a rational animal. So at least we have been told. Throughout a long life I have searched diligently for evidence in favor of this statement. So far, I have not had the good fortune to come across it.
>
> —Bertrand Russell[1]

> He that can carp in the most eloquent or acute manner at the weakness of the human mind is held by his fellows as almost divine.
>
> —Baruch Spinoza[2]

Homo sapiens means wise hominin, and in many ways we have earned the specific epithet of our Linnaean binomial. Our species has dated the origin of the universe, plumbed the nature of matter and energy, decoded the secrets of life, unraveled the circuitry of consciousness, and chronicled our history and diversity. We have applied this knowledge to enhance our own flourishing, blunting the scourges that immiserated our ancestors for most of our existence. We have postponed our expected date with death from thirty years of age to more than seventy (eighty in developed countries), reduced extreme poverty from ninety percent of humanity to less than nine, slashed the rates of death from war twentyfold and from famine a hundredfold.[3] Even when the ancient bane of pestilence rose up anew in the twenty-first century, we identified the cause within days, sequenced its genome within weeks, and administered vaccines within a year, keeping its death toll to a fraction of those of historic pandemics.

The cognitive wherewithal to understand the world and bend it to our advantage is not a trophy of Western civilization; it's the patrimony of our species. The San of the Kalahari Desert in southern Africa are one of the world's oldest peoples, and their foraging lifestyle, maintained until recently, offers a glimpse of the ways in which humans spent most of their existence.[4] Hunter-gatherers don't just chuck spears at passing animals or help themselves to fruit and nuts growing around them.[5] The tracking scientist Louis Liebenberg, who has worked with the San for decades, has described how they owe their survival to a scientific mindset.[6] They reason their way from fragmentary data to remote conclusions with an intuitive grasp of logic, critical thinking, statistical reasoning, causal inference, and game theory.

The San engage in persistence hunting, which puts to use our three most conspicuous traits: our two-leggedness, which enables us to run efficiently; our hairlessness, which enables us to dump heat in hot climates; and our big heads, which enable us to be rational. The San deploy this rationality to track the fleeing animals from their hoofprints, effluvia, and other spoor, pursuing them until they keel over from exhaustion and heat stroke.[7] Sometimes the San track an animal along one of its habitual pathways, or, when a trail goes cold, by searching in widening circles around the last known prints. But often they track them by reasoning.

Hunters distinguish dozens of species by the shapes and spacing of their tracks, aided by their grasp of cause and effect. They may infer that a deeply pointed track comes from an agile springbok, which needs a good grip, whereas a flat-footed track comes from a heavy kudu, which has to support its weight. They can sex the animals from the configuration of their tracks and the relative location of their urine to their hind feet and droppings. They use these categories to make syllogistic deductions: steenbok and duiker can be run down in the rainy season because the wet sand forces open their hooves and stiffens their joints; kudu and eland can be run down in the dry season because they tire easily in loose sand. It's the dry season and the animal that left these tracks is a kudu; therefore, this animal can be run down.

The San don't just pigeonhole animals into categories but make finer-grained logical distinctions. They tell individuals apart within a species by reading their hoofprints, looking for telltale nicks and variations. And they distinguish an individual's permanent traits, like its species and sex, from transient conditions like fatigue, which they infer from signs of hoof-dragging and stopping to rest. Defying the canard that premodern peoples have no concept of time, they estimate the age of an animal from the size and crispness of its hoofprints, and can date its spoor by the freshness of tracks, the wetness of saliva or droppings, the angle of the sun relative to a shady resting place, and the palimpsest of superimposed tracks from other animals. Persistence hunting could not succeed without those logical niceties. A hunter can't track just any gemsbok from among the many that have left tracks, but only the one he has been pursuing to exhaustion.

The San also engage in critical thinking. They know not to trust their first impressions, and appreciate the dangers of seeing what they want to see. Nor will they accept arguments from authority: anyone, including a young upstart, may shoot down a conjecture or come up with his own until a consensus emerges from the disputation. Though it's mainly the men who hunt, the women are just as knowledgeable at interpreting spoor, and Liebenberg reports that one young woman, !Nasi, "put the men to shame."[8]

The San adjust their credence in a hypothesis according to how diagnostic the evidence is, a matter of conditional probability. A porcupine foot, for instance, has two proximal pads while a honey badger has one, but a padprint may fail to register on hard ground. This means that though the probability that a track will have one padprint given that it was made by a honey badger is high, the inverse probability, that a track was made by a honey badger given that it has one padprint, is lower (since it could also be an incomplete porcupine track). The San do not confuse these conditional probabilities: they know that since two padprints could only have been left by a porcupine, the probability of a porcupine given two padprints is high.

The San also calibrate their credence in a hypothesis according to its prior plausibility. If tracks are ambiguous, they will assume they come from a commonly occurring species; only if the evidence is definitive will they conclude that they come from a rarer one.[9] As we shall see, this is the essence of Bayesian reasoning.

Another critical faculty exercised by the San is distinguishing causation from correlation. Liebenberg recalls: "One tracker, Boroh//xao, told me that when the [lark] sings, it dries out the soil, making the roots good to eat. Afterwards, !Nate and /Uase told me that Boroh//xao was wrong—it is not the *bird* that dries out the soil, it is the *sun* that dries out the soil. The bird is only *telling* them that the soil will dry out in the coming months and that it is the time of the year when the roots are good to eat."[10]

The San use their knowledge of the causal texture of their environment not just to understand how it is but to imagine how it might be. By playing out scenarios in their mind's eye, they can think several steps ahead of the animals in their world and devise intricate snares to trap them. One end of a springy branch is anchored in the ground and the stick is bent in half; the other is tied to a noose camouflaged with twigs and sand and held in place by a trigger. They place the snares at the openings of barriers they have built around an antelope's resting place, and guide the animal into the deadly spot with a hurdle the antelope must clear. Or they lure an ostrich to a snare by spotting its tracks under a camelthorn tree (whose pods are an ostrich delicacy) and leaving a conspicuous bone that's too big for the ostrich to swallow, which draws its attention to a smaller but still unswallowable bone, which leads to a still smaller bone, the bait in the snare.

Yet for all the deadly effectiveness of the San's technology, they have survived in an unforgiving desert for more than a hundred thousand years without exterminating the animals they depend on. During a drought, they think ahead to what would happen if they killed the last plant or animal of its kind, and they spare the members of the threatened species.[11] They tailor their conservation plans to the different vulnerabilities of plants, which cannot migrate but recover quickly when the rains return, and animals, which can survive a drought but build back their numbers slowly. And they enforce these conservation efforts against the constant temptation of poaching (everyone feeling they should exploit the scarce species, because if they don't, everyone else will) with an extension of the norms of reciprocity and collective well-being that govern all their resources. It is unthinkable for a San hunter not to share meat with an empty-handed bandmate, or to exclude a neighboring band driven from their drought-stricken territory, since they know that memories are long and some day fortunes may reverse.

THE SAPIENCE OF THE SAN makes the puzzle of human rationality acute. Despite our ancient capacity for reason, today we are flooded with reminders of the fallacies and follies of our fellows. People gamble and play the lottery, where they are guaranteed to lose, and fail to invest for their retirement, where they are guaranteed to win. Three quarters of Americans believe in at least one phenomenon that defies the laws of science, including psychic healing (55 percent), extrasensory perception (41 percent), haunted houses (37 percent), and ghosts (32 percent)—which also means that some people believe in houses haunted by ghosts without believing in ghosts.[12] In social media, fake news (such as JOE BIDEN CALLS TRUMP SUPPORTERS "DREGS OF SOCIETY" and FLORIDA MAN ARRESTED FOR TRANQUILIZING AND RAPING

ALLIGATORS IN THE EVERGLADES) is diffused farther and faster than the truth, and humans are more likely to spread it than bots.[13]

It has become commonplace to conclude that humans are simply irrational—more Homer Simpson than Mr. Spock, more Alfred E. Neuman than John von Neumann. And, the cynics continue, what else would you expect from descendants of hunter-gatherers whose minds were selected to avoid becoming lunch for leopards? But evolutionary psychologists, mindful of the ingenuity of foraging peoples, insist that humans evolved to occupy the "cognitive niche": the ability to outsmart nature with language, sociality, and know-how.[14] If contemporary humans seem irrational, don't blame the hunter-gatherers.

How, then, can we understand this thing called rationality which would appear to be our birthright yet is so frequently and flagrantly flouted? The starting point is to appreciate that rationality is not a power that an agent either has or doesn't have, like Superman's X-ray vision. It is a kit of cognitive tools that can attain particular goals in particular worlds. To understand what rationality is, why it seems scarce, and why it matters, we must begin with the ground truths of rationality itself: the ways an intelligent agent *ought* to reason, given its goals and the world in which it lives. These "normative" models come from logic, philosophy, mathematics, and artificial intelligence, and they are our best understanding of the "correct" solution to a problem and how to find it. They serve as an aspiration for those who want to be rational, which should mean everyone. A major goal of this book is to explain the most widely applicable normative tools of reason; they are the subjects of chapters 3 to 9.

Normative models also serve as benchmarks against which we can assess how human schlemiels *do* reason, the subject matter of psychology and the other behavioral sciences. The many ways in which ordinary people fall short of these benchmarks have become famous through the Nobel Prize–winning research of Daniel Kahneman, Amos Tversky, and other psychologists and behavioral economists.[15] When people's judgments deviate from a normative model, as they so often do, we have a puzzle to solve. Sometimes the disparity reveals a genuine irrationality: the human brain cannot cope with the complexity of a problem, or it is saddled with a bug that cussedly drives it to the wrong answer time and again.

But in many cases there is a method to people's madness. A problem may have been presented to them in a deceptive format, and when it is translated into a mind-friendlier guise, they solve it. Or the normative model may itself be correct only in a particular environment, and people accurately sense that they are not in that one, so the model doesn't apply. Or the model may be designed to bring about a certain goal, and, for better or worse, people are after a different one. In the chapters to come, we will see examples of all these extenuating circumstances. The penultimate chapter will lay out how some of today's florid outbursts of irrationality may be understood as the rational pursuit of goals other than an objective understanding of the world.

Though explanations of irrationality may absolve people of the charge of outright stupidity, to understand is not to forgive. Sometimes we can hold people to a higher standard. They can be taught to spot a deep problem across its superficial guises. They can be goaded into applying their best habits of thinking outside their comfort zones. And they can be inspired to set their sights higher than self-defeating or collectively destructive goals. These, too, are aspirations of the book.

Since a recurring insight of the study of judgment and decision making is that humans become more rational when the information they're dealing with is more vivid and relevant, let me turn to examples. Each of these classics—from math, logic, probability, and forecasting—exposes a quirk in our reasoning and will serve as a preview of the normative standards of rationality (and the ways in which people depart from them) in the chapters to come.

Three Simple Math Problems

Everyone remembers being tormented in high school by algebra problems about where the train that left Eastford traveling west at 70 miles per hour will meet the train that left Westford, 260 miles away, traveling east at 60 miles per hour. These three are simpler; you can do them in your head:

- A smartphone and a case cost $110 in total. The phone costs $100 more than the case. How much does the case cost?
- It takes 8 printers 8 minutes to print 8 brochures. How long would it take 24 printers to print 24 brochures?
- On a field there is a patch of weeds. Every day the patch doubles in size. It takes 30 days for the patch to cover the whole field. How long did it take for the patch to cover half the field?

The answer to the first problem is $5. If you're like most people, you guessed $10. But if that were right, the phone would cost $110 ($100 more than the case), and the total for the pair would be $120.

The answer to the second question is 8 minutes. It takes a printer 8 minutes to print a brochure, so as long as there are as many printers as there are brochures and they are working simultaneously, the time it takes to print the brochures is the same.

The answer to the third problem is 29 days. If the weed patch doubles every day, then by working backwards from when the field was completely covered, we may infer that it was half covered the day before.

The economist Shane Frederick gave these questions (with different examples) to thousands of university students. Five out of six got at least one of them wrong; one in three got them *all* wrong.[16] Yet each question has a simple answer that almost everyone understands when it's pointed out. The problem is that people's heads are turned by superficial features of the problem which they mistakenly think are relevant to the answer, such as the round numbers 100 and 10 in the first problem and the fact that the number of printers is the same as the number of minutes in the second.

Frederick calls his low-tech battery the Cognitive Reflection Test, and suggests that it exposes a cleavage between two cognitive systems, later made famous by Kahneman (his sometime coauthor) in the 2011 bestseller *Thinking, Fast and Slow.* System 1 operates rapidly and effortlessly, and it seduces us with the wrong answers; System 2 requires concentration, motivation, and the application of learned rules, and it allows us to grasp the right ones. No one thinks these are literally two anatomical systems in the brain; they are two modes of operation which cut across many brain structures. System 1 means snap judgments; System 2 means thinking twice.

The lesson of the Cognitive Reflection Test is that blunders of reasoning may come from thoughtlessness rather than ineptitude.[17] Even students at the math-proud Massachusetts Institute of Technology averaged only two out of three correct. Performance does correlate with math skill, as you'd expect, but it also correlates with patience. People who describe themselves as not impulsive, and who would rather wait for a larger payment in a month than get a smaller one right away, are less likely to fall into the traps.[18]

The first two items feel like trick questions. That is because they give details which, in the back-and-forth of conversation, would be relevant to what the speaker is asking, but in these examples are designed to lead the hearer astray. (People do better when the smartphone costs, say, $73 more than the case and the combination costs $89.)[19] But of course real life is also baited with garden paths and siren songs that lure us from good decisions, and resisting them is a part of being rational. People who fall for the alluring but wrong answers on the Cognitive Reflection Test appear to be less rational in other ways, such as turning down lucrative offers that require a bit of waiting or a bit of risk.

And the third problem, the one with the weed patch, is not a trick question but taps a real cognitive infirmity. Human intuition doesn't grasp exponential (geometric) growth, namely something that rises at a rising rate, proportional to how large it already is, such as compound interest, economic growth, and the spread of a contagious disease.[20] People mistake it for steady creep or slight acceleration, and their imaginations don't keep up with the relentless doubling. If you deposit $400 a month into a retirement account that earns 10 percent annually, how big will your nest egg be after forty years? Many people guess around $200,000, which is what you get by multiplying 400 by 12 by 110% by 40. Some know that that can't be right and adjust their guess upward, but never enough. Almost no one gets the correct answer: $2.5 *million*. People with a shaky grasp of exponential growth have been found to save less for retirement and to take on more credit-card debt, two roads to penury.[21]

A failure to visualize exponential blastoff can trap experts as well—even experts in cognitive biases. When Covid-19 arrived in the United States and Europe in February 2020, several social scientists (including two heroes of this book, though not Kahneman himself) opined that people were irrationally panicking because they had read about a gruesome case or two and got carried away by the "availability bias" and "probability neglect." The objective risk at the time, they noted, was lower than that of the flu or strep throat, which everyone accepts calmly.[22] The fallacy of the fallacy scolds was to underestimate the accelerating rate at which a disease as contagious as Covid can spread, with each patient not only infecting new ones but turning each of them into an infector. The single confirmed American death on March 1 grew in successive weeks to 2, 6, 40, 264, 901, and 1,729 deaths per day, adding up to more than 100,000 deaths by June 1 and soon making it the most lethal hazard in the country.[23] Of course the authors of these obscure op-eds cannot be blamed for the insouciance which lulled so many leaders and citizens into dangerous complacency, but their comments show how deeply rooted cognitive biases can be.

Why do people misunderestimate exponential growth, as George W. Bush might have put it? In the great tradition of the physician in the Molière play who explained that opium makes people sleepy because of its dormitive power, social scientists attribute the errors to an "exponential growth bias." Less circularly, we

might point to the ephemerality of exponential processes in natural environments (prior to historical innovations like economic growth and compound interest). Things that can't go on forever don't, and organisms can multiply only to the point where they deplete, foul, or saturate their environments, bending the exponential curve into an S. This includes pandemics, which peter out once enough susceptible hosts in the herd are killed or develop immunity.

A Simple Logic Problem

If anything lies at the core of rationality, it must surely be logic. The prototype of a rational inference is the syllogism "If P then Q. P. Therefore, Q." Consider a simple example.

Suppose the coinage of a country has a portrait of one of its eminent sovereigns on one side and a specimen of its magnificent fauna on the other. Now consider a simple if-then rule: "If a coin has a king on one side, then it has a bird on the other." Here are four coins, displaying a king, a queen, a moose, and a duck. Which of the coins do you have to turn over to determine whether the rule has been violated?

If you're like most people, you said "the king" or "the king and the duck." The correct answer is "the king and the moose." Why? Everyone agrees you have to turn over the king, because if you failed to find a bird on the reverse it would violate the rule in so many words. Most people know there's no point in turning over the queen, because the rule says "If king, then bird"; it says nothing about coins with a queen. Many say you should turn over the duck, but when you think about it, that coin is irrelevant. The rule is "If king, then bird," not "If bird, then king": if the duck shared the coin with a queen, nothing would be amiss. But now consider the moose. If you turned that coin over and found a king on the obverse, the rule "If king, then bird" would have been transgressed. The answer, then, is "the king and the moose." On average, only 10 percent of people make those picks.

The Wason selection task (named after its creator, the cognitive psychologist Peter Wason) has been administered with various "If P then Q" rules for sixty-five years. (The original version used cards with a letter on one side and a number on the other and a rule like "If there is a D on one side, there is a 3 on the other.") Time and again people turn over the P, or the P and the Q, and fail to turn over the not-Q.[24] It's not that they're incapable of understanding the right answer. As with the Cognitive Reflection Test, as soon as it is explained to them they slap themselves on the forehead and accept it.[25] But their unreflective intuition, left to its own devices, fails to do the logic.

What does this tell us about human rationality? A common explanation is that it reveals our *confirmation bias*: the bad habit of seeking evidence that ratifies a belief and being incurious about evidence that might falsify it.[26] People think that dreams are omens because they recall the time when they dreamt a relative had a mishap

and she did, but they forget about all the times when a relative was fine after they dreamt she had a mishap. Or they think immigrants commit a lot of crime because they read in the news about an immigrant who robbed a store, but don't think about the larger number of stores robbed by native-born citizens.

Confirmation bias is a common diagnosis for human folly and a target for enhancing rationality. Francis Bacon (1561–1626), often credited with developing the scientific method, wrote of a man who was taken to a church and shown a painting of sailors who had escaped a shipwreck thanks to their holy vows. "Aye," he remarked, "but where are they painted that were drowned after their vows?"[27] He observed, "Such is the way of all superstitions, whether in astrology, dreams, omens, divine judgments, or the like; wherein men, having a delight in such vanities, mark the events where they are fulfilled, but where they fail, although this happened much oftener, neglect and pass them by."[28] Echoing a famous argument by the philosopher Karl Popper, most scientists today insist that the dividing line between science and pseudoscience is whether advocates of a hypothesis deliberately search for evidence that could falsify it and accept the hypothesis only if it survives.[29]

How can humans make it through the day with an inability to apply the most elementary rule of logic? Part of the answer is that the selection task is a peculiar challenge.[30] It doesn't ask people to apply the syllogism to make a useful deduction ("Here's a coin with a king; what's on the other side?") or to test the rule in general ("Is the rule true of the country's coinage?"). It asks whether the rule applies specifically to each of a handful of items before them on the table. The other part of the answer is that people do apply logic when the rule involves the shoulds and shouldn'ts of human life rather than arbitrary symbols and tokens.

Suppose the Post Office sells fifty-cent stamps for third-class mail but requires ten-dollar stamps for Express Mail. That is, properly addressed mail must follow the rule "If a letter is labeled Express Mail, it must have a ten-dollar stamp." Suppose the label and the stamp don't fit on the same side of the envelope, so a postal worker has to turn envelopes over to check to see if the sender has followed the rule. Here are four envelopes. Imagine that you are a postal worker. Which ones do you have to turn over?

The correct answer once again is P and not-Q, namely the Express envelope and the one with the fifty-cent stamp. Though the problem is logically equivalent to the four-coin problem, this time almost everyone gets it right. The content of a logical problem matters.[31] When an if-then rule implements a contract involving permissions and duties—"If you enjoy a benefit, you must pay a cost"—then a violation of the rule (take the benefit, don't pay the cost) is equivalent to cheating, and people intuitively know what it takes to catch a cheater. They don't check up on people who aren't enjoying the benefit or people who have paid a cost, neither of whom could be trying to get away with something.

Cognitive psychologists debate exactly what kinds of content temporarily turn people into logicians. They can't be just any concrete scenarios, but must embody the kinds of logical challenges that we became attuned to as we developed into adults and perhaps when we evolved into humans. Monitoring a privilege or duty is

one of these logic-unlocking themes; monitoring danger is another. People know that to verify compliance with the precaution "If you ride a bicycle, then you must wear a helmet," they have to check that a child on a bicycle is wearing a helmet and that a child without a helmet does not get onto a bicycle.

Now, a mind that can falsify a conditional rule when the violations are tantamount to cheating or danger is not exactly a logical mind. Logic, by definition, is about the form of statements, not their content: how Ps and Qs are connected by IF, THEN, AND, OR, NOT, SOME, and ALL, regardless of what the Ps and Qs stand for. Logic is a crowning achievement of human knowledge. It organizes our reasoning with unfamiliar or abstract subject matter, such as the laws of government and science, and when implemented in silicon it turns inert matter into thinking machines. But what the untutored human mind commands is not a general-purpose, content-free tool, with formulas like "[IF P THEN Q] is equivalent to NOT-[P AND NOT Q]," into which any P and Q can be plugged. It commands a set of more specialized tools that bake together the content relevant to the problem with the rules of logic (without those rules, the tools wouldn't work). It isn't easy for people to extricate the rules and wield them in novel, abstract, or apparently meaningless problems. That's what education and other rationality-enhancing institutions are for. They augment the *ecological rationality* we are born and grow up with—our horse sense, our street smarts—with the broader-spectrum and more potent tools of reasoning perfected by our best thinkers over the millennia.[32]

A Simple Probability Problem

One of the most famous television game shows from the heyday of the genre from the 1950s to the 1980s was *Let's Make a Deal*. Its host, Monty Hall, achieved a second kind of fame when a dilemma in probability theory, loosely based on the show, was named after him.[33] A contestant is faced with three doors. Behind one of them is a sleek new car. Behind the other two are goats. The contestant picks a door, say Door 1. To build suspense, Monty opens one of the other two doors, say Door 3, revealing a goat. To build the suspense still further, he gives the contestant an opportunity either to stick with their original choice or to switch to the unopened door. You are the contestant. What should you do?

Almost everyone stays.[34] They figure that since the car was placed behind one of the three doors at random, and Door 3 has been eliminated, there is now a fifty-fifty chance each that the car will be behind Door 1 or Door 2. Though there's no harm in switching, they think, there's no benefit either. So they stick with their first choice out of inertia, pride, or anticipation that their regret after an unlucky switch would be more intense than their delight after a lucky one.

The Monty Hall dilemma became famous in 1990 when it was presented in the "Ask Marilyn" column in *Parade*, a magazine inserted in the Sunday edition of hundreds of American newspapers.[35] The columnist was Marilyn vos Savant, known at the time as "the world's smartest woman" because of her entry in the *Guinness Book of World Records* for the highest score on an intelligence test. Vos Savant wrote that you should switch: the odds of the car being behind Door 2 are two in three, compared with one in three for Door 1. The column drew ten thousand letters, a thousand of them from PhDs, mainly in mathematics and statistics, most of whom said she was wrong. Here are some examples:

> You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame!
>
> —Scott Smith, PhD, University of Florida

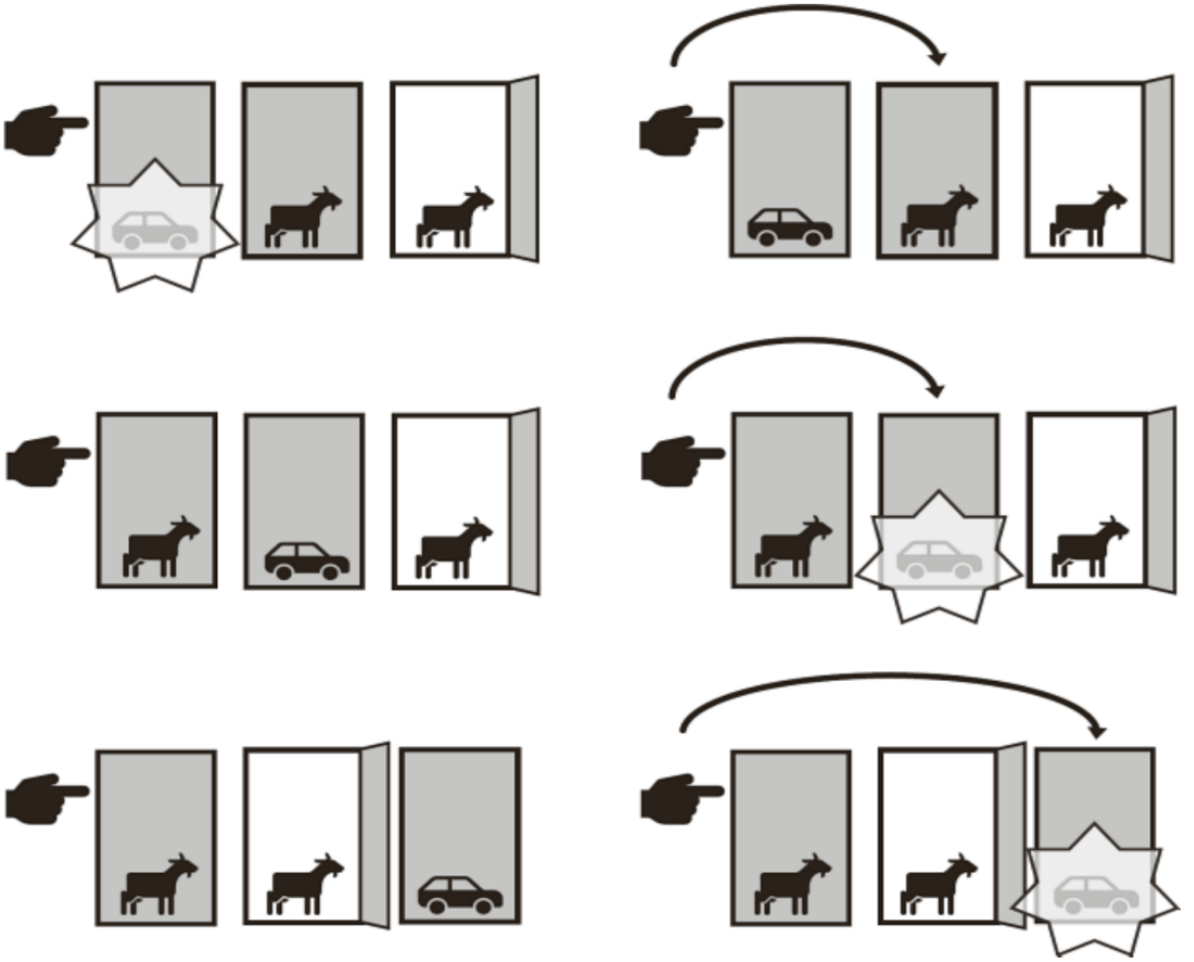
> I am sure you will receive many letters on this topic from high school and college students. Perhaps you should keep a few addresses for help with future columns.
>
> —W. Robert Smith, PhD, Georgia State University

> Maybe women look at math problems differently than men.
>
> —Don Edwards, Sunriver, Oregon[36]

Among the objectors was Paul Erdös (1913–1996), the renowned mathematician who was so prolific that many academics boast of their "Erdös number," the length of the shortest chain of coauthorships linking them to the great theoretician.[37]

But the mansplaining mathematicians were wrong and the world's smartest woman was right. You should switch. It's not that hard to see why. There are three possibilities for where the car could have been placed. Let's consider each door and count up the number of times out of the three that you would win with each strategy. You picked Door 1, but of course that's just a label; as long as Monty follows the rule "Open an unselected door with a goat; if both have goats, pick one at random," the odds come out the same whichever door you picked.

Suppose your strategy is "Stay" (left column in the figure). If the car is behind Door 1 (top left), you win. (It doesn't matter which of the other doors Monty opened, because you're not switching to either.) If the car is behind Door 2 (middle left), you lose. If the car is behind Door 3 (bottom left), you lose. So the odds of winning with the "Stay" strategy are one in three.



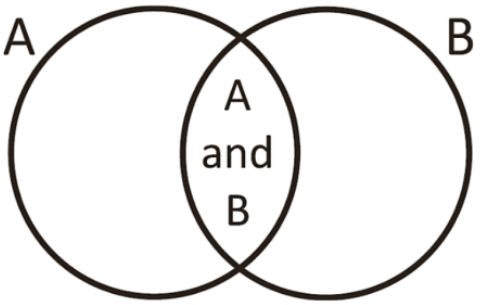
Now suppose your strategy is "Switch" (right column). If the car is behind Door 1, you lose. If the car is behind Door 2, Monty would have opened Door 3, so you would switch to Door 2 and win. If the car is behind Door 3, he would have opened Door 2, so you would switch to Door 3 and win. The odds of winning with the "Switch" strategy are two in three, double the odds of staying.

It's not rocket surgery.[38] Even if you don't work through the logical possibilities, you could play a few rounds yourself with cutouts and toys and tot up the outcomes, as Hall himself did to convince a skeptical journalist. (Nowadays, you can play it online.)[39] Or you could pursue the intuition "Monty knows the answer and gave me a clue; it would be foolish not to act on it." Why did the mathematicians, professors, and other bigshots get it so wrong?

Certainly there were failures of critical thinking coming from sexism, ad hominem biases, and professional jealousy. Vos Savant is an attractive and stylish woman with no initials after her name who wrote for a recipe- and gossip-filled rag and bantered on late-night talk shows.[40] She defied the stereotype of a mathematician, and her celebrity and bragging rights from *Guinness* made her a big fat target for a takedown.

But part of the problem is the problem itself. Like the teasers in the Cognitive Reflection and Wason selection tests, something about the Monty Hall dilemma is designed to bring out the stupid in our System 1. But in this case System 2 is not much brighter. Many people can't swallow the correct explanation even when it's pointed out to them. This included Erdös, who, violating the soul of a mathematician, was convinced only when he saw the game repeatedly simulated.[41]

probability of picking an even-numbered spade out of a deck of cards, for example (even and spade), has to be less than the probability of picking a spade, because some spades are not even numbers.



In each pair of world events, the second scenario is a conjunction of events, one of which is the event in the first scenario. For example, "Iran tests a nuclear weapon and Saudi Arabia develops a nuclear weapon" is a conjunction that embraces "Saudi Arabia develops a nuclear weapon" and must have a smaller chance of happening, since there are other scenarios in which the Saudis might go nuclear (to counter Israel, to flaunt hegemony over the Persian Gulf, and so on). By the same logic, Maduro resigning the presidency has to be more likely than Maduro resigning the presidency after a series of strikes.

What are people thinking? A class of events described by a single statement can be generic and abstract, with nothing for the mind to hold on to. A class of events described by a conjunction of statements can be more vivid, especially when they spell out a story line we can watch in the theater of our imagination. Intuitive probability is driven by imaginability: the easier something is to visualize, the likelier it seems. This entraps us into what Tversky and Kahneman call the conjunction fallacy, in which a conjunction is more intuitively probable than either of its elements.

The forecasts of pundits are often driven by vivid narratives, probability be damned.[47] A famous 1994 cover story in *The Atlantic* by the journalist Robert Kaplan predicted "The Coming Anarchy."[48] Kaplan forecasted that in the first decades of the twenty-first century, wars would be fought over scarce resources such as water; Nigeria would conquer Niger, Benin, and Cameroon; world wars would be fought over Africa; the United States, Canada, India, China, and Nigeria would break apart, whereupon American regions with many Latinos would erase the border with Mexico while Alberta would merge with Montana; crime would climb in American cities; AIDS would get worse and worse; together with a dozen other calamities, crises, and crackups. Yet as the article was becoming a sensation (including with

President Bill Clinton, who passed it around the White House), the number of civil wars, the proportion of people without access to clean water, and the rate of American crime were sinking like stones.[49] Within three years an effective treatment for AIDS would begin to decimate its death toll. And more than a quarter century later, national borders have barely budged.

The conjunction fallacy was first illustrated by Tversky and Kahneman with an example that has become famous as "the Linda problem":[50]

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Please indicate the probability of each of these statements:
>
> Linda is a teacher in elementary school.
> Linda is active in the feminist movement.
> Linda is a psychiatric social worker.
> Linda is a bank teller.
> Linda is an insurance salesperson.
> Linda is a bank teller and is active in the feminist movement.

Respondents judged that it was likelier that Linda was a feminist bank teller than that she was a bank teller: once again, the probability of A and B was judged to be higher than the probability of A alone. The dated vignette, with its baby-boomer "Linda," backhanded compliment "bright," passé protests, and declining occupation, betrays its early-1980s vintage. But as any psychology instructor knows, the effect is easily replicable, and today, highly intelligent Amanda who marches for Black Lives Matter is still deemed likelier to be a feminist registered nurse than a registered nurse.

The Linda problem engages our intuitions in a particularly compelling way. Unlike the selection task, where people make errors when the problem is abstract ("If P then Q") and get it right when it is couched in certain real-life scenarios, here everyone agrees with the abstract law "prob(A and B) ≤ prob(A)" but are upended when it is made concrete. The biologist and popular science writer Stephen Jay Gould spoke for many when he commented, "I know that the [conjunctive] statement is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—'but she can't just be a bank teller; read the description.'"[51]

That little homunculus can be exploited by skilled persuaders. A prosecutor with little to work with but a corpse washed up on a beach may spin a yarn on how her husband might, hypothetically, have smothered her and dumped the body so he could marry his mistress and start a business with the insurance money. The defense attorney could tell a competing shaggy-dog story in which she could, in theory, have been the victim of a late-night purse-snatching that went horribly awry. Each conjectural detail should make the scenario less likely, according to the laws of probability, yet each can make it more compelling. As Pooh-Bah says in *The Mikado,* it's all "merely corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative."[52]

The conjunction rule is a basic law of mathematical probability, and you don't need to think in numbers to understand it. This made Tversky and Kahneman pessimistic about people's intuitive sense of probability, which they argued is driven

by representative stereotypes and available memories rather than on a systematic reckoning of possibilities. They rejected the idea that "inside every incoherent person there is a coherent one trying to get out."[53]

Other psychologists are more charitable. As we saw with the Monty Hall dilemma, "probability" has several meanings, including physical propensity, justified strength of belief, and frequency in the long run. Still another sense is provided by the *Oxford English Dictionary*: "the appearance of truth, or likelihood of being realized, which any statement or event bears in the light of present evidence."[54] People faced with the Linda problem know that "frequency in the long run" is irrelevant: there's only one Linda, and either she is a feminist bank teller or she isn't. In any coherent conversation the speaker would supply biographical details for a reason, namely to lead the listener to a plausible conclusion. According to the psychologists Ralph Hertwig and Gerd Gigerenzer, people may have rationally inferred that the relevant meaning of "probability" in this task is not one of the mathematical senses in which the conjunction rule applies, but the nonmathematical sense of "degree of warrant in light of the present evidence," and they sensibly followed where the evidence pointed.[55]

In support of the charitable reading, many studies, beginning with ones by Tversky and Kahneman themselves, show that when people are *encouraged* to reason about probability in the sense of relative frequency, rather than being left to struggle with the enigmatic concept of the probability of a single case, they are likelier to obey the conjunction rule. Imagine a thousand women like Linda. How many of them do you think are bank tellers? How many of them do you think are bank tellers who are active in the feminist movement? Now the homunculus is quiet; a coherent person tries to get out. The rate of conjunction errors plummets.[56]

So is the conjunction fallacy, the quintessential demonstration of human probability blindness, an artifact of ambiguous wording and leading questions? Tversky and Kahneman insisted that it isn't. They noted that people commit the fallacy even when they are invited to *bet* on the possibilities (yes, a majority prefer to bet that Linda is a feminist bank teller than that she is a bank teller). And even when the question is couched in frequencies, where people could avoid a conjunction error just by counting the bank tellers in their mind's eye, a substantial minority commit it. This rises to a majority when people consider each alternative in isolation rather than seeing one next to the other, and so their noses are not rubbed in the absurdity of a subset outnumbering a superset.[57]

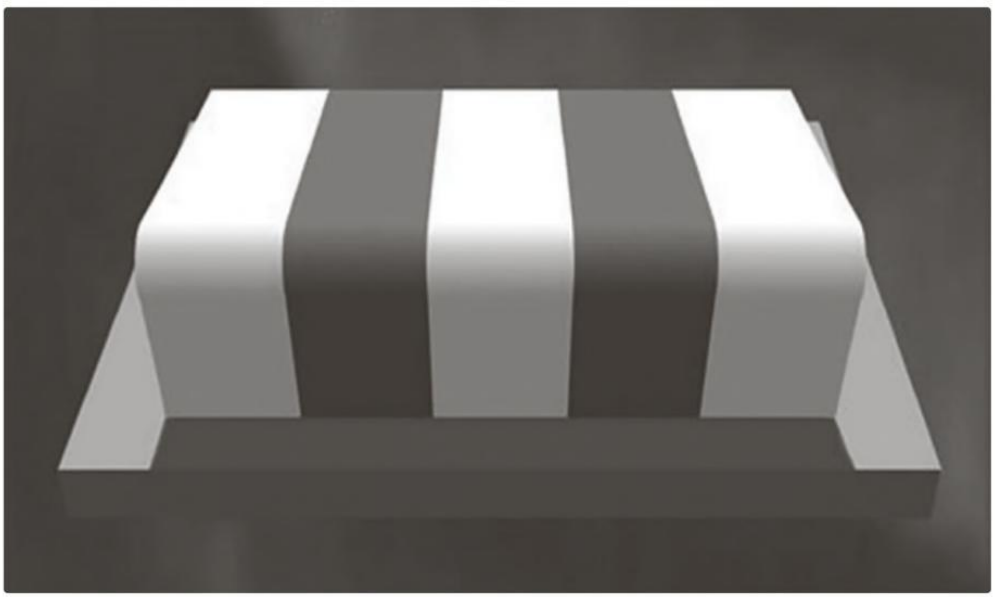
Kahneman has observed that humans are never so irrational as when protecting their pet ideas. So he advocated a new method for resolving scientific controversies to replace the time-honored custom of the rivals taking turns moving the goalposts and talking trash in volleys of rejoinders and replies. In an "adversarial collaboration," the disputants agree in advance on an empirical test that would settle the matter, and invite an arbiter to join them in carrying it out.[58] Fittingly, Kahneman collaborated with Hertwig to see who was right about the Linda problem, recruiting the psychologist Barbara Mellers to act as arbiter. The team of rivals agreed to run three studies that couched the problem in frequencies ("Of 100 people like Linda, how many are ...?") rather than asking about lone Linda. In their write-up of the complex results, the trio reported, "We did not think the experiments would resolve all the issues, nor did this miracle occur." But both sides agreed that people are prone to committing the conjunction fallacy, even when dealing with frequencies. And they agreed that under the right circumstances—the alternatives

are available for comparison side by side, and the wording of the alternatives leaves nothing to the imagination—people can think their way out of the fallacy.

The Moral from Cognitive Illusions

How do we reconcile the rationality that allows our species to live by its wits in ancient and modern environments with the bloopers and gaffes that these brainteasers reveal—the confirmation bias, the overconfidence, the distractibility by concrete details and conversational habits? The classic errors in reasoning are often called "cognitive illusions," and the parallels with the visual illusions familiar from cereal boxes and science museums are instructive. They run deeper than the obvious fact that our eyes and minds can trick us. They explain how our species can be so smart and yet so easily deluded.

Here are two classic illusions, brought to life by the neuroscientist Beau Lotto.[59] The first is a shading illusion. Believe it or not, the dark stripes on the top of the box and the white stripes on the front are identical shades of gray.

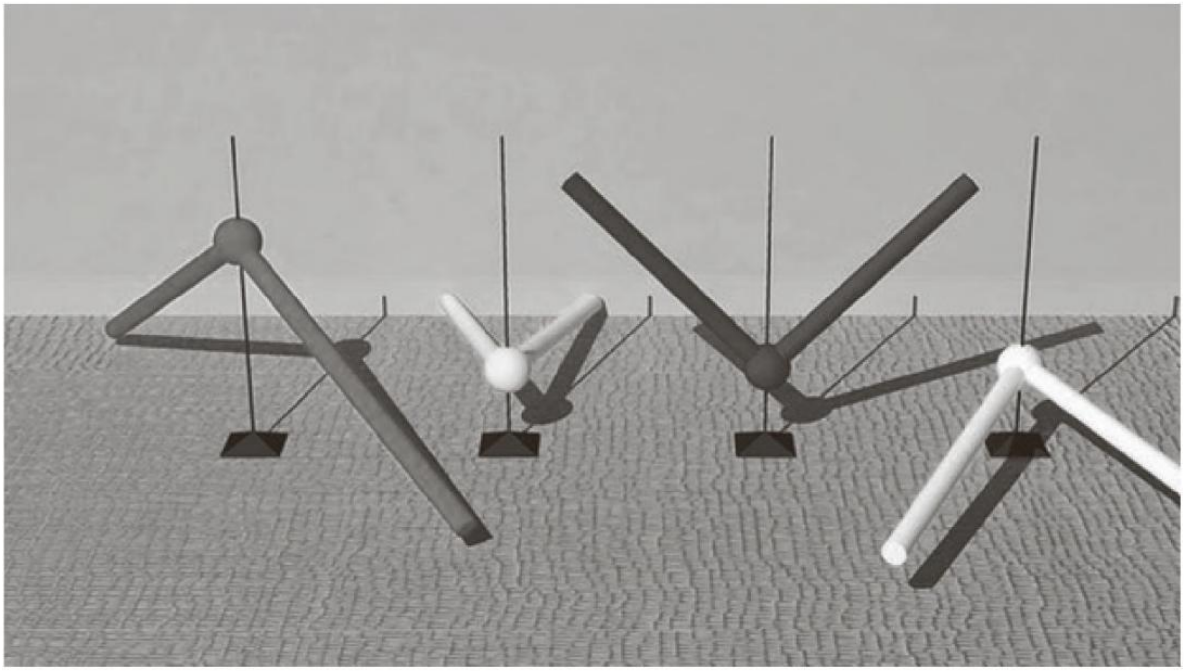


Used by permission of Beau Lotto

The second is a shape illusion: the angles of the four elbows are identical, 90 degrees.

The first takeaway is that we can't always believe our eyes, or, more accurately, the visual System 1 in our brains. The second is that we can recognize our errors using System 2—say, by punching two holes in an index card and laying it over the first figure, and by aligning the corner of the card with the elbows in the second.

But the wrong takeaway is that the human visual system is a buggy contraption that constantly fools us with figments and mirages. The human visual system is one of the wonders of the world. It is a precision instrument that can detect a single photon, recognize thousands of shapes, and negotiate rocky trails and high-speed autobahns. It outperforms our best artificial vision systems, which is why at the time of this writing autonomous vehicles have not been loosed on city streets despite decades of R&D. The vision modules of the robocars are apt to mistake a tractor trailer for a billboard, or a traffic sign plastered with stickers for a refrigerator filled with food.[60]



Used by permission of Beau Lotto

The shape and shading illusions are not bugs but features. The goal of the visual system is to provide the rest of the brain with an accurate description of the 3-D shapes and material composition of the objects in front of us.[61] This is a hard problem because the information coming into the brain from the retina doesn't reflect reality directly. The brightness of a patch on the retinal image depends not just on the pigmentation of the surface in the world but on the intensity of the illumination falling on it: a gray patch could have arisen from a black surface illuminated by a bright light or from a white surface illuminated by a dim one. (That's the basis for the illusion called #thedress, which took the internet by storm in 2015.)[62] A shape on the retina depends not just on the 3-D geometry of the object but on its orientation from a vantage point: an acute angle on the retina could be a sharp corner viewed straight on or a right-angled corner foreshortened. The visual system undoes the effects of these distortions, dividing out the intensity of the illumination and inverting the trigonometry of perspective to feed the rest of the brain with a representation that matches real shapes and materials in the world. The intermediate scratch pad in these calculations—the 2-D array of pixels coming in from our retina—is hidden from the reasoning and planning systems of the brain because they would just be distractions.

Thanks to this design, our brains are not very good light meters or protractors, but then they don't have to be (unless we are realist painters). The illusions emerge when people are asked to be just those instruments. The viewer is asked to notice how bright the stripe is, how sharp the angle, *in the picture*. The pictures have been confected so that simple properties—equal brightnesses, right angles—are buried in the scratch pads that the conscious mind ordinarily ignores. If the questions were about the things in the *world* captured in the pictures, our impressions would be correct. The gray stripe really is darker than the white stripe on both the lit and shaded faces of the box; the elbows poised at different tilts really are bent at different angles.

In the same way, cognitive illusions like the ones in this chapter may arise from our setting aside the literal statement of a question as it comes into our brains and

they knew the keys could not be, or if those beliefs could not be justified—if they came, say, from a drug-induced vision or a hallucinated voice rather than observation of the world or inference from some other true belief.

The beliefs, moreover, must be held in service of a goal. No one gets rationality credit for merely thinking true thoughts, like calculating the digits of π or cranking out the logical implications of a proposition ("Either 1 + 1 = 2 or the moon is made of cheese," "If 1 + 1 = 3, then pigs can fly"). A rational agent must have a *goal*, whether it is to ascertain the truth of a noteworthy idea, called theoretical reason, or to bring about a noteworthy outcome in the world, called practical reason ("what is true" and "what to do"). Even the humdrum rationality of seeing rather than hallucinating is in the service of the ever-present goal built into our visual systems of knowing our surroundings.

A rational agent, moreover, must attain that goal not by doing something that just happens to work there and then, but by using whatever knowledge is applicable to the circumstances. Here is how William James distinguished a rational entity from a nonrational one that would at first appear to be doing the same thing:

> Romeo wants Juliet as the filings want the magnet; and if no obstacles intervene he moves toward her by as straight a line as they. But Romeo and Juliet, if a wall be built between them, do not remain idiotically pressing their faces against its opposite sides like the magnet and the filings with the card. Romeo soon finds a circuitous way, by scaling the wall or otherwise, of touching Juliet's lips directly. With the filings the path is fixed; whether it reaches the end depends on accidents. With the lover it is the end which is fixed; the path may be modified indefinitely.[2]

With this definition the case for rationality seems all too obvious: do you want things or don't you? If you do, rationality is what allows you to get them.

Now, this case for rationality is open to an objection. It advises us to ground our beliefs in the truth, to ensure that our inference from one belief to another is justified, and to make plans that are likely to bring about a given end. But that only raises further questions. What *is* "truth"? What makes an inference "justified"? How do we know that means can be found that really do bring about a given end? But the quest to provide the ultimate, absolute, final reason for reason is a fool's errand. Just as an inquisitive three-year-old will reply to every answer to a "why" question with another "Why?," the quest to find the ultimate reason for reason can always be stymied by a demand to provide a reason for the reason for the reason. Just because I believe P implies Q, and I believe P, why should I believe Q? Is it because I also believe [(P implies Q) and P] implies Q? But why should I believe *that*? Is it because I have still another belief, {[(P implies Q) and P] implies Q} implies Q?

This regress was the basis for Lewis Carroll's 1895 story "What the Tortoise Said to Achilles," which imagined the conversation that would unfold when the fleet-footed warrior caught up to (but could never overtake) the tortoise with the head start in Zeno's second paradox. (In the time it took for Achilles to close the gap, the tortoise moved on, opening up a new gap for Achilles to close, ad infinitum.) Carroll was a logician as well as a children's author, and in this article, published in the philosophy journal *Mind,* he imagines the warrior seated on the tortoise's back and responding to the tortoise's escalating demands to justify his arguments by filling up a notebook with thousands of rules for rules for rules.[3] The moral is that reasoning with logical rules at some point must simply be *executed* by a mechanism that is

Indeed, some of our apparent goals are not even really *our* goals—they are the metaphorical goals of our genes. The evolutionary process selects for genes that lead organisms to have as many surviving offspring as possible in the kinds of environments in which their ancestors lived. They do so by giving us motives like hunger, love, fear, comfort, sex, power, and status. Evolutionary psychologists call these motives "proximate," meaning that they enter into our conscious experience and we deliberately try to carry them out. They can be contrasted with the "ultimate" motives of survival and reproduction, which are the figurative goals of our genes—what they would say they wanted if they could talk.[13]

Conflicts between proximate and ultimate goals play out in our lives as conflicts between different proximate goals. Lust for an attractive sexual partner is a proximate motive, whose ultimate motive is conceiving a child. We inherited it because our more lustful ancestors, on average, had more offspring. However, conceiving a child may not be among our proximate goals, and so we may deploy our reason to foil that ultimate goal by using contraception. Having a trusted romantic partner we don't betray and maintaining the respect of our peers are other proximate goals, which our rational faculties may pursue by advising our not-so-rational faculties to avoid dangerous liaisons. In a similar way we pursue the proximate goal of a slim, healthy body by overriding another proximate goal, a delicious dessert, which itself arose from the ultimate goal of hoarding calories in an energy-stingy environment.

When we say someone's acting emotionally or irrationally, we're often alluding to bad choices in these tradeoffs. It often feels great in the heat of the moment to blow your stack when someone has crossed you. But our cooler head may realize that it's better to put a lid on it, to achieve things that make us feel even greater in the long run, like a good reputation and a trusting relationship.

Conflicts among Time Frames

Since not everything happens at once, conflicts between goals often involve goals that are realized at different times. And these in turn often feel like conflicts between different selves, a present self and a future self.[14]

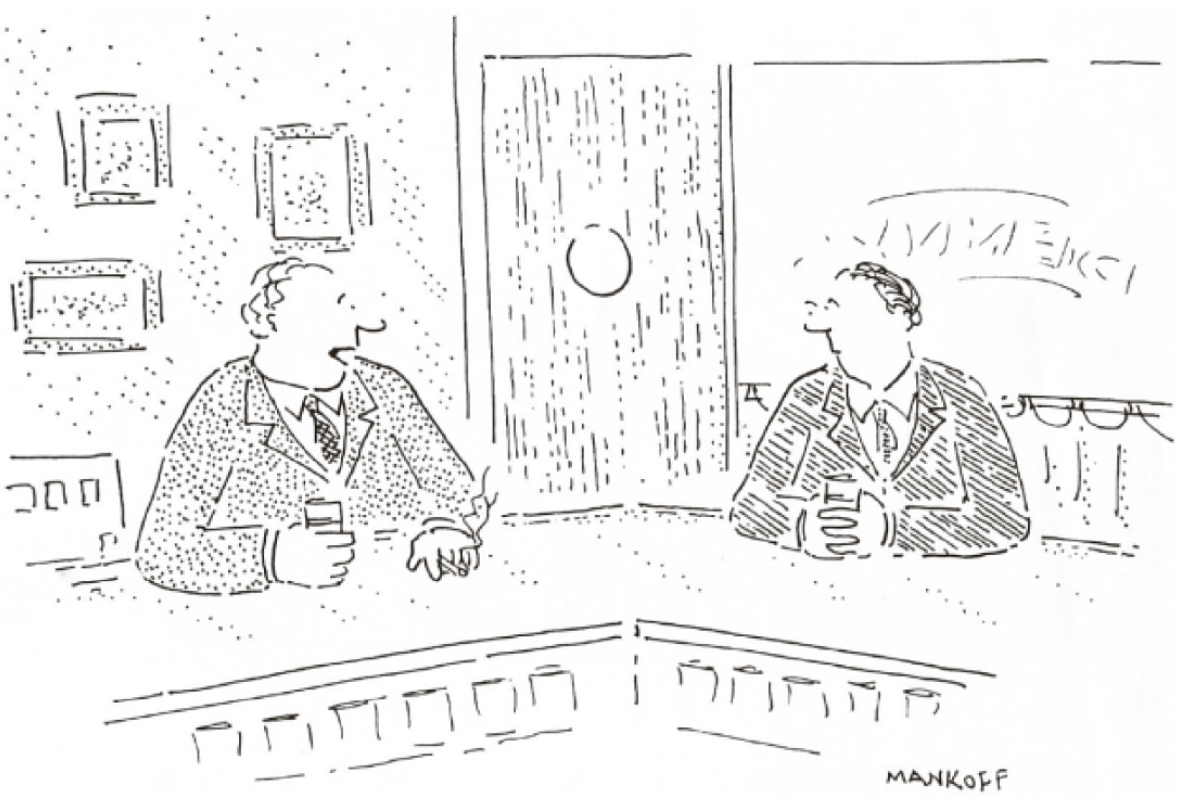
The psychologist Walter Mischel captured the conflict in an agonizing choice he gave four-year-olds in a famous 1972 experiment: one marshmallow now or two marshmallows in fifteen minutes.[15] Life is a never-ending gantlet of marshmallow tests, dilemmas that force us to choose between a sooner small reward and a later large reward. Watch a movie now or pass a course later; buy a bauble now or pay the rent later; enjoy five minutes of fellatio now or an unblemished record in the history books later.

The marshmallow dilemma goes by several names, including self-control, delay of gratification, time preference, and discounting the future.[16] It figures into any analysis of rationality because it helps explain the misconception that too much rationality makes for a cramped and dreary life. Economists have studied the normative grounds for self-control—when we *ought* to indulge now or hold off for later—since it is the basis for interest rates, which compensate people for giving up money now in exchange for money later. They have reminded us that often the rational choice is to indulge now: it all depends on when and how much. In fact this conclusion is already a part of our folk wisdom, captured in aphorisms and jokes.

First, a bird in the hand is worth two in the bush. How do you know that the experimenter will keep his promise and reward you for your patience with two marshmallows when the time comes? How do you know that the pension fund will still be solvent when you retire and the money you have put away for retirement will be available when you need it? It's not just the imperfect integrity of trustees that can punish delay of gratification; it's the imperfect knowledge of experts. "Everything they said was bad for you is good for you," we joke, and with today's better nutrition science we know that a lot of pleasure from eggs, shrimp, and nuts was forgone in past decades for no good reason.

Second, in the long run we're all dead. You could be struck by lightning tomorrow, in which case all the pleasure you deferred to next week, next year, or next decade will have gone to waste. As the bumper sticker advises, "Life is short. Eat dessert first."

Third, you're only young once. It may cost more overall to take out a mortgage in your thirties than to save up and pay cash for a house in your eighties, but with the mortgage you get to live in it all those years. And the years are not just more numerous but different. As my doctor once said to me after a hearing test, "The great tragedy in life is that when you're old enough to afford really good audio equipment, you can't hear the difference." This cartoon makes a similar point:



"See, the problem with doing things to prolong your life is that all the extra years come at the end, when you're old."

These arguments are combined in a story. A man is sentenced to be hanged for offending the sultan, and offers a deal to the court: if they give him a year, he will teach the sultan's horse to sing, earning his freedom. When he returns to the dock, a fellow prisoner says, "Are you crazy? You're only postponing the inevitable. In a

The public choices we face today, like how high a tax we should pay on carbon to mitigate climate change, depend on the rate at which we discount the future, sometimes called the social discounting rate.[19] A rate of 0.1 percent, which reflects only the chance we'll go extinct, means that we value future generations almost as much as ourselves and calls for investing the lion's share of our current income to boost the well-being of our descendants. A rate of 3 percent, which assumes growing knowledge and prosperity, calls for deferring most of the sacrifice to generations that can better afford it. There is no "correct" rate, since it also depends on the moral choice of how we weight the welfare of living people against unborn ones.[20] But our awareness that politicians respond to election cycles rather than the long term, and our sad experience of finding ourselves unprepared for foreseeable disasters like hurricanes and pandemics, suggest that our social discounting rate is irrationally high.[21] We leave problems to future Homer, and don't envy that guy.

There's a second way in which we irrationally cheat our future selves, called myopic discounting.[22] Often we're perfectly capable of delaying gratification from a future self to an even more future self. When a conference organizer sends out a menu for the keynote dinner in advance, it's easy to tick the boxes for the steamed vegetables and fruit rather than the lasagna and cheesecake. The small pleasure of a rich dinner in 100 days versus the large pleasure of a slim body in 101 days? No contest! But if the waiter were to tempt us with the same choice then and there—the small pleasure of a rich dinner in fifteen minutes versus the large pleasure of a slim body tomorrow—we flip our preference and succumb to the lasagna.

The preference reversal is called myopic, or nearsighted, because we see an attractive temptation that is near to us in time all too clearly, while the faraway choices are emotionally blurred and (a bit contrary to the ophthalmological metaphor) we judge them more objectively. The rational process of exponential discounting, even if the discounting rate is unreasonably steep, cannot explain the flip, because if a small imminent reward is more enticing than a large later one, it will still be more enticing when both rewards are pushed into the future. (If lasagna is more enticing than steamed vegetables now, the prospect of lasagna several months from now should be more enticing than the prospect of vegetables several months from now.) Social scientists say that a preference reversal shows that the discounting is *hyperbolic*—not in the sense of being exaggerated, but of falling along a curve called a hyperbola, which is more L-shaped than an exponential drop: it begins with a steep plunge and then levels off. Two exponential curves at different heights never cross (more tempting now, more tempting always); two hyperbolic curves can. The graphs on the next page show the difference. (Note that they plot absolute time as it is marked on a clock or calendar, not time relative to now, so the self who is experiencing things right now is gliding along the horizontal axis, and the discounting is shown in the curves from right to left.)

Admittedly, explaining weakness of the will as a reward gets closer by hyperbolic discounting is like explaining the effect of Ambien by its dormitive power. But the elbow shape of a hyperbola suggests that it may really be a composite of two curves, one plotting the irresistible pull of a treat that you can't get out of your head (the bakery smell, the come-hither look, the glitter in the showroom), the other plotting a cooler assessment of costs and benefits in a hypothetical future. Studies that tempt volunteers in a scanner with adult versions of marshmallow tests confirm that different brain patterns are activated by thoughts of imminent and distant goodies.[23]

Though hyperbolic discounting is not rational in the way that calibrated exponential discounting can be (since it does not capture the ever-compounding uncertainty of the future), it does provide an opening for the rational self to outsmart the impetuous self. The opening may be seen in the leftmost segment of the hyperbolas, the time when both rewards lie far off in the future, during which the large reward is subjectively more appealing than the small one (as rationally it should be). Our calmer selves, well aware of what will happen as the clock ticks down, can chop off the right half of the graph, never allowing the switchover to temptation to arrive. The trick was explained by Circe to Odysseus:[24]

PENGUIN BOOKS

UK | USA | Canada | Ireland | Australia
New Zealand | India | South Africa

Penguin Books is part of the Penguin Random House group of companies whose addresses can be found at global.penguinrandomhouse.com.

Penguin
Random House
UK

First published in the United States of America by Viking, an imprint of Penguin Random House LLC 2021
First published in Great Britain by Allen Lane 2021

Grateful acknowledgment is made for permission to reprint the following:
Excerpt on page 55 from *The Odyssey* by Homer, translated by Emily Wilson.
Copyright © 2018 by Emily Wilson. Used by permission of W. W. Norton & Company, Inc.

Charts rendered by Ilavenil Subbiah
Cover design: Pete Garceau

ISBN: 978-0-241-38030-7