INTRODUCTION

The Brain—is wider than the Sky—
For—put them side by side—
The one the other will contain
With ease—and You—beside—
The Brain is deeper than the sea—
For—hold them—Blue to Blue—
The one the other will absorb—
As Sponges—Buckets—do—
The Brain is just the weight of God—
For—Heft them—Pound for Pound—
And they will differ—if they do—
As Syllable from Sound

—Emily Dickinson

As the most important phenomenon in the universe, intelligence is capable of transcending natural limitations, and of transforming the world in its own image. In human hands, our intelligence has enabled us to overcome the restrictions of our biological heritage and to change ourselves in the process. We are the only species that does this.

The story of human intelligence starts with a universe that is capable of encoding information. This was the enabling factor that allowed evolution to take place. How the universe got to be this way is itself an interesting story. The standard model of physics has dozens of constants that need to be precisely what they are, or atoms would not have been possible, and there would have been no stars, no planets, no brains, and no books on brains. That the laws of physics are so precisely tuned to have allowed the evolution of information appears to be incredibly unlikely. Yet by the anthropic principle, we would not be talking about it if it were not the case. Where some people see a divine hand, others see a multiverse spawning an evolution of universes with the boring (non-information-bearing) ones dying out. But regardless of how our universe got to be the way it is, we can start our story with a world based on information.

The story of evolution unfolds with increasing levels of abstraction. Atoms—especially carbon atoms, which can create rich information structures by linking in four different directions—formed increasingly complex molecules. As a result, physics gave rise to chemistry.

A billion years later, a complex molecule called DNA evolved, which could precisely encode lengthy strings of information and generate organisms described by these "programs." As a result, chemistry gave rise to biology.

At an increasingly rapid rate, organisms evolved communication and decision networks called nervous systems, which could coordinate the increasingly complex parts of their bodies as well as the behaviors that facilitated their survival. The neurons making up nervous systems aggregated into brains capable of increasingly intelligent behaviors. In this way, biology gave rise to neurology, as brains were now the cutting edge of storing and manipulating information. Thus we went from atoms to molecules to DNA to brains. The next step was uniquely human.

The mammalian brain has a distinct aptitude not found in any other class of animal. We are capable of *hierarchical* thinking, of understanding a structure composed of diverse elements arranged in a pattern, repre-

HowToCreateAMind.com

senting that arrangement with a symbol, and then using that symbol as an element in a yet more elaborate configuration. This capability takes place in a brain structure called the neocortex, which in humans has achieved a threshold of sophistication and capacity such that we are able to call these patterns *ideas*. Through an unending recursive process we are capable of building ideas that are ever more complex. We call this vast array of recursively linked ideas *knowledge*. Only *Homo sapiens* have a knowledge base that itself evolves, grows exponentially, and is passed down from one generation to another.

Our brains gave rise to yet another level of abstraction, in that we have used the intelligence of our brains plus one other enabling factor, an opposable appendage—the thumb—to manipulate the environment to build tools. These tools represented a new form of evolution, as neurology gave rise to technology. It is only because of our tools that our knowledge base has been able to grow without limit.

Our first invention was the story: spoken language that enabled us to represent ideas with distinct utterances. With the subsequent invention of written language we developed distinct shapes to symbolize our ideas. Libraries of written language vastly extended the ability of our unaided brains to retain and extend our knowledge base of recursively structured ideas.

There is some debate as to whether other species, such as chimpanzees, have the ability to express hierarchical ideas in language. Chimps are capable of learning a limited set of sign language symbols, which they can use to communicate with human trainers. It is clear, however, that there are distinct limits to the complexity of the knowledge structures with which chimps are capable of dealing. The sentences that they can express are limited to specific simple noun-verb sequences and are not capable of the indefinite expansion of complexity characteristic of humans. For an entertaining example of the complexity of human-generated language, just read one of the spectacular multipage-length sentences in a Gabriel García Márquez story or novel—his six-page story "The Last Voyage of the Ghost"

HowToCreateAMind.com

is a single sentence and works quite well in both Spanish and the English translation.[1]

The primary idea in my three previous books on technology (*The Age of Intelligent Machines*, written in the 1980s and published in 1989; *The Age of Spiritual Machines*, written in the mid- to late 1990s and published in 1999; and *The Singularity Is Near*, written in the early 2000s and published in 2005) is that an evolutionary process inherently accelerates (as a result of its increasing levels of abstraction) and that its products grow exponentially in complexity and capability. I call this phenomenon the law of accelerating returns (LOAR), and it pertains to both biological and technological evolution. The most dramatic example of the LOAR is the remarkably predictable exponential growth in the capacity and price/performance of information technologies. The evolutionary process of technology led invariably to the computer, which has in turn enabled a vast expansion of our knowledge base, permitting extensive links from one area of knowledge to another. The Web is itself a powerful and apt example of the ability of a hierarchical system to encompass a vast array of knowledge while preserving its inherent structure. The world itself is inherently hierarchical—trees contain branches; branches contain leaves; leaves contain veins. Buildings contain floors; floors contain rooms; rooms contain doorways, windows, walls, and floors.

We have also developed tools that are now enabling us to understand our own biology in precise information terms. We are rapidly reverse-engineering the information processes that underlie biology, including that of our brains. We now possess the object code of life in the form of the human genome, an achievement that was itself an outstanding example of exponential growth, in that the amount of genetic data the world has sequenced has approximately doubled every year for the past twenty years.[2] We now have the ability to simulate on computers how sequences of base pairs give rise to sequences of amino acids that fold up into three-dimensional proteins, from which all of biology is constructed. The complexity of proteins for which we can simulate protein folding has been

HowToCreateAMind.com

steadily increasing as computational resources continue to grow exponentially.[3] We can also simulate how proteins interact with one another in an intricate three-dimensional dance of atomic forces. Our growing understanding of biology is one important facet of discovering the intelligent secrets that evolution has bestowed on us and then using these biologically inspired paradigms to create ever more intelligent technology.

There is now a grand project under way involving many thousands of scientists and engineers working to understand the best example we have of an intelligent process: the human brain. It is arguably the most important effort in the history of the human-machine civilization. In *The Singularity Is Near* I made the case that one corollary of the law of accelerating returns is that other intelligent species are likely not to exist. To summarize the argument, if they existed we would have noticed them, given the relatively brief time that elapses between a civilization's possessing crude technology (consider that in 1850 the fastest way to send nationwide information was the Pony Express) to its possessing technology that can transcend its own planet.[4] From this perspective, reverse-engineering the human brain may be regarded as the most important project in the universe.

The goal of the project is to understand precisely how the human brain works, and then to use these revealed methods to better understand ourselves, to fix the brain when needed, and—most relevant to the subject of this book—to create even more intelligent machines. Keep in mind that greatly amplifying a natural phenomenon is precisely what engineering is capable of doing. As an example, consider the rather subtle phenomenon of Bernoulli's principle, which states that there is slightly less air pressure over a moving curved surface than over a moving flat one. The mathematics of how Bernoulli's principle produces wing lift is still not yet fully settled among scientists, yet engineering has taken this delicate insight, focused its powers, and created the entire world of aviation.

In this book I present a thesis I call the pattern recognition theory of mind (PRTM), which, I argue, describes the basic algorithm of the

HowToCreateAMind.com

neocortex (the region of the brain responsible for perception, memory, and critical thinking). In the chapters ahead I describe how recent neuroscience research, as well as our own thought experiments, leads to the inescapable conclusion that this method is used consistently across the neocortex. The implication of the PRTM combined with the LOAR is that we will be able to engineer these principles to vastly extend the powers of our own intelligence.

Indeed this process is already well under way. There are hundreds of tasks and activities formerly the sole province of human intelligence that can now be conducted by computers, usually with greater precision and at a vastly greater scale. Every time you send an e-mail or connect a cell phone call, intelligent algorithms optimally route the information. Obtain an electrocardiogram, and it comes back with a computer diagnosis that rivals that of doctors. The same is true for blood cell images. Intelligent algorithms automatically detect credit card fraud, fly and land airplanes, guide intelligent weapons systems, help design products with intelligent computer-aided design, keep track of just-in-time inventory levels, assemble products in robotic factories, and play games such as chess and even the subtle game of Go at master levels.

Millions of people witnessed the IBM computer named Watson play the natural-language game of *Jeopardy!* and obtain a higher score than the best two human players in the world combined. It should be noted that not only did Watson read and "understand" the subtle language in the *Jeopardy!* query (which includes such phenomena as puns and metaphors), but it obtained the knowledge it needed to come up with a response from understanding hundreds of millions of pages of natural-language documents including Wikipedia and other encyclopedias on its own. It needed to master virtually every area of human intellectual endeavor, including history, science, literature, the arts, culture, and more. IBM is now working with Nuance Speech Technologies (formerly Kurzweil Computer Products, my first company) on a new version of Watson that will read medical literature (essentially all medical journals and leading medical blogs) to become

HowToCreateAMind.com

a master diagnostician and medical consultant, using Nuance's clinical language–understanding technologies. Some observers have argued that Watson does not really "understand" the *Jeopardy!* queries or the encyclopedias it has read because it is just engaging in "statistical analysis." A key point I will describe here is that the mathematical techniques that have evolved in the field of artificial intelligence (such as those used in Watson and Siri, the iPhone assistant) are mathematically very similar to the methods that biology evolved in the form of the neocortex. If understanding language and other phenomena through statistical analysis does not count as true understanding, then humans have no understanding either.

Watson's ability to intelligently master the knowledge in natural-language documents is coming to a search engine near you, and soon. People are already talking to their phones in natural language (via Siri, for example, which was also contributed to by Nuance). These natural-language assistants will rapidly become more intelligent as they utilize more of the Watson-like methods and as Watson itself continues to improve.

The Google self-driving cars have logged 200,000 miles in the busy cities and towns of California (a figure that will undoubtedly be much higher by the time this book hits the real and virtual shelves). There are many other examples of artificial intelligence in today's world, and a great deal more on the horizon.

As further examples of the LOAR, the spatial resolution of brain scanning and the amount of data we are gathering on the brain are doubling every year. We are also demonstrating that we can turn this data into working models and simulations of brain regions. We have succeeded in reverse-engineering key functions of the auditory cortex, where we process information about sound; the visual cortex, where we process information from our sight; and the cerebellum, where we do a portion of our skill formation (such as catching a fly ball).

The cutting edge of the project to understand, model, and simulate the human brain is to reverse-engineer the cerebral neocortex, where we do

HowToCreateAMind.com

our recursive hierarchical thinking. The cerebral cortex, which accounts for 80 percent of the human brain, is composed of a highly repetitive structure, allowing humans to create arbitrarily complex structures of ideas.

In the pattern recognition theory of mind, I describe a model of how the human brain achieves this critical capability using a very clever structure designed by biological evolution. There are details in this cortical mechanism that we do not yet fully understand, but we know enough about the functions it needs to perform that we can nonetheless design algorithms that accomplish the same purpose. By beginning to understand the neocortex, we are now in a position to greatly amplify its powers, just as the world of aviation has vastly amplified the powers of Bernoulli's principle. The operating principle of the neocortex is arguably the most important idea in the world, as it is capable of representing all knowledge and skills as well as creating new knowledge. It is the neocortex, after all, that has been responsible for every novel, every song, every painting, every scientific discovery, and the multifarious other products of human thought.

There is a great need in the field of neuroscience for a theory that ties together the extremely disparate and extensive observations that are being reported on a daily basis. A unified theory is a crucial requirement in every major area of science. In chapter 1 I'll describe how two daydreamers unified biology and physics, fields that had previously seemed hopelessly disordered and varied, and then address how such a theory can be applied to the landscape of the brain.

Today we often encounter great celebrations of the complexity of the human brain. Google returns some 30 million links for a search request for quotations on that topic. (It is impossible to translate this into the number of actual quotations it is returning, however, as some of the Web sites linked have multiple quotes, and some have none.) James D. Watson himself wrote in 1992 that "the brain is the last and grandest biological frontier, the most complex thing we have yet discovered in our universe." He goes on to explain why he believes that "it contains hundreds of billions of

HowToCreateAMind.com

cells interlinked through trillions of connections. The brain boggles the mind."[5]

I agree with Watson's sentiment about the brain's being the grandest biological frontier, but the fact that it contains many billions of cells and trillions of connections does not necessarily make its primary method complex if we can identify readily understandable (and re-creatable) patterns in those cells and connections, especially massively redundant ones.

Let's think about what it means to be complex. We might ask, is a forest complex? The answer depends on the perspective you choose to take. You could note that there are many thousands of trees in the forest and that each one is different. You could then go on to note that each tree has thousands of branches and that each branch is completely different. Then you could proceed to describe the convoluted vagaries of a single branch. Your conclusion might be that the forest has a complexity beyond our wildest imagination.

But such an approach would literally be a failure to see the forest for the trees. Certainly there is a great deal of fractal variation among trees and branches, but to correctly understand the principles of a forest you would do better to start by identifying the distinct patterns of redundancy with stochastic (that is, random) variation that are found there. It would be fair to say that the concept of a forest is simpler than the concept of a tree.

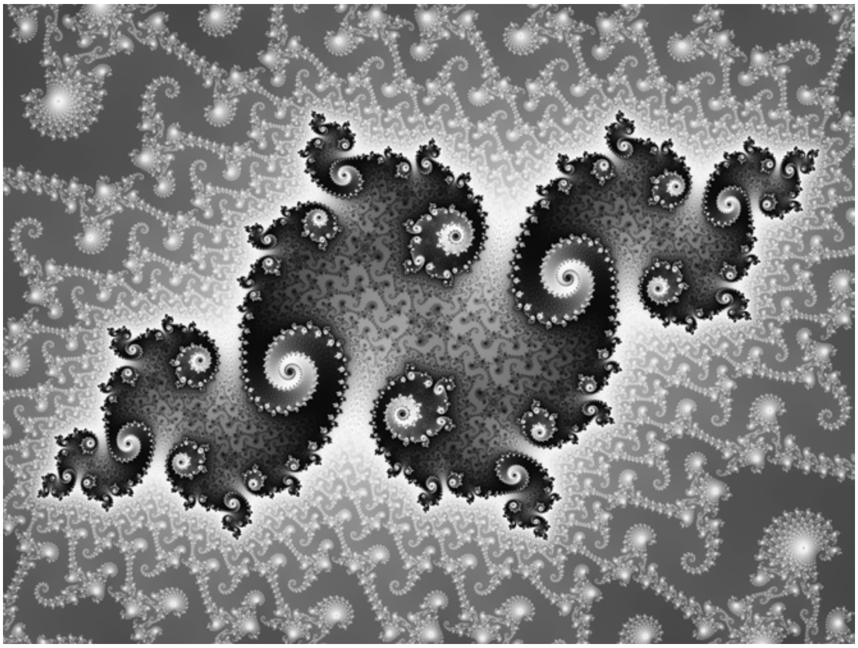
Thus it is with the brain, which has a similar enormous redundancy, especially in the neocortex. As I will describe in this book, it would be fair to say that there is more complexity in a single neuron than in the overall structure of the neocortex.

My goal in this book is definitely not to add another quotation to the millions that already exist attesting to how complex the brain is, but rather to impress you with the power of its simplicity. I will do so by describing how a basic ingenious mechanism for recognizing, remembering, and predicting a pattern, repeated in the neocortex hundreds of millions of times, accounts for the great diversity of our thinking. Just as an astonishing diversity of organisms arises from the different combinations of the values

HowToCreateAMind.com

of the genetic code found in nuclear and mitochondrial DNA, so too does an astounding array of ideas, thoughts, and skills form based on the values of the patterns (of connections and synaptic strengths) found in and between our neocortical pattern recognizers. As MIT neuroscientist Sebastian Seung says, "Identity lies not in our genes, but in the connections between our brain cells."[6]

We need to distinguish between true complexity of design and apparent complexity. Consider the famous Mandelbrot set, the image of which has long been a symbol of complexity. To appreciate its apparent complication, it is useful to zoom in on its image (which you can access via the links in this endnote).[7] There is endless intricacy within intricacy, and they are always different. Yet the design—the formula—for the Mandelbrot set couldn't be simpler. It is six characters long: $Z = Z^2 + C$, in which Z is a "complex" number (meaning a pair of numbers) and C is a constant. It is



One view of the display of the Mandelbrot set, a simple formula that is iteratively applied. As one zooms in on the display, the images constantly change in apparently complex ways.

not necessary to fully understand the Mandelbrot function to see that it is simple. This formula is applied iteratively and at every level of a hierarchy. The same is true of the brain. Its repeating structure is not as simple as that of the six-character formula of the Mandelbrot set, but it is not nearly as complex as the millions of quotations on the brain's complexity would suggest. This neocortical design is repeated over and over at every level of the conceptual hierarchy represented by the neocortex. Einstein articulated my goals in this book well when he said that "any intelligent fool can make things bigger and more complex . . . but it takes . . . a lot of courage to move in the opposite direction."

So far I have been talking about the brain. But what about the mind? For example, how does a problem-solving neocortex attain consciousness? And while we're on the subject, just how many conscious minds do we have in our brain? There is evidence that suggests there may be more than one.

Another pertinent question about the mind is, what is free will, and do we have it? There are experiments that appear to show that we start implementing our decisions before we are even aware that we have made them. Does that imply that free will is an illusion?

Finally, what attributes of our brain are responsible for forming our identity? Am I the same person I was six months ago? Clearly I am not exactly the same as I was then, but do I have the same identity?

We'll review what the pattern recognition theory of mind implies about these age-old questions.