Reason and Less

-1___
0___
+1___

Reason and Less

Pursuing Food, Sex, and Politics

Vinod Goel

The MIT Press
Cambridge, Massachusetts
London, England

___-1
___0
___+1

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Goel, Vinod, author.
Title: Reason and less : pursuing food, sex, and politics / Vinod Goel.
Description: Cambridge, Massachusetts : The MIT Press, [2022] |
Includes bibliographical references and index.
Identifiers: LCCN 2021017752 | ISBN 9780262045476 (paperback)
Subjects: LCSH: Decision making. | Reasoning. | Logic. | Cognitive neuroscience.
Classification: LCC BF448 .G64 2022 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2021017752

10 9 8 7 6 5 4 3 2 1

To the memory of my mother and father

For Kalpna, Amit, and Natasha,
who taught me there is often less
to life than reason

Contents

___-1
___0
___+1

___-1
___0
___+1

Preface

After more than 20 years of studying the neural basis of rationality, it dawned on me that there was very little consequential human behavior that I could explain. Nothing I have learned about rationality was relevant to understanding my teenage daughter. Nothing I have learned about rationality is relevant to explaining the behavior of my MAGA (Make America Great Again) Florida friends and neighbors who profess an unshakable faith in American exceptionalism (which I accept and have benefited from) but then deny and ridicule the sciences of vaccines and climate change emerging from exceptional American institutions. Nothing I have learned about rationality seems particularly relevant to explaining certain views of my ultraliberal friends and colleagues, such as gender being just a social construct, despite scientific evidence to the contrary. Nothing I have learned about rationality is relevant to explaining why intelligent, powerful men engage in sexual indiscretion, even assault, at great personal risk and harm to others. Nothing I have learned seems particularly relevant to explaining why I overindulge in chocolate cake and pizza, despite being overweight. Based on the standard models of reasoning, the only explanatory tools available are appeals to "heuristics," some form of "motivated reasoning," poor education, or perhaps cognitive deficiency. Such explanations may apply in specific individual cases, but they cannot account for all or even much of human behavior. I have come to believe that we are making a fundamental mistake in bringing only the tools of rationality to explain human behavior.

My main message is that, while we *are* rational animals, explaining real-world human behavior just in terms of reasoning does not get us very far. We have to recognize that nonreasoning systems also affect actual behavior. We need to look beyond (or below) reason to *noncognitive* factors to fully account for human behavior. Much human behavior that does not conform to our expectations of rationality is not irrational but rather *arational*,

by which I mean that it is not reason based. Some nonreasoning systems are initiating and/or modulating the behavior.

The goal of this book is to undertake a commonsense reconsideration and recalibration of theories of human behavior. Human behavior needs to be explained in terms of the workings of autonomic systems, instinctive systems, associative systems, and reasoning systems. Each of these systems has been extensively studied. How these systems communicate and interact to account for human behavior is rarely considered. I sketch out a proposal that I call *tethered rationality*, in which human behavior is a *blended response* incorporating inputs from each of these systems. The challenges are to provide empirical data for the blended response hypothesis, show how the tethering is supported by the neurophysiology, propose a common currency that would allow these systems to communicate and interact, and provide a control structure for the overall system. Meeting these challenges takes us on a fascinating journey through psychology (cognitive, behavioral, developmental, and evolutionary), neuroscience, philosophy, ethology, economics, and political science, among other disciplines.

One key insight that holds the model together is that *feelings*—generated in old, widely conserved brain stem structures—are evolution's solution to initiating and selecting all behaviors and provide the common currency for the four different systems to interact. Reason is as much about feelings as is lust and the taste of chocolate cake. All systems contribute to behavior and the overall control structure is one that maximizes pleasure and minimizes displeasure. Such an account drives human behavior back into the biology, where it belongs, and provides a richer set of tools to understand how we pursue food, sex, and politics.

Models not only explain behavior but also have consequences for changing it. The model of tethered rationality is no exception. For those engaged in changing behaviors—such as sexism, racism, cheating, or even climate change denial—tethered rationality may have the unwelcome message that such behaviors cannot be easily changed by changing beliefs through a few days of "sensitivity training." This is not to say that they cannot be changed at all, but rather that more drastic measures will be required, the nature of which will depend on the specific behavior in question. Having an accurate model of human behavior is the first step in this endeavor.

Utopia, Ontario, Canada

May 2021

Acknowledgments

I would like to thank the following colleagues from diverse areas of expertise for reading one or more chapters of the manuscript and offering valuable feedback: Ellen Bialystok, Christopher von Bülow, Ron Chrisley, Hugo Critchley, Wim De Neys, Shira Elqayam, Larry Fiddick, Jordan Grafman, Sam Gilbert, Ira Novak, Magda Osman, Jerome Prado, Steven Sloman, and Hongkui Zeng. Among friends and students, I'm grateful to Mark Hewitt, Ron Bean, Adam Burnett, and Claire Quenneville, for reading and commenting on multiple chapters. I thank Sophie Goss and Jenna Zorik for reading all chapters multiple times and forcing me to articulate the story more clearly. Finally, I am grateful to my American MAGA friends and neighbors for many months of discussions that provided insight into their underlying thought processes and confirmed what I had learned from my teenage daughter: there is indeed often less to life than reason.

-1___
0___
+1___

I The Rational Animal

> Man is the only animal capable of reasoning, though many others possess the faculty of memory and instruction in common with him.
>
> —Aristotle

> There's a logical explanation for everything, often mistaken for the reason it happened.
>
> —Robert Breault

To ask questions about the role of reason in human affairs is, in the broadest sense, to ask questions about our place in the universe. What is the nature of man? Who and what are we? We have struggled with such questions for as long as we have been able to think about such things. Are we reasoning animals? Are we only reasoning animals? Is reason necessary? Is it sufficient? What ever happened to the "animal passions"? Have socialization and culture—constructions of the reasoning mind—allowed us to rise above them (like Katharine Hepburn's character in the film *The African Queen* advocated [Huston, 1951]: "Nature, Mr. Allnutt, is what we were put on this world to rise above"), or do we need an account of human nature that reconciles the two? The reader will guess from the title of the volume that I make the case for the latter.

-1___
0___
+1___

1 Food, Sex, Politics, and the Rational Animal

> To proceed on this track, investigators would need to accept one grand but empirically robust premise—that higher aspects of the human mind are still strongly linked to the basic neuropsychological processes of "lower" animal minds.
>
> —Jaak Panksepp

Much of life is about pursuing food, sex, and politics. Any adequate theory of human behavior must be able to explain these pursuits.

By far the most popular academic accounts of human behavior place the *rational* mind front and center (Cassirer, 1944; Durkheim, [1895] 2014; Simon, 1955). Humans bring the tools of reason to bear on these problems. Reason sets us apart from other animals. It allows us to successfully pursue not only food, sex, and politics but also art, science, and technology. This model is often referred to as the standard cognitive or social science reasoning model of human behavior (Tooby & Cosmides, 1995). After more than 20 years of trying to understand human decisions and choices just through the lens of reason, I have become skeptical of the explanatory scope of this standard model.

I'm convinced that reason is an integral part of who and what we are. I'm also convinced that, on its own, it is inadequate to explain much, if not most, real-world human behavior. It is only half the story. We do not have to look very far to understand what is missing. There is a commonsense model of behavior, embedded in the Western-Christian intellectual tradition, that recognizes not only reason but also "animal passions" (often characterized as the four Fs: feeding, fornicating, fighting, and fleeing) as determinants of human behavior. Our choices and decisions are a function of both. Not only is this much more intuitive, but we will see that the data demand such a model.

Despite common sense and data, such a model no longer gets serious consideration in large segments of modern society, including much of

academia. I worry that the main reason is that many people, some academics included, hold variations on the meritless belief that "humans no longer need to rely on instinct to survive, not when we have education, technology, and social norms" (Pomeroy, 2011). The goal of this book is to push back against this widespread misconception, and articulate a commonsense model of human nature, called *tethered rationality*, that preserves the basic intuitive insight of the Western-Christian model—that both reasoning and nonreasoning systems are in play in human behavior—and can be discharged without divine intervention.

The "animal passions," or nonreasoning behaviors in technical parlance, include autonomic behaviors, instinctive behaviors, and associative learning behaviors. These behaviors and their underlying mechanisms have been studied extensively over the past hundred years. They differ not only from reasoned behaviors but also from each other. They are hierarchically organized in terms of appearance on the evolutionary tree, are integrated, and are widely available across species, including humans. Humans also exhibit reasoning or rational behavior, which (I will argue) is unique to us. However, it does not supplant the evolutionarily older behaviors. Reason evolved on top of them, but it does not "float" untethered above them; it is tightly integrated with both bottom-up and top-down connections. This means that human behavior is a blended function of all these systems, not just reason (or any other individual system). Humans have a reasoning mind, but it is tethered to and modulated by evolutionarily older associative, instinctive, and autonomic minds.

I begin this chapter by introducing five examples of real-world decisions that are widely thought to be explained by reason. Before we can consider whether these examples are actually explained by models of reasoning, we need to introduce the notion of reason and rationality. This is initially done informally. With this preliminary understanding of reasoning in hand, I then evaluate each example to see if it can be explained just in terms of reason. I conclude that four of the five examples cannot be so explained. Satisfactory explanations for these require the introduction of evolutionarily older nonreasoning systems. A roadmap is then provided to foreshadow the argument for the model of tethered rationality and guide the reader through the subsequent chapters.

Examples of Reasoning in the Real World

-1___
0___
+1___

Let's begin by considering five real-world examples of reasoning and decision-making scenarios.

The first example is climate change, the ultimate existential issue of our time. The best science we have agrees that human activity is contributing to rising temperatures, which will reshape planetary weather patterns and geography and have detrimental, even catastrophic, effects on all life on earth. The scientific models could be wrong by either overestimating or underestimating the changes that will occur, but they provide the best information we currently have. Most governments and citizens accept the science and are willing to take some (limited) steps to mitigate the impact of human activity. However, the forty-fifth president of the United States, a number of US senators, and 40% of the American public believe that "man-made global warming is the greatest hoax ever perpetrated on the American people" (Revkin, 2003). They claim, without evidence, that the scientific models are incorrect. Even among the other half of Americans who do accept the science, there is considerable reluctance to undertake full remedial measures. This example illustrates two separate issues: that many people simply deny the science, without evidence to the contrary, and others seem to accept the science but fail to act on it. There seems to be a lack of rationality in both cases.

The second example involves weight management. Last year, I went to my doctor's office for my annual checkup. After I stepped on the scale, my doctor advised me to lose 30 pounds. I agreed but complained that my busy schedule did not allow time to eat healthy meals and exercise regularly. My doctor replied, "What fits your busy schedule better, eating healthy and exercising one hour a day or being dead 24 hours a day?" Many of us have been in this situation, but few of us actually manage to follow our doctor's advice. Notice that we do not question the doctor's judgment. There seems to be considerable evidence linking obesity with the onset of various diseases (e.g., diabetes and heart disease) and premature mortality. Most of us do not have a death wish. Given that we want to live a long, healthy life, and given that we accept that obesity will impair and even shorten our lives, the rational, reasonable thing to do would be to lose weight. So, why don't many of us comply with our doctor's advice?

For our third example we turn to sex. In December 2006, John Edwards, a handsome, charismatic lawyer and politician, announced his candidacy for the 2008 Democratic nomination for president of the United States. He was among the frontrunners, along with Barack Obama and Hillary Clinton, for the nomination. In March 2007, it was revealed that his wife, Elizabeth, was suffering from stage IV breast cancer. Shortly thereafter, it came to light that he was having an affair with one of his campaign workers. In what world was this a rational choice? He was running for the highest office in

the world, in a country that contains some of the most socially conservative, prudish, judgmental, evangelical voters. He must have known that if there was any hint of infidelity—even in the best of circumstances—his campaign was over. His circumstances were such that his wife was dying of cancer and receiving enormous emotional and moral support from the public. Any hint of infidelity in such circumstances would be suicidal. Evidence of the affair emerged in early 2008 and ended his candidacy overnight. How do we explain his choices?

The fourth example concerns healthcare, a topic that often comes up in discussions with my American friends. The conversations often take the following form:

Me: Given your very high premiums and the large deductible in your private healthcare plan, why don't you support overhauling your healthcare system into a universal Canadian/European-type system whereby everyone can receive good equivalent healthcare at a lesser cost?

My American friend: Affordable healthcare would certainly be a great benefit to me. However, you see that guy over there? Yes, that one. He doesn't work. He doesn't pay taxes. He is a freeloader. If we had universal healthcare, he would get the same healthcare that I do, but he doesn't *deserve* it. Therefore, I cannot support a universal system. (Another interesting response is the admission that, "yes, that would probably be better than what we have," followed by passing shame and a disappointed sigh, "but that would be socialism.")

My friend is willing to forgo a benefit for himself just so that someone "undeserving" does not receive an equivalent benefit. Again, it is hard to see the rationality in this choice.

For the fifth example we turn to a drug warning issued to doctors and patients by the UK Committee on Safety of Medicines in 1995. The warning stated that the third generation of birth control pills doubled (i.e., increased by 100%) the risk of life-threatening blood clots in the legs and/or lungs. Unsurprisingly, this caused great anxiety among women and resulted in a sharp increase in unwanted pregnancies and abortions in subsequent years. A closer examination of the study showed that for every 7,000 women who took the second-generation pill, one developed thrombosis. By contrast, for every 7,000 women who took the third-generation pill, two developed thrombosis. So, while the relative risk did increase by 100% as advertised, the absolute risk was an increase of 1 in 7,000 women (Gigerenzer, 2015). This hardly seems to warrant the panic that ensued, so how can we explain it?

These are five (very different) examples of everyday, real-world decisions or choices. Other examples will be introduced throughout the book. Even though I have not yet formally introduced the idea of "rationality," I'm confident most readers will agree that each example illustrates a choice that seems less than fully rational. I will not go so far as to say that they are *irrational*. In the cases of examples one through four, I will argue that they are *arational*—that is, they involve noncognitive factors.

The most popular academic models that we have for explaining these behaviors are the cognitive reasoning and decision-making models, buttressed by distinctions between analytic and "heuristic" reasoning, such as the "fast and slow" thinking model popularized by Daniel Kahneman (2012), or by notions of motivated reasoning (Kunda, 1990) or even sloppy reasoning (Pennycook & Rand, 2019). Such models will be introduced and considered in chapters 7 and 13. They provide satisfactory explanations for a number of phenomena, including example five, but lack the requisite machinery to deal convincingly with examples one through four, which are the ones of interest in this book.

To make sure we are all on the same page, I offer an initial introduction to the notion of rationality and decision-making and then return to address the preceding examples.

What Is Rationality?

Man is widely considered to be the "reasoning" or rational animal. But what does this mean? To invoke reason or rationality is to say that human behaviors or actions are explained by postulating beliefs and desires and a principle of *coherence* that guides our pursuit of the latter in the context of the former. By coherence I mean roughly "making sense." Coherence is a relationship that holds between thoughts, propositions, or sentences. In the first instance, it is a basic, primitive, intuitive notion, though it can be considerably enhanced with education. For example, if I believe that all Americans are intelligent, and all Fox News viewers are American, then it would be coherent or reasonable for me to infer that all Fox News viewers are intelligent. Given the same beliefs, it would not be coherent to infer that no Fox News viewers are intelligent. This example illustrates a particularly extreme case of coherence found in deductive arguments, referred to as *validity*, where the truth of the given information (or beliefs) is sufficient to guarantee the truth of the conclusion, but it is worth noting that validity does not evaluate the veracity of the premises that all Americans are

intelligent and all Fox News viewers are Americans; it merely determines whether a conclusion follows from or is entailed by them. We can consider validity as coherence in the narrow sense of the term and additionally have a broader sense of the term, corresponding to *soundness* in logic, that also takes into consideration the veracity of the premises. In this broader use of the term, we would step back and evaluate (and either accept or reject) the truth of the premises before drawing the inference.

On a recent trip to New Delhi, India, one afternoon I observed Indian fruit bats dangling from tree branches like so many brown and black cloth sacks. Based on this observation, I formulated the belief that Indian fruit bats spend the afternoon dangling from tree branches. This is a plausible or coherent inference based on my observations, but notice that it lacks the certainty of the preceding inference about Fox News viewers. Further observations (or consultation with bat experts) might reveal that this behavior is a peculiar habit of fruit bats in this particular region of India. In this case, I would have to modify my belief for it to be consistent with the facts in the world. Absent additional information, it is coherent for me to believe that Indian fruit bats spend afternoons dangling from tree branches. Given the same evidence, it would be incoherent for me to conclude that Indian fruit bats do not spend the afternoon dangling from tree branches or spend the afternoons diving for crayfish in shallow rivers.

Coherence relations between premise and conclusion are disrupted by inconsistency, indeterminacy, or irrelevance. Inconsistency is illustrated where the conclusion "No Fox News viewers are intelligent" is drawn from the beliefs that "All Americans are intelligent" and "All Fox News viewers are Americans." An example of indeterminacy occurs if I tell you "Mary is taller than George, and Mary is taller than Michael" and ask you the height relationship between George and Michael. The premises do not provide sufficient information to draw any inferences about the relative heights of George and Michael. An example of failure of coherence through irrelevance would occur if, given the belief that all Americans are intelligent and the belief that Indian fruit bats spend afternoons dangling from tree branches, I conclude that global warming is caused by human activity. In this case, the issue of coherency does not even arise, because the three propositions are unrelated.

From Rationality to Decision Theory

-1___
0___
+1___

Reasoning is about maintaining coherence in belief networks. Life is about actions. Reason mediates action by determining choices consistent with

specific goals, given specific beliefs. Choice selection is studied by decision theory. We get from reasoning to decision-making by overlaying some model of human goals on top of the model of rationality. These models are usually based on maximizing self-interest. A historically popular one is the *Homo economicus* model. In this account, man is intrinsically a self-interested utility maximizer as a consumer and a self-interested profit maximizer as a producer.[1] These become the goals of the individual. Rational actions are those that are expected to advance goals in light of beliefs.

I will illustrate this standard model of decision-making with the controversial US decision to invade Iraq in March 2003. While I have no privileged access to the particulars of the decision-making process, its overall *form* would be something like that depicted in figure 1.1. It would begin with a goal or desire that needs to be achieved, such as securing the Iraqi oil leases. This goal would be explored or expanded via subgoals. One subgoal option might be negotiation. Another might be to take the oil by force if certain conditions can be met, such as: assurance of success, clean surgical intervention and withdrawal, that the value of the oil leases be greater than the cost of the invasion, and that Iraq be able to pay for its own reconstruction costs. In this example, these conditions are believed to be met

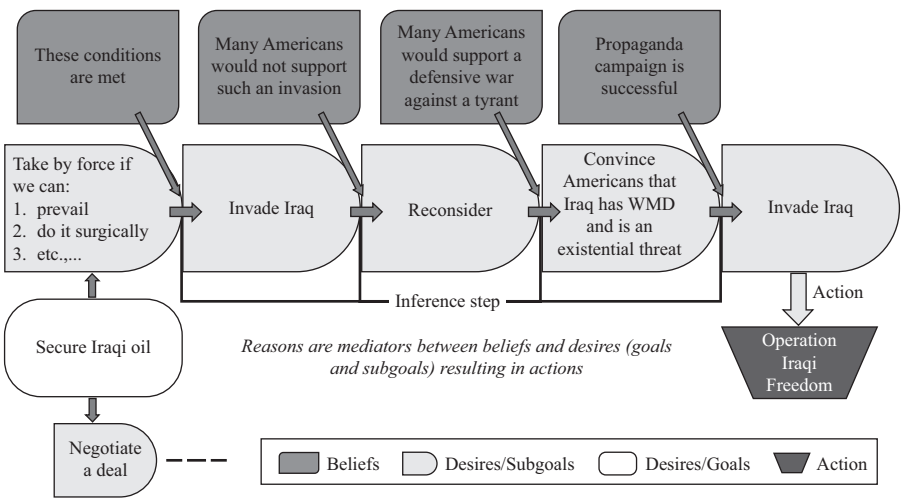

Figure 1.1
An example of the rational mind at work using a hypothetical reconstruction of the US decision to invade Iraq in 2003. Each subgoal follows coherently from the preceding goal or subgoal plus beliefs, eventually resulting in an action. The integration of goals and subgoals plus beliefs via the coherence relation is the nexus of the reasoning step.

___-1
___0
___+1

(and negotiation is not considered feasible or cost-effective), leading to the subgoal of invading Iraq. However, there are accompanying beliefs that suggest most Americans (and the world community) will not support an unprovoked invasion, even if it means access to cheap oil. This results in another subgoal to pause and reconsider. There are accompanying beliefs that most Americans (and the world community) would support a defensive war against a tyrant. This generates another subgoal of launching a campaign to vilify Saddam Hussein and convince Americans that Iraq has weapons of mass destruction that are an imminent threat to the United States (which has more weapons of mass destruction than all other countries combined) and its allies. It is determined that the propaganda campaign is successful and there is sufficient support within the country for the invasion. Given all this, the rational decision is to invade Iraq; each step follows coherently from the previous goal or subgoal plus beliefs.

However, this model is an oversimplification. It assumes that the beliefs or information at hand are complete and certain. But how certain are we that Iraq can repay its own reconstruction costs? 100%? 10%? 73%? Are there any constraints on the desire to take the oil by force? If the financial cost of the war equals or exceeds the benefits of the oil, do we still want to pursue this desire? In real-world situations, information is always incomplete and uncertain, and even the relative utility of different desires cannot be confidently ascertained and ordered. These complications transform the problem of inferential coherence from the realm of logic to the realm of probability theory (see figure 1.2). Coherency is then determined by applying the probability calculus to the model. The rational choice is the one with the highest utility value. One consequence of this shift is that the criterion of coherence morphs to an optimality criterion. However, for our current purposes, these complications are not material. It is still coherent to select the option with the highest expected utility (see figure 1.2). I have chosen to use the concept of coherence rather than utility as central to rational decision-making throughout the book.

This example is offered as a simplified illustration of the machinery of standard decision-making models. There are two points worth noting. First, the postinvasion justification (when no weapons of mass destruction were found)—that American lives and resources were expended so the Iraqi people could benefit from regime change and democracy—is irrational because it violates the basic tenets of maximizing self-interest. Second, I'm not claiming that this rational model is sufficient to explain the invasion of Iraq. On the contrary, I'm certain that a number of nonrational factors considered in this book were significant factors in making the decision.

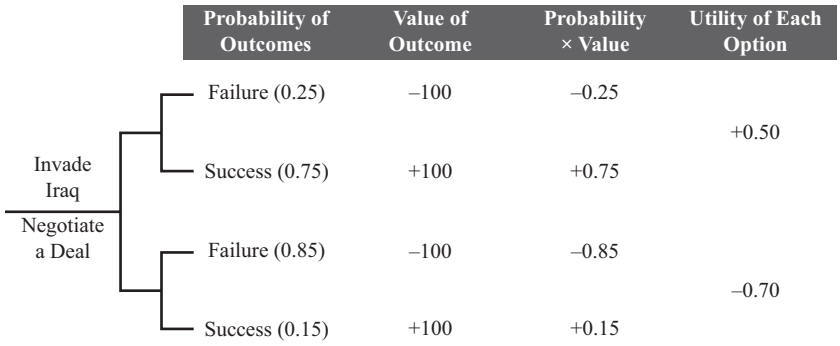

Figure 1.2
Simple decision tree and utility function. One might model the decision to invade or make a deal as follows. The chances of a successful invasion are 0.75, while the chances of failure are 0.25. The chances of a successful negotiation are 0.15, while the chances of failure are 0.85. The value assigned to both the successful invasion and successful negotiation is +100. The value assigned to a failed invasion and failed negotiation is –100. Based on these values, the utility of invasion is +0.5 and the utility of negotiation is –0.7 (utility = $\Sigma(\text{probability}_{\text{outcome}} \times \text{value}_{\text{outcome}})$). Notice that decision theory provides no guidelines for assigning probabilities of outcomes and the value of the outcomes, but once these numbers are (magically) assigned, simple probability theory allows us to coherently calculate expected utility. The rational choice is the one with the highest expected utility.

More generally, I'm claiming that such standard models of rationality cannot adequately account for much of human behavior, including the invasion of Iraq and four of the five examples introduced earlier. Understanding this claim requires reviewing each example more closely, beginning with global warming.

Rationality in the Real World: Global Warming Example

The basic questions around climate change are "Is the earth warming?" and "Is human industrial activity contributing to it?" Most scientists answer "yes" to both questions (The National Academy of Sciences & The Royal Society, 2020). Many members of the public agree, but at least 40% of Americans vehemently disagree. The same data are available to all. We are all rational, so why the discrepancy in opinion? Let us consider the argument and the various sources of dissent to see rationality working, failing, and being irrelevant.

The argument climate scientists make for man-made climate change is summarized as follows by the National Aeronautics and Space Administration

___-1
___0
___+1

(2020): data indicate global temperatures have been steadily rising since the 1800s (the start of the Industrial Revolution), resulting in melting of the polar ice caps and rising sea levels. There can be many natural sources for temperature increases, such as variation in solar activity, volcanic activity, and even slight shifts in Earth's trajectory around the sun, and these have indeed resulted in past climatic changes. But the timescale and "fingerprint" of the changes we are currently experiencing are not consistent with any of these natural causes. Examination of ice cores from Antarctica reveals that carbon dioxide levels have been relatively stable throughout the past 800,000 years but have shot up dramatically over the past hundred years. When we incorporate the data about excess introduction of carbon dioxide into the atmosphere as a result of human fossil fuel activity and disruption of the natural carbon-oxygen cycles, the projected greenhouse effect is very similar to what we are actually experiencing. Therefore, it is reasonable to believe that human activity (such as carbon dioxide emissions) is a large causal factor in global warming.

This conclusion is plausible, perhaps even compelling, but not certain. One can probe, question, and doubt. Let's examine some possible "reasons" for rejecting the argument offered by nonbelievers by reviewing a question-and-answer session on climate change, held in June 2010 at the University of New South Wales, called "The Sceptics" (2010). It was moderated by Jenny Brockie and featured climate scientist Professor Stephen Schneider from Stanford University and some ardent skeptics from the Australian general public. The first skeptic questioned by the moderator was Tania.

Moderator: Tania, do you believe in man-made climate change?

Tania: Man-made? Not at all.

Moderator: Why?

Tania: No one has proven to me that it's man-made at all. What I say is it's a big hysteria just for money. The only reason you're getting grant money is because of climate change. The planet is warming is the only reason you're getting grant money. If we didn't have this hysteria there would be no grants. There would be no people making money at all.

In this case, the argument for climate change is not actually in play. Tania's objection does not consider the relation between the evidence and the conclusion. Tania is attributing a disingenuous or malicious motive to climate scientists and is offering an ad hominem response. Scientists are simply lying to pad their pockets with grant money. This objection is a case of disagreeing with a conclusion but for reasons that have nothing to do with the coherence of the argument. Many real-world disagreements fall

into this category. A similar technique can be used to endorse arguments that are offered by friends and people that one admires. The argument itself does not matter. Coherence relations between evidence and conclusion are not in play. Therefore, such objections (or endorsements) do not belong to the realm of the rational. One might think that educating Tania about the individualistic and competitive nature of science and scientific grant funding may dissuade her from her misconception, but as we will see in chapter 13, it probably will not.

More valid reasons to reject the climate change argument would be to question the data and/or measurement techniques. Another skeptic, John, voiced the concern that he had read that 89% of the thermometers were placed too close to artificial heat sources, such as buildings, and this was artificially inflating temperatures. The accuracy of the methods of calculating temperature changes from thousands of years ago using tree rings data was also questioned. If these concerns are correct, whatever coherency the initial argument had would need to be reevaluated. Schneider acknowledged the challenges of accurate historical measurements, corrected John's belief about 89% of thermometers being placed near heat sources, and explained some of the techniques scientists use to ensure accuracy of the data (e.g., pulling out from the record those temperature readings affected by urban heat sources and covarying population growth with temperature increase).

Case, another skeptic, had just returned from a trip to Alaska and raised two issues regarding glacial melt. On his trip, he had learned that in 1750 Glacier Bay was completely occupied by a glacier. By 1860, half of it had melted. This melting occurred prior to any significant human industrial activity, so how can we assume that the melting of glaciers is proof of global warming? Furthermore, the Alaskan glaciers that originate at low altitudes are indeed receding, but those that originate at higher altitudes are actually advancing. Both these observations seem inconsistent with the global warming models. If so, the models need to be revisited to make them cohere with the data.

Professor Schneider replied that it is not correct to say that human activity was inconsequential prior to the 1800s. We have been involved in large-scale agricultural land clearing for thousands of years, and this has had an impact on CO_2 accumulation, albeit on a much smaller scale than present industrial activity. Schneider explained that if we average across all glaciers around the world and the rates of melt, the data show accelerated rates of melting in the twentieth century relative to prior centuries. With respect to glaciers actually building up and advancing at higher altitudes, that is exactly what the models predict. If you begin with a very cold temperature, say –10°C, and you warm it up by 5°C, to –5°C, the warmer atmosphere

___-1
___0
___+1

will hold more moisture, resulting in increased snowfall and ice buildup, until it warms up past 0°C and starts to melt. This has been observed not only in Alaska but also in Antarctica and Greenland. All this is consistent with expectations of the theory. So, the apparent contradiction was based on some incorrect information in the belief network of the dissenter. Once this misinformation is corrected, the inconsistency should disappear and coherence emerge.

The final skeptic we will consider is Ian. He raised the following objection: "I understand that carbon dioxide that man produces is 3% of what nature produces. How can small changes to our production of CO_2 impact upon something as large as the Earth? It seems absurd." Schneider responded to Ian by briefly explaining the annual carbon cycle, whereby carbon dioxide is taken up from the atmosphere by vegetation during photosynthesis in the spring and summer growth seasons and released back into the atmosphere in the autumn and winter when the leaves fall and decompose. The amount of carbon involved here is much greater than that generated by human activity, but critically, the cycle is in balance. Burning of fossil fuels by humans disrupts the balance of the cycle by adding CO_2.

Ian: Sorry to butt in on this. Look, you're not answering the question. I said that we produce approximately 3% of natural production. You haven't really addressed that. You've given us some prevaricative answer.

Prof. Schneider: I mean perhaps you do not understand the answer. What I said is the amount of carbon dioxide coming from the atmosphere goes in and out and it's larger than what we inject. But it's in balance.

Ian: It's 3% carbon dioxide of the total production of carbon dioxide. It's still a small percentage. If we reduce our carbon dioxide by 50% and send ourselves back to the Stone Age we've made very little difference. Could you answer that question? I did understand what you said perfectly.

Prof. Schneider: Let me give you an example. If you have a bathtub, you can turn it on so you are getting a gallon coming in a minute, right? Now the drain is opened up to the point where a gallon is going out in a minute. So, there's a flow in and there's a flow out. That's an analogy to the fact that there is a very large flow of carbon dioxide naturally going into the system in the summertime and coming out in the winter. Much larger than the 3%, I agree with that. However, it's in balance. The amounts are the same, so when you add the 3%, it's 3% this year and next year and next year. . . . And it accumulates. So, if all of a sudden, I go to the bathtub and I make the one gallon into 1.2 gallons and I don't change the drain size in the bottom, the water in the bathtub is going to rise [and overflow].

-1___
0___
+1___

In this particular case, the dissenter does not assail the motives of climate change scientists or question the measurement techniques, nor is he giving indications of harboring erroneous beliefs, but he nonetheless simply refuses to accept the coherency of Schneider's argument. Most of us can readily see the rationale in the bathtub analogy. Even if the human contribution of CO_2 is a small fraction of the naturally occurring amount, as long as it is in addition to the natural input/output cycle, such that the input becomes greater than the output, we can readily understand that an overflow will eventually occur (figure 1.3). But this skeptic simply fails to understand or acknowledge the coherency of the argument. If this is a genuine failure of coherence (rather than a contrived stance), it is not clear what more can be said to convince the dissenter. Simple coherence relations are primitive intuitive notions. Either you "see it" or you don't. Everyone with normal cognitive capacity should be able to "see" that the bathtub will overflow.

These exchanges between Schneider and the skeptics illustrate various sources of disagreement in real-world arguments, including assigning

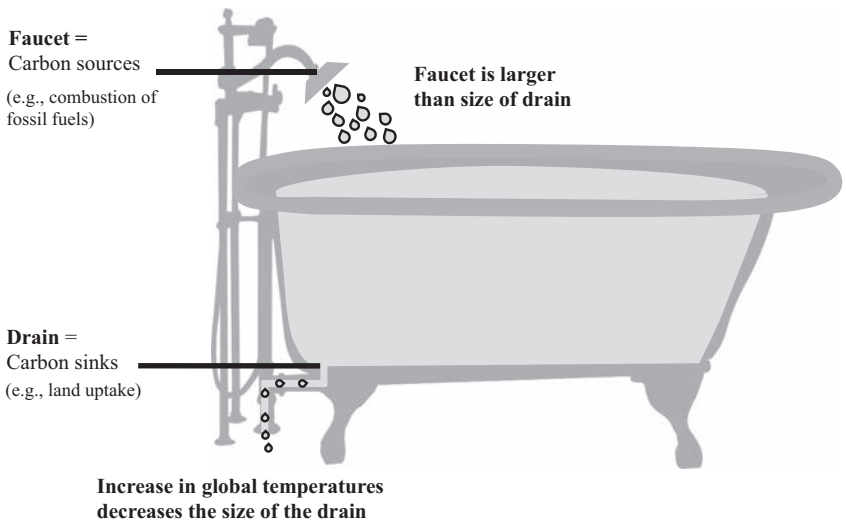

Figure 1.3
The carbon bathtub analogy. If more water is dripping into the bathtub than is leaving via the drain, no matter how small the difference, the coherent conclusion is that the bathtub will eventually overflow. This is an example of a basic, intuitive coherency judgment. If one fails to acknowledge it (in good faith), it is not clear what more can be said to change one's mind. This would constitute a cognitive failure in detecting coherency.

___-1
___0
___+1

disingenuous motives to the individual putting forward the argument and therefore simply not believing it; questioning measurement techniques and the accuracy of data; having false beliefs about data or misunderstanding parts of the argument; and failure of coherency judgments. Professor Schneider did provide evidence and reason-based answers regarding measurement techniques and corrected false beliefs among participants. But at the end of this exchange, only one individual changed their mind from "sitting on the fence" to accepting the reality of climate change. *The other 20 or so skeptics were equally as skeptical at the end of the session as at the beginning*. The reasoned responses provided by Schneider had no impact on their beliefs. The skeptics did not question the evidence and arguments he presented. They did not offer corrections or additional evidence to the contrary. They simply refused to change their beliefs. This is not rational. How can this be explained?

The most ubiquitous explanation for failures of reasoning is the prominence of "heuristics" over analytical reasoning (Evans & Over, 1996; Kahneman, 2003; Sloman, 1996). We will see in chapter 7 that heuristics come in several different flavors and can play useful roles in theories of reasoning, but they are not particularly relevant to explaining the types of examples under consideration here. They are part of the machinery of reasoning and are sensitive to coherence relations. The heuristic explanation is often combined with the "sloppy reasoning" and "motivated reasoning" explanations.

The "sloppy reasoning" explanation is exactly what it sounds like (Pennycook & Rand, 2019). While coherence itself is a basic, intuitive notion, determining coherence between data and theory need not be a trivial matter. We often need to call on the formal apparatus of logic and mathematics to guide coherence determinations in complex cases, highlighting the value of education, training, and effort in honing and developing basic, intuitive coherency judgments. In the sloppy reasoning account, one would say that the audience did not have sufficient education and training to understand the argument. As we will see in later chapters, when we take up this issue in earnest, this may be true in individual cases but cannot explain the overall phenomenon.

A third explanation is that the skeptics have a vested interest in some status quo and were engaging in motivated reasoning (Kunda, 1990). Whereas ideal reasoning involves going from data to conclusions in a disinterested manner, motivated reasoning is guided by a preexisting goal or desire (e.g., continuing to burn fossil fuels) that serves to filter the data in order to support the preferred conclusion (man-made climate change is a hoax). This

explanation also falls short. In fact, motivated reasoning is genuine reasoning and is part of scientific reasoning.

Indeed, science is rife with motivated reasoning. But scientific disagreements usually involve issues of metatheoretical frameworks, such as technical methodological differences having to do with study design and analysis (e.g., confounding variables, underpowered studies, appropriate statistical techniques) or assigning different weightings to existing beliefs, theories, or data points that favor one's preferred theory. For example, climate scientists may disagree on the relative roles of solar radiation and atmospheric aerosol concentrations versus greenhouse gases in causing global warming (Hansen & Lacis, 1990), the most accurate method of reconstructing preindustrial global temperatures (Holland, 2007), or the numerical values that should be assigned to some of the assumptions built into the computational models (Lindzen, 1994). These judgments will undoubtedly be affected by one's pre-existing theoretical commitments. But even in such cases data are collected, vetted, and interpreted to maximize overall coherence with existing knowledge and only then added to the knowledge base. Incorrect beliefs are revised or discarded. This is how the reasoning mind works. Why wasn't this the case among Schneider's audience? This issue will be revisited in chapter 13, once we have described the machinery necessary for tethered rationality. We will see that neither heuristics, motivated reasoning, nor sloppy reasoning can explain Schneider's inability to change minds among his audience. We require an explanation involving nonreasoning systems.

It was noted earlier that there were two issues involved in the climate change example: (1) accepting the scientific conclusions and (2) acting on them. The preceding discussion dealt with some of the challenges involved in getting people to accept the science. Getting people to *act* on the science raises a different set of issues. Societal participation in actions to combat climate change constitutes what economists refer to as a "tragedy of the commons" dilemma (Hardin, 1968). The dilemma is that as an individual you receive a higher benefit from not cooperating (using excess energy, continuing to pollute) than from cooperating, irrespective of what other members do, but if everyone cooperates, everyone is better off. These are nontrivial problems, but as we will see, there are some known solutions. I will take up this issue in chapter 9 and argue that a model that recognizes a blended response, incorporating both reasoning and nonreasoning systems, takes us further than just a reason-based model in understanding this failure to act.

We now turn to the four other examples from the introduction and see that the standard rational model fares no better on three of the four.

Rationality in the Real World: Other Examples

In the weight management example, it would seem more advantageous for me to eat less and exercise more rather than risking poor health outcomes. One complicating factor is that the reward for a long and healthy life is in the future, while modified eating and exercise habits need to be implemented in the present. In the decision-making literature, these types of situations are often framed and analyzed as temporal discounting problems (Frederick, Loewenstein, & O'Donoghue, 2002; Reuben, Sapienza, & Zingales, 2010). In this account, we assign a value to a present utility (or profit) and a value to future utilities (or profits). Distant utilities or profits are always discounted (after all, "a bird in the hand is worth two in the bush"). So, for example, if I am giving away money and give you the choice of receiving $10 today or $12 next week, most people would opt for the $10 today, for obvious reasons: it can be spent or invested immediately, serving to maximize utility or profit. By accepting the $10 today, you reduce your chances of receiving nothing in case you do not see me next week, I change my mind, or some other reason. However, if the choice is between $10 today and $100 next week, many people will bypass the $10 today and wait for the $100 next week, calculating that it is more beneficial to delay gratification and take the risk associated with waiting for the larger sum. Where monetary rewards are concerned, this type of explanation often makes sense. Present and future values of monetary sums can be quickly and accurately calculated, given the rate of inflation, interest rates, and other factors. Where individuals diverge in terms of the future value they will trade for the current value, we can explain this in terms of the shape of personal preference or discounting functions and cognitive differences in ability to carry out temporal discounting calculations.

How does this type of explanation fare with my overweight problem? The problem can certainly be formulated as a temporal discounting problem. It could be argued that I do not have the cognitive ability to carry out the temporal discounting calculations or that I have a "skewed preference function." I think this formulation is ultimately unsatisfactory. Even if I don't have the cognitive ability to do the temporal discounting calculations, others making the same choices I do will, so this cannot be the correct general explanation. This leaves the "skewed preference function" explanation, which is just to say that I didn't make the expected, rational choice.

I overeat chocolate cake and pizza because they *taste good*. The decision theorist may want to associate this craving with the skewed discounting

function. That is fine, but it just begs the question. A more satisfactory answer requires a description of the systems driving the craving and how there can be individual differences, which can then explain the different discounting functions. Furthermore, the debilitating health consequences of overeating behaviors are not only distant but abstract and do not have any immediate *feelings*—pleasurable or not—like the taste of chocolate cake, associated with them (until they are actually realized). I will argue that without feelings the consequences cannot even enter into the temporal discounting function to actually impact my decision making. These issues are taken up in considerable detail in chapters 11 and 12.

Consider the third example, where sexual gratification jeopardizes the prized goal of the presidency of the United States. This also can be cast as a temporal discounting problem, and we can postulate that perhaps Edwards was not smart enough to do the calculation. But this is simply not convincing. Furthermore, there are some interesting differences between this problem and the one involving weight management. The goal in this scenario is not an abstract commodity in some distant future. Someone running for the presidency of the United States must *taste* it, *feel* it, *crave* it, every living day they are engaged in the pursuit. So there is an affective component associated with both the immediate sexual gratification and the path to the presidency. If questioned, I do not imagine that Edwards would find greater utility in a current transitory sexual encounter than in the future prize of the presidency. So why did he choose the former and jeopardize the latter? It is possible that he thought he could get away with it. If he had good reasons to believe so, we might consider the choice rational, but as an experienced politician, he should have known better. It is possible that he may have deluded himself into believing this, but then the question becomes, what is the source of the delusion? I think a better explanation for his behavior is offered by the old joke attributed to Mae West: "God gave man two heads, but only enough blood to use one at a time." Jokes aside, any convincing explanation of Edwards's behavior requires an acknowledgment of rationality tethered to evolutionarily older biological systems. This type of behavior is also discussed in chapters 11 and 12.

The fourth example involves the reasoning of my American friend with respect to healthcare. He is willing to incur personal cost or forgo personal benefits just so that guy over there—that one—who doesn't pay taxes (i.e., is a freeloader) doesn't get any benefits. According to Kane (2012), some of the relevant facts about healthcare costs are as follows. In 2012, Americans spent on average $8,233 per person per year on healthcare, or 17.6% of GDP. Countries that have a universal single-payer system spent much less for equivalent

or better healthcare. Canadians and Germans spent $4,400 per person, or 12% of GDP, over the same period, for equivalent or better healthcare.[2] The French and the Japanese spent even less. It is not rational to pay $8,000 rather than $4,000 per person per year for equivalent or inferior healthcare.

This situation is not unlike one I encountered several years ago when my children were teenagers. They were squabbling and fighting over the TV. Failing to restore peace and quiet with simple requests and threats, I offered each of them five dollars if they would stop fighting. My son turned down the offer, stating that his little sister "did not *deserve* five dollars." This was a real choice made in real time, but it makes no sense in terms of maximizing utility. Our theories of rationality cannot account for it. From the perspective of rationality, my son should have been concerned about the fact that he is getting five dollars, irrespective of whether his sister was getting one dollar, five dollars, or five hundred dollars! But his sister's behavior had outraged his sense of justice, and he was determined to punish her (by withholding the five dollars from her) *even at the expense of losing five dollars himself*. When I remind him of this today, he realizes that it was a stupid decision. It would have been more advantageous to take the five dollars. Similarly, many Americans are willing to incur costs or forgo benefits just so someone they feel is not deserving doesn't get any benefits. How do we account for this behavior?

When it is laid out in economic terms, a universal single-payer healthcare system also has a tragedy of the commons component, but as already noted, there are solutions. This is again a situation where evolutionarily older, nonreasoning systems (and their highly affective manifestations) are short-circuiting the rational decision-making process. These systems are considered in greater detail in chapters 9 and 13.

Now let's consider the fifth and final oral contraceptive example. Such problems are discussed in chapter 7. I will agree with Gerd Gigerenzer that the distinction between relative risk and absolute risk, and our preference for natural frequencies over conditional probabilities, go a long way in explaining the poor decision-making in this example. Such explanations implicate issues internal to the reasoning mind. What I have to say in this book does not impact the work on these types of problems. Conversely, this research does not address the concerns that I'm raising.

These examples (excluding the oral contraceptive problem) have three common features that I would like to highlight: (1) the problem or decision seems to lie within the realm of rationality; (2) seemingly "irrational" choices are being made; and (3) there are underlying nonreasoning mechanisms such as autonomic, instinctive, and associative systems modulating

the behavior. In subsequent chapters, we will encounter other examples, but all will share these three features.

In discussing these examples, I have emphasized reasoning and rationality because these are by far the most popular models in the academic literature for construing and analyzing such scenarios. At this stage, it is worth pointing out a second, very different academic account of human behavior offered by sociobiologists and evolutionary psychologists, most recently popularized by Steven Pinker (1997). It emphasizes the continuity between human and nonhuman animals and postulates similar mechanisms to explain the behavior of both. In this type of model, human behavior, like the behavior of nonhuman animals, is not a function of reason but rather a function of a large collection of instincts, which in this literature are referred to as "modules." The most popular version of this model goes by the name of "massive modularity" and states that any particular situation that we encounter will trigger one or more instincts or modules, resulting in a particular choice or behavior. There is not much role for rationality in this model (Cosmides & Tooby, 1994a, 1994b). Some proponents argue that rationality may even be an illusion. This is very much a minority position. I address it in chapter 9.

Given that I'm questioning the explanatory scope of the reasoning models and appealing to evolutionarily older nonreasoning systems (including instinctive systems), the reader may be thinking that I will be advocating a massive modularity type model. This is not the case. I believe that evolutionary psychologists provide a critical insight that needs to be incorporated into the solution. But despite my appeal to nonreasoning systems, I am confident that we are not simply steered by them. We do have the ability to reason and make choices. As an illustration, consider the following example in which I used my rational mind to modulate my (nonrational) eating behavior.

A few years ago, I was attending a conference on reasoning sponsored by the Parmenides Foundation, held on the isle of Elba. The host organization was taking very good care of us, offering food and drink on every possible occasion. After several days of this, I was satiated and determined to limit my food intake for the sake of my health. After a particularly interesting talk on the neurobiology of addiction, I began conversing with the presenter. (I was fascinated by the claim that addictive behavior is not a choice.) It was lunchtime so we were all walking toward the beach, where lunch would be served. I said to my colleague, "I will come with you so we can continue our conversation, but I'm satiated so I won't eat anything." We sat down and continued talking about his presentation. The waiter brought menus. I thought, I will not order anything, but I'll just look at the menu. Upon

examining the menu, the pizza looked very appetizing. I said to myself, "Well, you know what, I will order the pizza but I will only eat half of it." The pizza arrived; I cut it in half and ate the first half. I then ate the other half. I'm overweight and suffer from the typical consequences. I do not need an extra slice of pizza. Why did I eat it? I derived such pleasure from it that I could not help myself.

The next day, at lunchtime, my wife was with me. We walked to the beach where lunch was being served, and I said to her, "The pizza is very good, but don't order your own. I will order one and we will split it." I ordered the pizza. When it arrived, I cut it in half and ate my half. I then looked for the other half, but my wife had already eaten it. So, reluctantly, I did without. I used my reasoning abilities to put myself in a situation where I would not be confronted with the temptation of eating the other half of the pizza, and I was thus able to control my food intake.

The point of this anecdotal story is to highlight that the rational mind is able to exert some control over behaviors through various strategies to avoid being totally at the mercy of deep-seated evolutionarily older mechanisms. To prevent overeating, I placed myself in a situation where food was not readily available. Some of the questions we will need to explore are: To what extent is this possible? What is the nature of the interaction between reasoning and nonreasoning systems? What is the common language used for communication across different systems? How do we account for individual differences in behavior? Who is in charge of the tethered mind?

The idea that reason alone is not sufficient to account for human behavior is being voiced by an increasing number of researchers, particularly in the social, economic, and political sciences (Kahan, 2016; Oliver & Wood, 2018; Young, 2019). These researchers frequently contrast reason with "heuristics," "emotions," "gut feelings," and "unconscious processes." They are trying to account for their intuitions and data, but this vocabulary lacks substantive conceptual machinery to allow them to say what they want and need to say. (The first two terms are part of the reasoning mind, and the latter two are undefined or unhelpful.) It is not their fault. They are not in the business of developing the models to explain behavior but rather applying machinery developed by cognitive scientists to their respective problem domains. The cognitive sciences have come up short. I will argue that tethered rationality provides a much richer repertoire of conceptual machinery to explain their intuitions and data.

Finally, models of human behavior are not only necessary for explaining political, economic, social, and moral behaviors but also have consequences for changing these behaviors. Tethered rationality is no exception.

Some readers may be disappointed in its implications. It suggests that many behaviors—such as racism, sexual harassment, cheating, adherence to false beliefs despite counter evidence, and overeating resulting in obesity—however unacceptable, are often driven by early maturing autonomic, instinctive, and associative neural systems and cannot be easily changed simply by changing beliefs. Even behaviors based on reasoned social constructs can become deeply entrenched once neural systems mature. Attempts at belief revision through a weekend of "sensitivity training" will be ineffective. This does not mean that such unacceptable behaviors cannot be changed at all, but it does mean we will need to understand the underlying biology of each specific behavior and apply behavior-specific remedies. Even then, there may be limits.

Organization of the Book

This volume is organized into six parts. Part I introduces the rational animal and the enigma of rationality. It is accepted that we *are* the rational animal but our rationality is not disembodied. It is tightly tethered to evolutionarily older autonomic, instinctive, and associative systems. Before we can tell the story of the tethered mind we need to have a common understanding of each of these systems. Part II is devoted to characterizing autonomic, instinctual, associative, and reasoning behaviors and systems. Each is characterized in terms of the following five dimensions: (1) function of the behavior, (2) tightness of causal coupling between stimulus and response, (3) origin of behavior, (4) underlying mechanisms, and (5) brain structures involved. The behaviors are found to differ along these five dimensions and are accordingly assigned to different systems or "kinds of minds." Considerable effort is made to explain what behaviors each type of mind can and cannot explain.

The characterizations of each type of mind are reasonably standard. The autonomic mind (chapter 3) is characterized in the manner found in most biology textbooks. The instinctive mind (chapter 4) draws on the models of Konrad Lorenz, Nikolaas Tinbergen, and other ethologists. The characterization of the associative mind (chapter 5) follows that of B. F. Skinner and other behaviorists, enriched by the insights of William James. The exposition of the reasoning mind (chapter 6) draws on the ideas of twentieth-century philosophers and cognitive scientists, including Ernst Cassirer, Donald Davidson, John Searle, Noam Chomsky, Herbert Simon, Allen Newell, Jerry Fodor, and Zenon Pylyshyn. I will always use the terms *autonomic*, *instinctive*, *associative*, and *reasoning* in the manner specified in the corresponding chapters. What is new in this book is my assertion that

___-1
___0
___+1

all these systems are in play (to various degrees) in *all* human behaviors; that is, human behavior is a blended response. This will be a blatant truism for many general readers. It will be less obvious to many of my colleagues. It was expunged from us during graduate school.

Once an understanding of the kinds of minds associated with each type of behavior has been established, part III reviews the theoretical frameworks of reasoning, built from the conceptual machinery of the cognitive mind. Chapter 7 considers models of formal reasoning, particularly various dual mechanism accounts of reasoning resulting in the widely accepted distinction between heuristic and analytical systems. I flag a number of sources of confusion in this literature, but by and large, the literature is not relevant to the types of issues of interest in this volume, so it is set aside. The reader not encumbered with the belief that heuristics explain the examples raised earlier could bypass chapter 7.

Chapter 8 reviews the literature on conceptual coherence (inductive reasoning). It begins by raising a number of issues largely of concern to philosophers and cognitive scientists, and then shifts to considering real-world problems from the realm of science (Galileo's arguments about motion) and politics (first impeachment of Donald Trump). Once we reach the latter we are confronted with a whole set of issues that the cognitive science literature cannot address. Any satisfactory explanation requires an appeal to the engagement of nonrational systems.

Part IV begins the development of the positive account of the tethered mind. Chapter 9 provides the behavioral data for the "blended response" hypothesis. It begins by considering instincts in their modern reincarnation as "modules," from the work of Leda Cosmides and John Tooby. While I reject the massive modularity model, I find value in their insights regarding the role of instincts in human behavior, specifically their explanation of behavioral data from a famous reasoning task in terms of "cheater detection" instincts, rather than coherence relations. An exploration of the related concepts of self-interest maximization, fairness, trust, cheating, and punishment suggests that they are reasonable candidates for instincts, albeit all but the first may be specific to humans. Given the types of decision-making examples of interest (e.g., climate change, universal healthcare), I turn to the work of a small but influential group of economists and mathematical biologists, such as Ernst Fehr and Martin Nowak, who explain human choices on such problems as interactions of the instincts noted above. A careful examination of the data from this literature shows that these instinctual systems are modulated by reasoning systems and that a full account of the data requires postulating a blended response involving both instincts and reason;

that is, a model of tethered rationality. The appendix to chapter 9 reiterates the distinction between reason and instincts and discusses the conceptual pitfalls in trying to account for human behavior just in terms of "modules" or instincts. The reader more interested in the positive account of tethered rationality could bypass this appendix without sacrificing continuity.

Chapter 10 turns to comparative neuroanatomy for the neural underpinnings of tethered rationality. The challenge here is to show that the hierarchy of evolved behaviors (autonomic, instinctive, associative, reasoning) is mapped onto a hierarchy of evolved brain structures. The interconnections between brain structures supporting tethering are readily apparent at the level of anatomy and physiology. A further challenge is to illustrate the differences in brains of organisms that can reason and those that cannot.

Once we have a story of hierarchically organized behaviors, underwritten by hierarchically organized (but interconnected) brain structures, it is necessary to account for how these various systems contribute to behavioral responses. The tethered rationality model allows each of the four systems to generate responses to any environmental perturbation, but the organism is restricted to a single behavioral response at a time. This requires some global integration function that takes input from each of the systems and generates a blended response. Chapter 11 advances the speculative conjecture that what is common across each system is *feelings*. Feelings are generated in old, widely conserved brain stem structures, and are evolution's solution to initiating and selecting behaviors. Reason is as much about feelings as lust and the taste of chocolate cake. Feelings provide the common currency that allows communication across systems and the calculation of an overall blended response. This controversial solution has the additional benefit of bridging the divide between the cognitive and noncognitive and driving reason back into the biology, where it belongs. The works of neuroscientists such as Jaak Panksepp, Kent Berridge, and Morten Kringelbach play a central role in putting together some of these ideas.

Chapter 12 considers the control structure for tethered rationality. Who is in charge of the tethered mind? Reason is not the CEO. In fact, I will conclude that there is no CEO. All four systems affect the resulting behavior in the currency of feelings. The system is set up to maximize pleasure and minimize displeasure. The model is illustrated with several examples and, in particular, my difficulty in losing weight. At this point, I return to complete the explanations of some of the other examples introduced throughout the book.

Part V takes up the question of why it is so hard to change certain beliefs, ranging from climate change being a hoax, to the MMR vaccine causing

autism, to gender identity being a socially constructed choice independent of biology, to "Democrats are evil people, they hate America," among others, despite evidence to the contrary. Insofar as reason is viewed as untethered to the biology (as in mind/body dualism or the cognitive science computer program/hardware metaphor), it should have an unfettered ability to update and revise beliefs (perhaps constrained only by time and memory resources). This is not the case. Driving reason back into the biology provides some answers.

Chapter 13 provides one answer to this puzzle by applying the tethered rationality model: reason is only one component of the system. The other systems reason is tethered to may prevent belief revision or belief revision may not be sufficient for behavioral change. The introduction of the in-group/out-group instinct allows us to complete the explanations for climate change denial, the impeachment debate, and some Americans' aversion to universal healthcare. The other constraint on belief revision is neural maturation. This phenomenon is independent of the tethering of reason and largely comes into play where large-scale global belief systems, known as worldviews, need to be revised late in life. In chapter 14 I propose that with the maturation of the association cortex in adulthood there may not be sufficient neuronal resources left for large-scale architectural neural reorganization, making global belief revision challenging.

Part VI briefly considers some of the consequences of the tethered mind and concludes the volume. Different models of human behavior come with different control structures and have different social and legal consequences. One consequence of the tethered rationality model is that changing certain deeply seated behaviors (however socially unacceptable) is not a matter of just changing beliefs. To consider remedies beyond belief revision, we need to understand the biological underpinnings of the specific behaviors. This sensitive topic warrants a separate volume. However, in chapter 15 I very briefly consider some concerns and consequences of the tethered mind, and conclude the volume by offering a few closing thoughts to colleagues and the general reader.

Let us begin the journey by stepping back and reconceptualizing the problem of rationality and human behavior in a broader context.