1
Beliefs Matter!

We human beings — especially, those of us who live in industrial societies — are, on the whole, pretty rational and practical in our decision making, aren't we? Don't we consider every situation we fall into pragmatically, weigh our alternative courses of action rationally, decide which of them will best achieve our goals, and then do it?

No, we often don't make decisions that way, and I'll give a few examples. These examples come both from experiments by psychologists, sociologists, and educators, and from daily life including politics. There are so many possible examples of the themes discussed here that the choices have to be somewhat arbitrary; readers will no doubt think of equally good or better illustrations. But we will see that these many different aspects of life all illustrate a few common principles about human behavior. Rather than being under rational control, our perceptions and actions are heavily influenced by attitudes and beliefs that are derived from observations colored by our emotions. Moreover, these attitudes and beliefs may have arisen from events that have little or nothing to do directly with the current situation.

Should the importance of emotionally derived beliefs make us despair for our future? Does it mean that underneath our vaunted civilization we are just "naked apes," doomed to petty tribal warfare because of our prejudices? Aside from the slur on our ape brothers and sisters, this book will argue quite the contrary! We can gain better understanding of human nature by facing up to the complexity of our beliefs and how they influence our behavior, and use this understanding to improve our lot. Some of the common patterns of behavior are rational and some are irrational, but all can be studied with the help of modern methods in both the natural and social sciences.

When we understand our beliefs, and more generally our mental constructs, more deeply, we can then integrate them with what modern science too often neglects: our *values*! Then we can try to decide

which of our widely held beliefs we do or don't wish to encourage. Some widely held conventional beliefs about human personalities and mental states, I will argue, are harmful to human relationships and need to be challenged in order to create a genuinely cooperative society. Insights from the social sciences combined with understanding of how the human brain performs cognitive functions will suggest, surprisingly to the cynics among us, that these harmful beliefs aren't supported by scientific analysis. Such beliefs are what this book calls *common nonsense*. As the folk saying goes, it's not what you don't know that hurts you, but what you know that ain't so. But we will also see that other widely held folk beliefs are both helpful to society and well founded in science. These are what this book calls *common sense*.

We will also see that brain science can help us find ways to distinguish common sense from common nonsense, in the way this book uses these terms. The scientific tools we use combine experimental findings from psychology and neurophysiology laboratories and clinics with the growing use of mathematical and computational theories of how the brain works, by means of theoretical constructs called *neural networks*. That mathematics, a field that is widely considered "precise," "rational," or even "machine-like," can yield insights into the irrational, emotional, and even spiritual aspects of human behavior will probably come as a surprise to many readers. But that is one conclusion that will emerge from this book. In fact, my own career in neural network research has been motivated by this search for connection between the precise and the imprecise.

To chronicle human irrationality, let us start with some situations in which having a belief makes things happen that confirm the belief!

Case 1: Self-Fulfilling Prophecies

The notion of *self-fulfilling prophecy*, now widely used in the behavioral sciences, was described by the Twentieth Century sociologist Robert Merton.[1] Merton cited a variety of previous thinkers who had written about the same idea, among them Karl Marx, Sigmund Freud, and the earlier sociologist W. I.

Thomas, but he seems to have invented the term. Basically, it means that when people believe something to be true, their belief makes them act in ways that lead to consequences that support their belief.

One example Merton gave of self-fulfilling prophecy deals with a common occurrence during the Great Depression of the 1930s. A large number of people would hear a rumor that their local bank was about to collapse, and so rush to draw their funds out of the bank. In turn, this run on the bank would *cause* the very collapse that these bank depositors had feared.

Another example Merton gave deals with beliefs about war. Though he wrote at a time when the two superpowers (the United States and Soviet Union) were enmeshed in a Cold War that is now over, his warning remains timely about present and future international conflicts:

> ... it is believed that war between two nations is inevitable. Actuated by this conviction, representatives of the two nations become progressively alienated, apprehensively countering each "offensive" move of the other with a "defensive" move of their own. Stockpiles of armaments, raw materials, and armed men grow larger and eventually the anticipation of war helps create the actuality.[2]

As Albert Einstein put it more succinctly, "You cannot simultaneously prevent and prepare for war." And another sociologist, C. Wright Mills, described how once this kind of war psychology gets going, it is reinforced by the tendency to label anyone in a position of power who opposes it as "impractical."[3]

Merton's third example deals with racial prejudice in American labor unions, as discussed more deeply by Gunnar Myrdal.[4] In the period between the two world wars, many union workers and organizers justified excluding African-Americans from their unions on the grounds of personality traits that they observed among their black fellow citizens. Blacks, these unionists argued, are strikebreakers. They are undisciplined in the art of collective bargaining. They have low standards of living so they rush to take jobs at low wages. All these observations combined to make the white union workers believe that African-Americans were naturally hostile to the working class. These white workers were not, by and large, nasty

people, and their observations were *accurate* — in regards not all but many blacks. But the whites failed to see that the traits they observed were produced by the exclusion itself! Blacks became low-wage strikebreakers *because* they were out of work and excluded from the unions that would have enabled them to get high-wage, skilled jobs. In industries where blacks succeeded in getting admitted to unions, their "strikebreaking behavior" disappeared.

In the three examples just given, self-fulfilling prophecy had harmful effects. But there are other occasions when it can have helpful or neutral effects. An example of people being helped by a self-fulfilling prophecy occurred in a famous experiment on elementary education. The psychologists Robert Rosenthal and Lenore Jacobson randomly divided grade school students in South San Francisco into two groups that were roughly comparable in performance on intelligence tests.[5] The experimenters misled the teachers, deliberately telling them that according to an intelligence test, one group of students had higher potential for achievement than the other. Based on this bogus information, the teachers gave more encouragement to the students they believed had more ability. A year later, in response to their teachers' treatment, the supposedly "higher potential" children had actually done better in school! This was called *the Pygmalion effect* after the legendary Greek sculptor who could teach a statue to speak.

More generally, Merton observed, self-fulfilling prophecy is an offshoot of our belief in the continuity and predictability of the world. This is a belief that we learn as children when we discover that our own explorations and manipulations of the environment can produce expected effects and reinforce our behavior;[6] some theories of the brain are starting to address this development process.[7] Belief in continuity can keep us entrenched in maladaptive habits, but also gives us the reassurance we need to believe that our actions are effective.

In a series of experiments by the social psychologists Mark Snyder and William Swann,[8] self-fulfilling prophecy has a neutral effect but still biases perceptions. Snyder and Swann showed that if people are told to test a specific hypothesis about another person — for example, find out whether Alice has an outgoing personality — they will tend to ask questions designed to *confirm* the hypothesis. In the

"outgoing" case, for example, people would ask Alice questions like "What would you do to liven things up at a party?" or "Where do you go to meet interesting people?"

Snyder and Swann noted that the ways people usually act in testing theories about others are contrary to accepted scientific method. When one is testing a scientific theory, the general practice is to subject the theory to the most rigorous trial possible. That is done by looking for evidence that would *disprove* the theory, not evidence that would prove it. By contrast, people in real life situations are more likely to look for evidence that will *prove* the theories they have already developed.

Snyder and Swann suggested that people are biased in favor of any hypotheses they somehow hear or learn about, that is, hypotheses that are cognitively *available* to them.[9] These psychologists found that the bias they created toward seeing people as outgoing would remain even when the experimenters told the subjects that it was very uncertain whether the person really was outgoing. The bias also remained if the subjects were told the person belonged to a group (e.g., a sorority) in which a majority of members were not outgoing.

Often, beliefs bias actions, actions perpetuate the beliefs, and what engineers call a "positive feedback cycle" gets going. Later we will discuss a theory, based on mathematical and computer models of brain function, for how such feedback might occur. As this happens in individuals, so can this kind of self-perpetuating feedback happen in societies, and I will give a few examples in the next section.

Some readers will object that societies have dynamics all of their own that can't be reduced to the dynamics of their individual members. How much this is true is a subject of long-standing controversy among sociologists.[10] The pioneering sociologist Émile Durkheim was particularly noted for his statements that society is a separate level from the individual and obeys its own laws.[11] But Durkheim also regarded a society as somewhat analogous to a biological organism, with particular needs that must be fulfilled for healthy functioning.

Moreover, the majority of sociologists believe there is constant mutual influence between individuals and societies. A given society's norms for behavior constrain the possible actions of each individual. On the other hand, the needs of individuals constrain the possible structures and customs that a society can

form. The rules governing mutual influence between individuals and societies have been addressed by many social theorists, such as George Herbert Mead, Jurgen Habermas, Randall Collins, and Anthony Giddens.[12] This book deals more with psychology than sociology, and will not discuss these theorists further. However, their work hints that at least *some* behaviors of societies reflect the averages of individuals within them. Other behaviors of societies reflect general dynamic principles that are common to all complex systems — whether the complex systems are societies composed of people, individual brains composed of cells, individual cells composed of molecules, et cetera. For these reasons, I believe that the same general concepts, such as self-fulfilling prophecy, are meaningful when applied to individuals, groups, and societies. Also, this book *will* discuss the influence of social contexts on individual psychology — which makes one of my main points, the tremendous variability of each of our mental lives, even stronger.

Case 2: Entrenched Social Systems

The anthropologist Marvin Harris devoted two books to the origins of some otherwise baffling social customs in many cultures.[13] One of Harris' studies dealt with societies such as Pharaonic Egypt and Ming China, which took off into advanced civilization through massive building projects such as huge pyramids, waterworks, or fortresses: to these he gave the name "hydraulic societies."[14] These projects, while they often had practical value, were also meant to magnify the perceived power of the ruling family. Their completion required the labor of large numbers of slaves or subordinates. The building projects provided a focus for an entire society, and while they were going on attracted a great deal of enthusiasm from the people as a whole. So in their early years, these societies enjoyed high morale and coherence, with widespread loyalty to a strong central government.[15]

The same factors leading to early success of the hydraulic societies, however, were harmful to their development as time went on. The building projects had spawned forms of administrative and social organization that were rigidly stratified and lasted for hundreds of years. Once in place, these hierarchical structures were almost impossible to dislodge, leading to continued efficiency but slowly sapping average

people's initiative. Among average people, a "positive feedback loop" had developed between their lowly status and their belief in their own inferiority to the royal families. This was evidenced by practices of humility such as the kowtowing of Chinese peasants. In other words, people's entrenched lack of belief in their own worth as individuals became self-fulfilling.* Eventually, each society declined, and rigid centralization was one of the main reasons.

Harris regards the development of rigid hierarchies as a successful response to genuine economic factors, but speculates that an equally successful and more egalitarian response might have been possible before the social structure was entrenched. Yet this sort of positive feedback between beliefs and practices, like self-fulfilling prophecy, is not always harmful. Harris traced the origin of other customs that have outlived their original "rational" purpose, but still perform a beneficial function in keeping societies or groups together.[16] These include, for example, dietary restrictions in the Jewish and Hindu religions. Harris claims that the ancient Israelites' restriction against eating pig meat was put forth for economic reasons. Pig herding wasn't viable for a nomadic culture in a semi-arid terrain, but people were tempted to raise pigs anyway because their flesh tasted good. The only way the leaders could stop the people from doing it was by prohibiting it on religious grounds.**

Harris also proposed economic and ecological reasons for the Hindu veneration of cows and prohibition against eating their meat (and later, by extension, all meat). This occurred because Indian peasants valued cattle for purposes other than food: as draft animals and as a source of dung to burn for fuel.

---

* Perhaps modern America is suffering now from the opposite self-fulfilling prophecy: too much individualism has bred a disbelief in the possibility of close community.

** Some other scientifically educated people now believe that the prohibition against pork was made for health reasons, because of the threat of trichinosis from insufficiently cooked pork. If that were the case, it wouldn't change Harris', or my, basic argument: that customs instituted for practical reasons often survive long after the practical reasons are no longer valid.

The examples given in this section, and the last one, give ample evidence that human behavior isn't rationally based. Instead, it is determined by a blend of reason, emotion, and habit, and the relative influences of these factors can change over time or with a change in context. Moreover, reason isn't always "better" than emotion or habit: instead, all three have their separate useful functions. Emotion, for example, powers love and creative impulses, and determines many of the goals that motivate behavior. Also, detailed reasoning about all possible alternatives often makes decisions too slow for real-life contingencies, and emotion (for better or for worse) facilitates coming to conclusions.[17] Habit can be valuable in encouraging continuity of behavior, which is needed to establish cultures, communities, and relationships.

Let us now shift gears to a different type of irrational decision making: one that involves processing information about probabilities or certainties of events.

Case 3: Judgment Under Uncertainty

The late Twentieth Century psychologists Amos Tversky and Daniel Kahneman performed a long series of experiments in which subjects were asked to envision themselves in various situations and make hypothetical choices among alternatives offered to them.[18] These choices involved different amounts of gains or losses (of money, or something else valuable such as health) with different probabilities of occurring. The results of these experiments have revolutionized thinking about how people make real-life decisions under conditions of uncertainty or risk. Tversky and Kahneman's results provide evidence that humans don't always "calculate what's best" and do it. What's more, they found that professors of statistics are no more rational in their own decision making than anybody else! But for now, in keeping with the theme of "beliefs matter," let us concentrate on one aspect of these psychologists' work: the role of context and wording in the choices people make.

Tversky and Kahneman's clearest demonstration of how wording can influence choices was their experiment about giving people a choice of two possible treatments for an epidemic.[19] The same problem

was posed to two groups of subjects, with over 150 people in each group, and both selected from the same population (students at Stanford University and the University of British Columbia):

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

The next part of the statement, the presentation of the two alternatives, is what was different between the two groups. The alternatives presented to the first group were:

> If *Program A* is adopted, 200 people will be saved.
> If *Program B* is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Look for a moment at the consequences of *Programs A* and *B*. Since 1/3 of 600 is 200, the two programs have the same effect of saving 200 people "on the average" (this can be formalized mathematically but it's not needed here). But you probably feel that the effects are very different. Under *Program A*, you are sure of saving at least some of the people who would otherwise die. Under *Program B*, there is a chance of saving a lot more, but the chance is less than even, and otherwise you save nobody. So *Program A* probably sounds more attractive, right? If you favor *A* over *B*, you agree with a majority (72 percent, to be exact) of Tversky and Kahneman's subjects.

Now the second group of subjects were given a different choice of alternative programs:

> If *Program C* is adopted, 400 people will die.
>
> If *Program D* is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

In this case, the 400 dying under *Program C* and the 600 dying (most probably) under *D* don't feel that different, psychologically. But under *D*, there is at least a *chance* (and not a slim one, but a 1 in 3 chance) that nobody will die. So if that glimmer of hope makes *D* seem more attractive to you than *C*, you agree with an even larger majority (78 percent) of Tversky and Kahneman's subjects.

So of the four programs mentioned, a majority of people prefer *A* to *B and* prefer *D* to *C*. But wait a minute! Isn't 400 equal to 600 minus 200? So if 200 are saved, 400 will die. That is, the actual consequences of programs *A* and *C* are *exactly the same*! Likewise, if none are saved, 600 die, and vice versa. So the consequences of programs *B* and *D* are also exactly the same. This means that the two sets of preferences by the two groups of subjects are logically inconsistent.

How do we explain this inconsistency? This is where beliefs enter in. The way the two sets of consequences were worded created different frames of reference, that is, different baselines of expectations in the two groups of experimental subjects. The subjects choosing between *A* and *B* had their choices presented in terms of people *being saved*. This means that there was an implied baseline reference point, the situation of all 600 dying, and the two alternatives were evaluated as to how much better they felt than that baseline. The subjects choosing between *C* and *D*, by contrast, heard their choices presented in terms of people *dying*. This meant that the two alternatives were now being compared with a different baseline reference point, the situation of nobody dying. For the reasons given in the discussion above, the relative attractiveness of the two alternatives is different when there is a different reference base.

Is there any way to "measure" such subjective preferences scientifically? A growing group of researchers is trying to do so, by building bridges between the precise techniques of mathematics and the imprecise phenomena studied in psychology. This involves a difficult translation from fluid concepts into simplified concepts with definite boundaries. So the types of theoretical constructs that do the bridging,

which are called *models*, depend on a kind of selective approximation. Researchers constructing models need to abstract out from a vast and confusing set of behavioral data a few variables (either measurable or artificial) that may be the key ones to understanding the phenomena being studied. Then these researchers either study their simplified constructions mathematically or simulate them on a computer. The models that arise are called *neural network*[*] models: these are based on structures of connecting regions that are in some way similar to parts of the brain.[**]

A mathematical model of the Tversky-Kahneman framing effect was developed by two neural network theorists, Stephen Grossberg and William Gutowski.[20] How this model works precisely is beyond the scope of this book, and not essential to our discussion. But a few words are in order about how emotional concepts like preference can be represented mathematically at all.[***]

Grossberg and Gutowski's basic idea is that to any anticipated event, or outcome of action, is assigned some positive or negative number representing its "emotional value." For example, a very delightful event, or one that satisfies a basic drive such as hunger, might have a value of +10. (This number is arbitrary: as yet we don't know enough to assign it to a precise biochemical variable in the brain, even though brain regions that calculate emotional values have already been identified, such as the *hypothalamus* and *amygdala* shown in Figure 1.1.) A less important or intense but still pleasant event might have a value of +2, whereas an unpleasant or dangerous event might have a value of !10. These quantities, in the mathematical model, are based not only on the anticipated event itself, but in how much it is a *change* from

---

[*] Some readers will be familiar with the term "neural network" as used for a class of machines that can perform "intelligent" functions — for example, face recognition; handwriting classification; diagnosing faults in automobiles; building autonomous robots; predicting the stock market, et cetera. The two uses of the term aren't really different. This is because the same set of computational techniques, with variations, can be applied *both* to understanding how our brains perform cognitive and behavioral functions, and to mimicking those functions in machines.

[**] As we will see in later chapters, *all* mental functions affect, and are affected by, biological processes in the brain — regardless of whether those functions were originally produced by interactions with society or by ingestion of biochemical substances.

[***] For a more detailed description of Grossberg and Gutowski's network link to "Grossberg and Gutowski"

the current, or expected near future, state of the environment. Expectation, which is a type of belief, is a key element in this process. Recent past events create an expectation, represented by certain levels of chemical neurotransmitter substances, at certain synapses. This in turn creates a kind of "mood" which colors the emotional desirability or undesirability of any anticipated changes in the environment.

So the psychological mechanism that explains Tversky and Kahneman's "illogical" data seems to be the one that selectively processes unexpected or changing events. While they differ in a great many details, it seems reasonable to assume that a lot of psychological phenomena involving novelty or surprise have some common elements — and perhaps, though we are less certain of this, some common brain mechanisms.

Like the self-fulfilling prophecy tendencies discussed above, the tendency to enhance novel or surprising events (or thoughts) can be helpful or harmful in different circumstances. For example, enhancing novelty can be harmful when it leads to seeking a new thing just because it's a fad. It can be harmful when it leads to sudden changes in lifestyle or in religious or political ideas, just for the stimulation those changes produce. But enhancing novelty can be helpful when it leads to creativity. Imaginative solutions to problems, whether in daily life or in the scientific research laboratory, often involve testing new combinations of old ideas, or noticing and exploiting elements that have changed in the context.

Irrationality and the Brain

All three of the cases discussed above strongly indicate that actions, decisions, and perceptions are inextricably, and dynamically, linked with emotions, beliefs, and attitudes. The types of mathematical models used in neural networks, far from restricting human behavior to be purely rational, reflect the interconnectedness of all these psychological elements in behavior. In the terms of this book's title, the belief that we make decisions on a purely rational, pragmatic basis — weighing the "pros and cons" of our different alternatives — is seen as common nonsense.

As an illustration, I had to decide, as a senior in college, among offers of scholarships in mathematics at several graduate schools. All of the schools were comparable in prestige, and the amounts of money offered were in the same range. After some of the possible schools had been rejected on definable grounds — unpleasant atmosphere, difficulty in meeting women, or being too far from home — the choices had narrowed down to Columbia University or the University of Chicago. On "rational" grounds Columbia seemed better: it had more professors working in what was then my field, and it was close to my parents who then lived in New York City. But having lived in the neighborhood of the University of Chicago during childhood and early adolescence, I was intuitively driven there by a powerful yet ill-defined "homing instinct," and went to Chicago instead. In retrospect, my experience there was quite positive and, anyway, my professional direction changed not much more than a year later — so making the intuitive, "nonrational" choice worked out well.

Scientists have made some conjectures about the brain's role in combining reason, emotion, and intuition to make decisions. The late behavioral neuroanatomist Walle Nauta said that reason and intuition frequently conflict, with intuition winning.[21] Often, Nauta said, we decide that one plan is better than another on the basis of objective reasoning, but thinking about the first plan makes us *sick* — quite literally sick, via the various pathways of what is called the *autonomic nervous system* connecting the brain to the heart, endocrine glands, and digestive organs. If that happens, we reject the "reasonable" plan in favor of the other plan. Nauta called this phenomenon *interoceptive censorship of plans.*[*]

The brain's frontal lobes (see Figure 1.1), which are mentioned a lot in this book, serve to "mediate" between the "emotional" and the "rational" parts of our brain. (Quotation marks are added

[*] *Interoceptive* is the term neurophysiologists use for perception of the states of one's own internal organs.

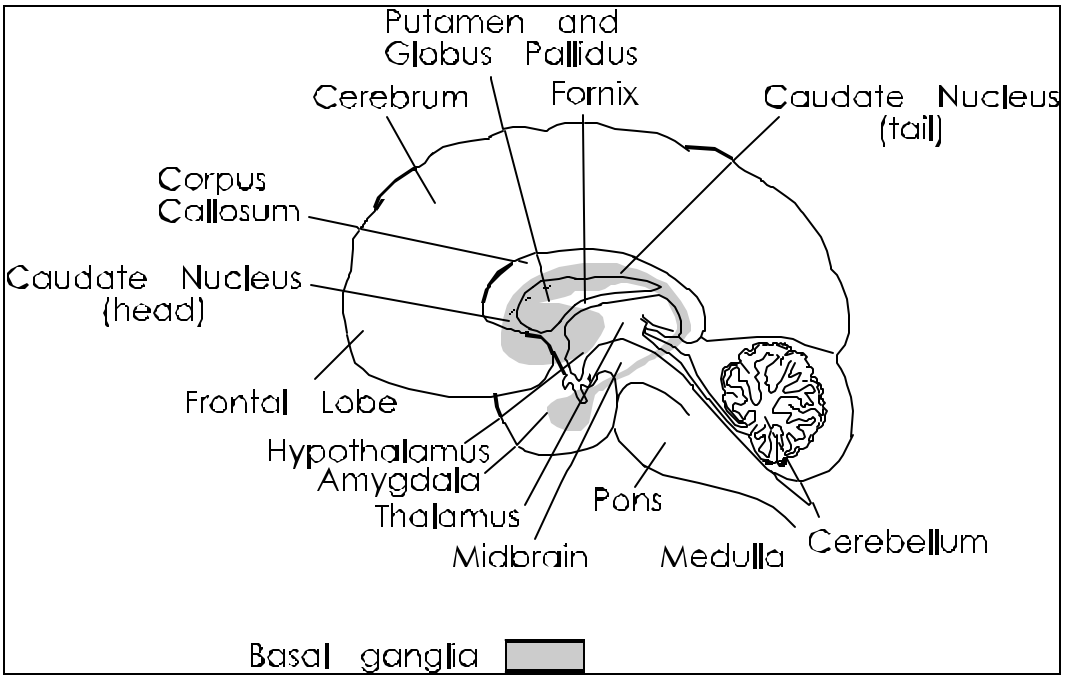

Figure 1.1. View of the human brain from its center line, showing locations of some of its major subdivisions. The *fornix* is a pathway linking the hypothalamus with parts of the limbic system. The *corpus callosum* is a pathway linking the left and right sides of the cerebral cortex or outer part of the brain. (Adapted from Thompson, 1967, with permission of Harper and Row Publishing Company).

because all psychological functions, including emotion and reason, really involve pathways, not separate areas. But some brain regions do have specific roles which we discuss later). The clinical neuroscientist Antonio Damasio discusses patients with damage to a part of the frontal lobes (called the *orbital* part) that connects, via nerve pathways, with areas below the cortex (surface of the brain) that are involved in emotional expression.[22] Patients with orbital frontal damage have largely normal memory and cognitive functions. They even do as well as normal people on a card sorting test used by clinical neuropsychologists, which is a test of damage to another part of the frontal lobes. But these orbitally damaged patients lack the usual emotional responses to sensory events. For example, one of Damasio's patients drove to his office

on a very icy road and, judging from the way he recounted having done so, did not feel scared like most normal people do. Another of Damasio's patients, with damage to the same brain area, could readily arrive at many different

options for dealing with any situation, but couldn't make an effective decision between these options because he found minor difficulties with each one. When trying to decide which of several restaurants to eat at, for example, he would drive to each restaurant to help him decide but still not be able to make up his mind because of minor objections to each. This happened because his emotional response wasn't appropriately gauged to the situation.

Brains, Beliefs, Attitudes, and Society

How might all this knowledge about our brains affect our views of society? Since beliefs have a great influence on behavior, changing society for the better doesn't just involve "pragmatic" political and economic programs: it also involves "cleaning up" our beliefs and attitudes. A growing number of people are recognizing that the early Twenty-first Century world is in a crisis largely of our own making, and that some shifts in our basic beliefs are necessary to survive this crisis.[23] The challenges of the present era — operating a global economy with world-wide communication links, and at the same time keeping alive a sense of community, minimizing violence, preserving our environment, and trying to eliminate poverty — make our attitudes and beliefs more critical than ever. Some of our prevailing common nonsense, our "conventional wisdom" that isn't really wise, about human behavior and personality, stands in the way of the paradigm shifts we sorely need to make to deal with these challenges.

The implications of our psychological theories for politics, and for building the future, will be developed in the later chapters of this book. First, the earlier chapters will use ideas from science — in particular, experimental psychology and sociology, neurobiology, and neural network theory — to develop a view of human nature. This view will suggest how to judge which beliefs are common sense, which are common nonsense, and what changes are possible for society. The discussion will wander between

different fields, and between academic subjects and mundane life. But in all these different areas, it will keep arguing variations of a few points. Among these points are the following:

- Modern methods in brain science and neural network theory provide insights into why we humans don't always meet our potential. These insights also suggest how social institutions, therapy, education, et cetera might be structured to enhance human potential. As brain science develops further, its social applications will form a growth field. This *doesn't* mean the natural sciences will replace the social sciences or the helping professions. Nor will science replace folk psychology or the intuition ("common sense") of average people. Instead, science can be a partner with all these other pursuits in joint work toward social progress. At best, scientific insights can destroy the "cult of the expert" by bringing out the creative "scientist" in everybody!
- Human nature is neither "good" nor "evil" in the senses that most people think of these words. The same person can have radically different and mutually inconsistent behaviors in different contexts. When two or more conflicting behavioral patterns coexist in the same adult, Darwinian natural selection isn't much help in choosing between them.
- There is a standard for the best possible human functioning, often called *self-actualization*, which has some biological basis that is yet to be well understood. But human behavior doesn't always self-actualize or optimize potential.
- The fact that we don't always act optimally should be a cause for hope, not despair. This is because it means that war, poverty, environmental pollution, and rigid dominance hierarchies are *not* the result of optimal functioning, so they can be improved on.
- Common nonsense falls into one or more of a few repeating patterns. Some of our common nonsense beliefs can be characterized as promoting dominator relationships based on hierarchical rankings, as opposed to partnership relations based on equality and mutual respect.[24] Others fall into a pattern of making false "either-or" distinctions (e.g., one can be socially responsible or pursue pleasure but not both). Still others result from confusing the *average attributes* of a group

with the attributes of *every individual* in the group, or confusing the way someone acts *most of the time* with the way he or she acts *all the time.* Our common sense, creative responses to problems often come from seeing through appearances to the variability in each person or group. *What you get is more than what you see*!.

! "What you get is more than what you see" leads to some norms for behavior. It suggests that we need to try to avoid judging people as *incapable* of doing something (whether it's being good parents or learning mathematics or anything else) just because they aren't *currently* doing it. It also suggests that a long-term perspective needs to replace a short-term one, whether in politics, economics, science, psychotherapy, or health care, for example. Attitudes we take on in any one of these spheres, whether helpful or harmful, tend to carry over into other spheres.

The novelist Aldous Huxley was embarrassed that late in life he had no more advice to offer on improving the human condition than "Try to be a little kinder." This book will have the temerity to suggest, with the help of science, various things that "being a little kinder" means operationally. Since emotions and cognitions are deeply interconnected in the brain, we will see that kindness is shown not just in our feelings but also in our thoughts, our categorizations, and our expectations of people.

These points will be developed not through airtight logical arguments, but through a mixture of rational and intuitive arguments. First they will be related to some age-old philosophical questions.

"The Good I Will I Do Not ..."

As an old limerick reads:

God's plan made a hopeful beginning,
But we spoiled our chances by sinning.
We trust that the story

Will end in God's glory
But at present the other side's winning.

This book isn't committed to any particular theological outlook. The desirable future world outlined at the end of this book includes an argument for the importance of religion, but religion that is flexible in its creed and includes a sense of humor.

Still, most of the world's major religions are somehow concerned with the gap between the human potential for cooperation, and the reality of human life that includes war and cruelty. Even when humans *consciously desire* to be more cooperative and compassionate toward others, they aren't always. Why is this so?

We can ask roughly the same question in several other ways. Why are racism, sexism, and rigid dominance hierarchies, for example, so common when they are such a waste of human potential — and so many people *realize* they are a waste? Why doesn't might always make right, nor right always make might? In the words of St. Paul (Romans 7:19): "The good that I would I do not: but the evil that I would not, I do."

Many modern scientists and secularists are uncomfortable with the words "good" and "evil" because those words carry connotations of judgment by a supernatural or divine power. So this book doesn't use those words much. But I like to interpret Paul's use of "good" and "evil" as metaphors for more naturalistic concepts. "Good" can be interpreted not as angelic or saintly, but as tending to enhance the quality of life of oneself and other human beings. "Evil" can be interpreted not as deliberately malicious, but as tending to lower the quality of life.

But what does *quality* mean? Robert Pirsig in *Zen and the Art of Motorcycle Maintenance* made a valiant effort to define quality philosophically, and concluded that it was primary and undefinable.[25] I suspect, though (based more on personal intuition than on solid evidence) that as we probe into brain organization, we may eventually find a measurable variable somewhere that's analogous to what most of

us think of as quality (without necessarily being a "definition" of quality). Later we will discuss a neural network model of self-actualization that may be a primitive first step toward defining this variable.

Anyway, there are some repeatable patterns to how and why humans fall short of their best potential. Whenever we see regular patterns in nature, we're tempted to try and explain them scientifically. The branches of science likely to be relevant to these problems are neuroscience — the study of brain physiology, anatomy, and chemistry — and psychology. Brain and behavioral science are still at an early stage of grappling with important human problems, but already some underlying principles are starting to be visible. In the past, many neuroscientists have avoided issues of value and motivation as too vague. But now many world-class neuroscientists find that their attempts to understand the brain compel them to consider motivational and emotional issues. Some, like Antonio Damasio and Walter Freeman,[26] are writing popular books on how the brain relates to emotions or values. And neural network modelers — as theoretical bridge builders between neuroscience and psychology — are also getting into the act.

Many writers have used different names for behavior that isn't optimal or self-actualizing. Edgar Allan Poe called it "the imp of the perverse,"[27] meaning by those words the impulse not to do what one knows "deep down" is the right thing to do. In the mental health field, it is called neurotic behavior. In societies or educational systems, aspects of it are described as obeying entrenched rules (compare the earlier examples of Pharaonic Egypt and Ming China).

The overriding question of this book is sometimes described by clinical psychologists as: why don't we always self-actualize? The term *self-actualization* was coined by the psychologist Abraham Maslow to mean achievement of one's full potential. Maslow relates the concept to a fairly definable set of characteristics, which is discussed later.

Though self-actualization is considered an aspect of individual behavior and functioning, this question is analogous to a sociological question dealing with rules of interpersonal conduct. In most interactions between people, some rules are developed (often unconsciously) that govern what actions and words are permissible. (Sometimes close friends interact spontaneously enough that they don't go by rules, but this is the exception.) As in bureaucracies, these rules are developed over time, often arbitrarily and

bearing little relation to anybody's ideas about how people should work together or organizations should function well. Rules are found in every sphere of life, from authority to sexual behavior to casual joking. Sometimes the rules had a human-enhancing purpose at one time but have survived long after anyone even remembers why they were instituted. So why do we develop or preserve "stupid" rules for social interactions, and why do we at the same time see these rules as stupid?

In the field of animal behavior, these questions can be related to some long-standing debates on human biological nature. Is our nature competitive or cooperative, or (as this book suggests) both or neither? Writers emphasizing our competitive side include the popular essayists Robert Ardrey and Desmond Morris who stress the instinct to acquire and hold on to territory, even when food and sexual needs have been met.[28] These authors use the territorial instinct to explain war and social competition. Ardrey and Morris have been shown to be on shaky ground scientifically[29] but their views are partly supported by well-known behavioral biologists like Konrad Lorenz and Edwin Wilson.[30] Wilson is particularly known as one of the founders of *sociobiology*, the school of thought that believes there are genetic predispositions to all prevalent forms of behavior.

Writers emphasizing the cooperative side of our nature, by contrast, include the anthropologist Ashley Montagu, who stressed mutual interdependence as the foundation of society.[31] Montagu saw empathy between people as the driving force that leads to interdependence. Along these same lines, the psychologist Alfie Kohn argued that social bonding has been one of the main directing forces in human evolution.[32] Kohn added that European and American mainstream scholarship has overemphasized the nastier side of human nature. He quoted a statement by twenty psychologists, anthropologists, and animal behaviorists (the Seville Statement of 1986) that the belief in "biological" causes of war is scientifically incorrect. The cooperative theme is also strong in the work of the psychologist Abraham Maslow, who devoted several books to describing the far reaches of human potential and argued that there is a biological drive for self-actualization.[33] Finally, several animal behaviorists have found at least one primate species whose social structure facilitates cooperation and bonding. These are the bonobos, an African species of pygmy chimpanzees.[34]

This book argues scientifically that both sets of writers on human nature are partially correct. I agree with the view (often associated with political liberals) that human potential for good is practically limitless, and institutions, societies, and economies should be reshaped to bring out the best in people. But I also agree with the view (often associated with political conservatives) that war, poverty, and power inequalities can't be blamed on institutions alone. After all, institutions were created *by human beings*! Since unequal hierarchical institutions have been common over the centuries, there must be something in human nature that causes them to keep on being built. But the desire for cooperation and harmony, for loving and caring connections between people, is universal. So there is something else in human nature that suffers from unequal hierarchical institutions and is frustrated by them. Maslow and the anthropologist Ruth Benedict both described this feeling as a craving for *synergy*. Synergy (Greek *syn*=together, *ourgos*=work) is defined roughly as the sense of people working together in relative harmony or "feeling with" each other.[35]

The Brain and Human Nature

What can brain science say about human nature? Our understanding of brain function and the relations of brain to behavior has exploded in the late Twentieth Century. Exciting progress has occurred in all aspects of laboratory neuroscience — neurophysiology, neurochemistry, neuroanatomy, electroencephalography, clinical neuropsychology — and experimental psychology. In addition, the growth of computers has created the new field of brain simulation via neural networks. These advances put partial scientific answers to some ancient philosophical questions within reach for the first time in recorded human history. The answers we get to these questions, we hope, can help us fill the long-lamented gap between our technology and our ethics.

A word of caution must be stated about applying science to philosophical questions. Most scientists working on the brain are ultimately interested in using their work to answer questions about the human mind and human behavior. But when they try to do so, these scientists are often charged with

reductionism. Reductionism is the materialist philosophical outlook that sees the world in "nothing but" terms; for example, it holds that behaviors are "nothing but" interactions among molecules in the nervous system. More recently, however, brain scientists have developed a collective defense that gives their speculations more credibility. The defense is that instead of looking for brain *explanations* of emotions, memory, consciousness, motivation, etc., they are looking for brain *correlates* of those things. The word "correlate" comes from the statistical term "correlation"; it means there is some imprecision, a high probability of association but not necessarily a certainty.* For example, the emotion of pleasure can't be *reduced* to a series of measurable biochemical and physiological events, but there are some measurable biochemical and physiological events repeatedly associated with pleasurable feelings.

The idea of correlation allows us to rephrase our overarching question about self-actualization in two parts, as follows.

> What events in the brain or a model neural network correlate with providing long-term positive reward?
>
> What events in the brain correlate with the "imp of the perverse"? That is, what in our biological nature might lead us toward actions that stand in the way of long-term reward?

Some contemporary philosophers of science are uncomfortable asking these questions. The philosophers Patricia Churchland and Stephen Stich, and the late neuroscientist J. Z. Young, have hinted that the concepts from neurobiology, as they emerge, will replace those from folk psychology.[36] By contrast, my belief is that *some* concepts from folk psychology (the ones I call common nonsense) will be replaced, but others (the ones that are common sense) will pave the way toward finding organizing

* Some scientists, instead of probability, prefer the metaphor of fuzzy mathematics (see Kosko, Bart, *Fuzzy Thinking: The New Science of Fuzzy Logic*, Hyperion, New York, 1993).

principles for understanding the brain. In particular, I reject the extreme reductionist position, that questions of value and emotion will be side-stepped as brain science progresses. Many leading contemporary neuroscientists, such as Antonio Damasio, Joseph LeDoux, and Edmund Rolls, are in fact deeply interested in the neural correlates of emotion and beginning to speculate on how they may relate to values.[37] So are many neural network theorists who develop computer models of brain function.

The state of the art in neuroscience and neural networks hasn't yet advanced to the point of yielding precise answers to the two questions about long-term reward and the imp of the perverse. But we already know enough about the brain to give us many pieces of the puzzle. We know some of the brain areas and chemical neurotransmitters involved in seeking and obtaining rewards. We also know some of the brain areas involved in planning, but little about how that works. But we do know, from experimental psychology as well as neurophysiology, that the short- and long-term decisions we make are heavily influenced by, and influence, how we perceive and categorize events.

So the clues to answering the self-actualization questions will come from our emotions, our cognitive functions, and how emotions and cognitions interact. Since my examples come from daily life as well as laboratories, this book's exposition touches on many disciplines: biology, mathematics, psychology, philosophy, religion, politics, sociology, economics, and literature among others. At times it moves quickly from one field to another; some of its main points, in fact, involve analogies between different fields. This reflects the universal involvement of the brain and mind, which means that theories about mental function apply to everything we do. It also reflects my own belief in the unity of knowledge and impatience with conventional boundaries between fields.

Moreover, parts of this book flow imperceptibly between hazy concepts (e.g., feelings, stereotypes, habits), and more precise ones (e.g., amounts of a chemical substance, mathematical equilibrium states). Other parts flow between academic subjects and mundane life. Some readers may be at first disconcerted by these apparent shifts of emphasis. To that I again invoke the unity of nature, a unity that can be described on many levels at once. At any given moment we understand only parts of the picture precisely. For a complete picture, we need imagination to fill in other parts that are not yet well analyzed. Some parts

of the picture may never be precise by their nature. But in other cases, this year's fuzzy introspection may become next year's laboratory measurement.

Interdisciplinary studies are now growing all over the world, and traditional intellectual boundaries are loosening. We need to accelerate this trend in order to break the stranglehold of specialization[38] and encourage average human beings to participate in the major decisions that affect their lives. One-time French President Georges Clemenceau is reputed to have said that war is too important to leave to the generals. In the same vein, science is too important to leave to the scientists, religion to leave to the clergy, and so forth. So without covering any one traditional academic field in depth, I draw on sources from many disciplines as they relate to the book's major arguments. I will lead the reader on an intellectual adventure, in which many of the most exciting results are still in progress, but enough have been obtained already to give us useful intuitions.

Our incomplete scientific descriptions help us know what's really going on mentally in all our daily actions and interactions. Some of our popular beliefs will prove to be wrong — those are the common nonsense of this book's title. Other popular beliefs will prove right and illuminate how the brain works — those are the book's common sense.

With these questions about behavior, beliefs, and attitudes in mind, we can sketch some of the scientific techniques that may be useful in answering them. The next chapter gives a brief nontechnical overview of neural networks, emphasizing those aspects which are most relevant for behavior and beliefs.

Chapter 1: Beliefs Matter!

1. Merton, 1967.

2. Ibid, p. 477.

3. Mills, 1976.

4. Myrdal, 1962.

5. Rosenthal and Jacobson, 1968.

6. Piaget, 1952.

7. Elman et al., 1996; Ö–men and Prakash, 1997.

8. Snyder and Swann, 1978a; Snyder and Swann, 1978b; Snyder, 1983.

9. The concept of availability is introduced in the psychological literature in Tversky and Kahneman, 1973.

10. Turner, 1991.

11. Durkheim, 1895/1938.

12. Collins, 1975; Giddens, 1984; Habermas, 1984; Mead, 1934.

13. Harris, 1974, 1977.

14. Harris, 1977, Chapter 13.

15. See also Bradley, 1987, for a study of modern urban communes that achieved high morale and coherence through centralized charismatic leadership.

16. Harris, 1974.

17. The process by which emotions are involved in decision making is described in Damasio, 1994. Damasio adds that patients with damage to the orbital part of the frontal lobes lack that emotional involvement. These patients take overly long to make decisions and often get stuck.

18. Tversky and Kahneman, 1974; Tversky and Kahneman, 1981.

19. Tversky and Kahneman, 1981.

20. Grossberg and Gutowski, 1987.

21. Personal communication, from a class at Massachusetts Institute of Technology, Spring, 1972.

22. Damasio, 1994.

23 For a summary of statements about the crisis and opportunity of the new millennium by over sixty religious, political, scientific, and literary leaders, see the January/February 2000 issue of *Tikkun*.

24. Eisler, 1995.

25. Pirsig, 1974.

26. Damasio, 1994; Freeman, 1995.

27. Stedman, 1894, Vol. II, 37-47.

28. Ardrey, 1966; Morris, 1968.

29. Cook, 1975; Kitcher, 1985.

30. Lorenz, 1966; Wilson, 1978.

31. Montagu, 1974.

32. Kohn, 1990.

33. Maslow, 1968, 1972.

32. A summary of the bonobo studies with references is found in Eisler, 1995, 40-48.

35. Benedict, 1970.

36. Churchland, 1986; Stich, 1983; Young, 1951. The belief that folk psychology is altogether unnecessary has been called *eliminative materialism*; see Churchland, 1988, 43-49, for a detailed description of this viewpoint, and Searle, 1992, 58-63, for a refutation of it.

37. Damasio, 1994: LeDoux, 1991; Rolls, 1994.

38. The bad effects of overspecialization on Western culture are described well in Saul, 1992.