PENGUIN BOOKS
Published by the Penguin Group
Penguin Group (USA) Inc., 375 Hudson Street, New York, New York 10014, U.S.A. • Penguin Group (Canada), 90 Eglinton Avenue East, Suite 700, Toronto, Ontario, Canada M4P 2Y3 (a division of Pearson Penguin Canada Inc.) • Penguin Books Ltd, 80 Strand, London WC2R 0RL, England • Penguin Ireland, 25 St. Stephen's Green, Dublin 2, Ireland (a division of Penguin Books Ltd) • Penguin Group (Australia), 250 Camberwell Road, Camberwell, Victoria 3124, Australia (a division of Pearson Australia Group Pty Ltd) • Penguin Books India Pvt Ltd, 11 Community Centre, Panchsheel Park, New Delhi–110 017, India • Penguin Group (NZ), cnr Airborne and Rosedale Roads, Albany, Auckland 1310, New Zealand (a division of Pearson New Zealand Ltd) • Penguin Books (South Africa) (Pty) Ltd, 24 Sturdee Avenue, Rosebank, Johannesburg 2196, South Africa

Penguin Books Ltd, Registered Offices: 80 Strand, London WC2R 0RL, England

First published in the United States of America by Viking Penguin, a member of Penguin Group (USA) Inc. 2006
Published in Penguin Books 2007

ISBN: 978-1-1012-0127-5

CIP data available

CONTENTS

INTRODUCTION

Everything is made of one hidden stuff.

—Ralph Waldo Emerson

Civilization is doomed.

That's probably not the first thing you want to read when you pick up a book, but it's true. Humanity—and all life in the universe—is going to be wiped out. No matter how advanced our civilization becomes, no matter if we develop the technology to hop from star to star or live for six hundred years, there is only a finite time left before the last living creature in the visible universe will be snuffed out. The laws of information have sealed our fate, just as they have sealed the fate of the universe itself.

The word *information* conjures visions of computers and hard drives and Internet superhighways; after all, the introduction and popularization of computers came to be known as the information revolution. However, computer science is only a very small aspect of an overarching idea known as information theory. While this theory does, in fact, dictate how computers work, it does much, much more than that. It governs the behavior of objects on many different scales. It tells how atoms interact with each other and how black holes swallow stars. Its rules describe how the universe will die, and they illuminate the structure of the entire cosmos.

Even if there were no such thing as a computer, information theory would still be the third great revolution of twentieth-century physics.

The laws of thermodynamics—the rules that govern the motion of atoms in a chunk of matter—are, underneath it all, laws about information. The theory of relativity, which describes how objects behave at extreme speeds and under the strong influence of gravity, is actually a theory of information. Quantum theory, which governs the realm of the very small, is a theory of information as well. The concept of information, which is far broader than the mere content of a hard drive, ties together all these theories into one incredibly potent idea.

Information theory is so powerful because information is physical. Information is not just an abstract concept, and it is not just facts or figures, dates or names. It is a concrete property of matter and energy that is quantifiable and measurable. It is every bit as real as the weight of a chunk of lead or the energy stored in an atomic warhead, and just like mass and energy, information is subject to a set of physical laws that dictate how it can behave—how information can be manipulated, transferred, duplicated, erased, or destroyed. And everything in the universe must obey the laws of information, because everything in the universe is shaped by the information it contains.

The idea of information was born from the ancient art of codemaking and codebreaking. The ciphers that hid state secrets were, in fact, methods of obscuring information and transporting

it from place to place. When the art of code cracking was combined with the science of thermodynamics—the branch of physics that describes the behavior of engines, the exchange of heat, and the production of work—information theory was the result. This new theory of information was an idea as revolutionary as quantum theory and relativity; it instantly transformed the field of communications and paved the way to the computer age, but that was just the beginning. Within a decade, physicists and biologists began to understand that the ideas of information theory govern much more than the bits and bytes of computers and codes and communications: they describe the behavior of the subatomic world, all life on Earth, and even the universe as a whole.

Each creature on Earth is a creature of information; information sits at the center of our cells, and information rattles around in our brains. But it's not just living beings that manipulate and process information. Every particle in the universe, every electron, every atom, every particle not yet discovered, is packed with information—information that is often inaccessible to us, but information nonetheless, information that can be transferred, processed, and dissipated. Each star in the universe, each one of the countless galaxies in the heavens, is packed full of information, information that can escape and travel. That information is always flowing, moving from place to place, spreading throughout the cosmos.

Information appears, quite literally, to shape our universe. The motion of information may well determine the physical structure of the cosmos. And information seems to be at the heart of the

deepest paradoxes in science—the mysteries of relativity and quantum mechanics, the origin and fate of life in the universe, the nature of the ultimate destructive power of the black hole, and the hidden order in a seemingly random cosmos.

The laws of information are beginning to reveal the answers to some of the most profound questions of science, but the answers are, in some ways, more disturbing and more bizarre than the paradoxes they solve. Information leads to a picture of a universe speeding toward its own demise, of living creatures as slaves to parasites within, and of an incredibly byzantine cosmos made up of an enormous collection of parallel universes.

The laws of information are giving physicists a way to understand the darkest mysteries that humanity has ever pondered. Yet those laws are painting a picture that is as grim as it is surreal.

REDUNDANCY

Gentlemen don't read other gentlemen's mail!

—Henry L. Stimson

"AF is short of water." These five words sank the Japanese fleet.

In the spring of 1942, the U.S. military was reeling from an unbroken series of defeats. The Japanese navy was supreme in the Pacific, and it was pushing ever closer to American territories. Though the situation was dire, the war was not lost. And U.S. cryptanalysts were about to use a weapon as important as bombs and guns: information.

U.S. codebreakers had cracked JN-25, a cipher used by the Japanese navy. It was a tough code to break, but by May cryptanalysts had completely pried open the mathematical vault of the cipher and revealed the information hidden within.

According to the intercepted and decrypted messages, an American base, code-named *AF,* was shortly to be the object of a major naval assault. American analysts knew that AF was an island in the Pacific (quite possibly Midway Island), but they didn't know

for certain which one it was. If the analysts guessed wrong, the navy would defend the incorrect island, and the enemy would be able to invade the true target unopposed. But if they could figure out which island AF really was and anticipate the destination of Japan's armada, the Americans could concentrate their fleet and maul the invading force. Everything—the war in the Pacific—hinged on one missing piece of information: Where was AF?

Commander Joseph Rochefort, head of the navy's cryptography center at Pearl Harbor, came up with a scheme to get that one last piece of information. He ordered the base at Midway to transmit a phony request for help. The transmission stated that the water distillery on Midway Island had been damaged and the base was nearly out of freshwater. The Japanese, who were eavesdropping on Midway's transmissions, heard the broadcast, too. This is precisely what Rochefort was counting on. Not long after the phony message, Navy Intelligence picked up the faint signals of a Japanese transmission on the airwaves: "AF is short of water." Rochefort had his last bit of information. AF was Midway.

The U.S. fleet gathered to defend the island. On June 4, 1942, Admiral Isoroku Yamamoto's invading force ran directly into Admiral Chester Nimitz's waiting fleet. During the battle, four Japanese aircraft carriers—the *Hiryu, Soryu, Akagi,* and *Kaga*—went to the bottom; in return, only one U.S. carrier was lost. The crippled Japanese fleet steamed home. Japan had lost the battle—and the war in the Pacific. The Japanese navy never again seriously threatened American territory, and the United States began the long, difficult drive to the Japanese homeland. A priceless piece of

information, the target of Yamamoto's invasion, leaked through the protection of codes and ciphers and gave America its crucial victory.[1]

World War II was the first information war. As U.S. cryptographers extracted information from the Japanese JN-25 and Purple ciphers, an elite group of British and Polish codebreakers unraveled Germany's (supposedly) uncrackable Enigma cipher. And just as information allowed the United States to defeat Japan, the Enigma information gave the Allies a way to defeat the Nazi U-boats that were choking Great Britain.

Just as the struggle over information left its imprint on the face of the war, the war left its imprint on the face of information. During World War II, cryptography began to change from an art to a science. The codebreakers in the sweaty code rooms in Hawaii and on a quaint estate in England would be the heralds of a revolution known as information theory.

Codemaking and codebreaking were always closely related to what would become the theory of information. However, for millennia, cryptographers and cryptanalysts had no idea that they were making tentative forays into an entirely new field of science. After all, encryption is older than science. Over and over again, since antiquity, monarchs and generals have relied upon the information hidden by the fragile security of a cipher or a hidden message—awkward attempts to circumvent the dangers of information transfer.

Codemaking goes back to the dawn of Western civilization. In 480 BC, ancient Greece was nearly conquered by the much stronger Persian Empire, but a secret message, hidden under the wax of a writing tablet, warned of an impending invasion. Alarmed at the message, the Greeks immediately began preparing for war. The forewarned Greeks roundly defeated the Persians at the battle of Salamis, ending the Persian threat and ushering in the golden age of Greece. But for that hidden message, the fragile collection of Greek city-states would not have been able to resist the much more powerful Persian navy; Greece would likely have become a Persian conquest, and Western civilization would have turned out quite differently.

Sometimes, a failed attempt to transfer information changes history, too. Heads have literally rolled because a secret message or a cipher has been discovered and decrypted. In 1587, Mary, Queen of Scots, went to the executioner's block because of a bad code. Mary, in prison, was forming a conspiracy to murder Queen Elizabeth and seize the English throne. But since all objects that went in and out of the prison were inspected, Mary had to resort to cryptography to keep in touch with her supporters. She and her coconspirators devised a code and traded little enciphered messages hidden in the bungs of beer barrels. Unfortunately for Mary, Sir Francis Walsingham, England's spymaster, discovered the messages and had them deciphered. He even planted a fake message from Mary to the conspirators, inducing the traitors to reveal the names of all the men in their cabal. When Queen Mary stood trial for treason, the messages were the prime exhibit. A

broken code—and two strokes of an axe—sealed her fate.

Codes and ciphers have many different forms, but they all have the same purpose: to transfer information from one person to another. At the same time, they must be secure; they must prevent an eavesdropper from getting that information if the message is intercepted.

Through most of history, codes weren't terribly secure. A smart codebreaker could unravel even the most sophisticated code with just a little bit of concentration; even so, monarchs and generals had to rely upon these rickety codes. Often, an intercepted and decrypted message meant death or defeat. Sending sensitive messages was always dangerous, but it was a necessary risk to take and a fundamental part of the business of diplomacy and war.

No matter how cryptographers fiddle with words or symbols or numbers or codebooks, no matter how cleverly they hide the messages in bungholes or pumpkins or within poems, there is an unavoidable risk of discovery as crucial information moves from place to place. Just as generals must move troops and arms and supplies from home to the front and back again, so too must they transfer information. And, in its way, information is every bit as palpable as the weight of a bullet, every bit as tangible as the heft of an artillery shell—and every bit as vulnerable as a freighter full of ammunition.

This fundamental property is the hardest thing to accept about information: information is as real and concrete as mass, energy,

or temperature. You cannot see any of these properties directly, but you accept them as real. Information is just as real. It can be measured and manipulated just as the weight of an apple can be gauged with a scale or redistributed with a knife. This is why leaders, generals, and diplomats always took such risks with rickety ciphers. Information must travel from the sender to the receiver just as a hunk of gold bullion would have to travel from Fort Knox to the Mint. There's no magical way of transmitting the information instantly, just as there's no way to teleport the gold directly from vault to vault. Even the most advanced computers must find a way to transfer information from place to place—it can go over a telephone line or through a coaxial cable or even through the air via a wireless connection—but if you want to transfer information from computer to computer, it must travel physically somehow from one computer to another.

Because the information in an object is a concrete, measurable property like mass, this means that information can be misplaced or stolen in the same way mass can be. Just as someone who wishes to move gold from one place to another must brave the risks of highwaymen or thieves, a leader who wants to exchange information must brave the risks of interception and decryption. Information, like gold, must be moved around in order to have any value to humans.

Underneath all the cloak-and-dagger frillery, good codemakers and codebreakers are experts in manipulating information. A cryptographer designing a cipher is trying to ensure that information gets from a sender to a receiver without allowing

anyone else to access that information; the information must not "leak" out of the encrypted message. Conversely, a codebreaker who intercepts an enemy's message is trying to extract information from a seemingly meaningless jumble of letters or symbols. This can only work if the cipher is imperfect—if information leaks out despite the codemaker's best efforts. But not even the best codemaker can make a message miraculously appear at the place it is needed; it must be transported. That's where it is most at risk of discovery.

This idea that something as seemingly abstract as information is actually measurable—and tangible—is one of the central tenets of information theory. This theory was born in the years right after World War II, when mathematicians laid out a set of rules that defined information and described its behavior. This theory has a mathematical certainty that is seldom seen in the sloppy, experimental world of science; its tenets are as inviolable as the laws of thermodynamics that prevent inventors from building a perpetual motion machine. Even though information had been around for centuries, it was only during World War II that cryptographers began to feel around the edges of information theory.

The science of cryptography holds the first clues to the nature of information. It will not yield the full story, but it will give an idea of how information is real and measurable and must be carried from place to place like a brick of gold. For one of the banes of a

cryptographer—redundancy—is closely related to the concept of information, and understanding redundancy can help explain why information can be as palpable as an atom in a chunk of matter.

Whenever you receive a message, even something as simple as "the sky is blue," you must take the series of words and process them to understand the meaning of the message. You receive a series of marks on paper (or sounds in the air) and extract the meaning encoded in those marks. Your brain takes the raw set of lines and curves that spell out "the sky is blue" and manipulates those symbols until it understands that the message is a statement about the color of the heavens outside. This process, this extraction of meaning from a set of symbols, is an unconscious one. It's just something a human brain has been training to do from the very moment that parents make goo-goo noises at a baby in the cradle; the process of becoming fluent in a language is, in a sense, nothing more than learning how to pull meaning from symbols. However, this unconscious process—taking a stream of symbols and extracting meaning from it—is crucial to our ability to use language. And so is the concept of redundancy, because it is redundancy that makes a message easy to understand.

Redundancy is the extra clues in a sentence or a message that allow the meaning to be understood even when the message is somewhat garbled. As it turns out, every sentence in any language is highly redundant. A sentence of English—or of any other language—always has more information than you need to decipher it. This redundancy is easy to see. J-st tr-t-r--d th-s s-nt-nc-. The previous sentence was extremely garbled; all the vowels in the

message were removed.[2] However, it was still easy to decipher it and extract its meaning. The meaning of a message can remain unchanged even though parts of it are removed. This is the essence of redundancy.

To humans, redundancy is a good thing, because it makes a message easier to comprehend, even when the message is partially scrambled by the environment. You can still understand a friend speaking in a crowded restaurant or talking on a staticky cell phone because of redundancy. Redundancy is a safety mechanism; it makes sure that a message gets through even if it gets slightly damaged in transit. All languages have these built-in safety nets composed of patterns and structures and sets of rules that make them redundant. You aren't usually aware of those rules, but your brain unconsciously uses them as you read, speak, listen, and write—anytime you are receiving a message from somebody in a natural language. Even though these rules aren't obvious, they are there nonetheless, and you can feel their influence if you play around with language a little bit.

Consider, for example, the nonsense word *fingry*. *Fingry* sounds as if it could be an English word. In fact, it sounds like an adjective. ("Gee, Bob, your boss looks like he's mighty fingry today.") But what if I create another nonsense word: *trzeci*? Unlike *fingry*, *trzeci* doesn't sound like a valid English word at all.[3] This is because of these implicit rules—in this case, rules specific to the English language. The letter *z* is fairly rare in English and never follows the letters *tr*. Furthermore, it's pretty uncommon to end a word with *i*, so *trzeci* doesn't feel like a real English word—it breaks the

unwritten rules about the attributes of valid English words. *Fingry*, on the other hand, has the right pattern of letters (and sounds) to make it seem like a real English word, and the ending *-gry* tends to signal that the word is an adjective.

The human brain automatically learns these rules and uses them to do a validity check on all the messages that it receives. This is how we distinguish a message with meaning from a meaningless string of symbols or syllables.

All languages have rules within rules within rules. The *trzeci* versus *fingry* rules operate on the level of letters and sounds; they determine what letters and sounds are likely to follow others. But lots of other rules operate on different levels as well. Though they all function unconsciously, you can sense them when something is wrong with a message, because they automatically raise an alarm. For example, there are rules that determine what words are likely to follow other words and phrases; your brain, continuously monitoring the language rules, lets you know if words order are wrong used in the. There are also rules that check on the meaning of a message as you process it. Even a perfectly valid sentence will sound strange if it is not precisely what your brain expects. When this happens, an ill-chosen word can stick out like a sore earlobe.[4]

These rules are everywhere. They tell you the difference between a meaningless grunt and a meaningful consonant, between a nonsense word and a real one, or a silly sentence and one full of information. Some rules are valid across many human languages; there are only a handful of sounds that are allowed to

carry meaning in human speech. Some rules are more specific to individual languages; Polish words look and sound very different from English ones because the corresponding "valid word" rules are very different. But all languages have a vast set of these rules, and it is this set of rules that gives a language its structure—and its redundancy.

When your brain raises an alarm about a broken rule, a word that does not sound English or a sentence that has the wrong word, it is telling you that the stream of letters (or sounds) you are receiving doesn't meet its expectation for a valid message. Something is out of place; something is garbled. By using these rules and working backward, your brain can often correct the problem, such as when a word is missspeledd. Without missing a beat, your brain applied the proper rules of spelling and corrected the garbled stream of symbols. You extracted the meaning of the sentence despite an error. This is nothing more than redundancy in action.

These rules are also what allowed you to read the no-vowels sentence. The implicit rules of the English language instantly told you that "th-s" was probably *this* rather than *thms* or even *thes.* Thanks to the rules, you can still extract the meaning of a message even if I pare away bits of sentences...so long as I haven't removed too much. But there is a point beyond which a sentence can no longer be garbled or compressed without losing comprehensibility. Pare away too many letters and you begin to lose the meaning in the message. When you get rid of all the redundancy in a string of letters, what is left is a concrete, measurable, incompressible

nucleus. That is information: the central, irreducible something that sits at the heart of every sentence.

This is a rough definition, and it's not a complete one, but it's accurate. Information and redundancy are complementary; when you remove the redundancy from a string of letters, or symbols for that matter, what's left is information. Computer scientists are well aware of this irreducible nub in every message. It is important when writing, say, a program for compressing computer files. Compression programs squash files—such as the ones that contain the text of this book—so that they take less room on a hard drive or similar storage devices. These programs are extremely good, but there is little mystery about how they do their job: they work by removing (almost) all of the redundancy from a file, leaving the nub of a file behind. A standard commercial compression program might take a text file and squash it by more than 60 percent. But what's left over is incompressible. Run that compression program again and the file won't squash down any further. (Try it yourself!) It can't be made any smaller unless you are willing to lose some of the meaning of the message, some of the information in the text file. If someone tries to sell you a program that can make such incompressible nubs even smaller, call the FBI to report a case of fraud.

Computer scientists aren't the only people concerned with redundancy. A key challenge of cryptography is to remove or mask the redundancy in a message while retaining that essential information at its heart. No matter how the cryptographers or computer scientists try to mask or shrink a message, though,

there's still an incompressible chunk that must travel from the sender to the recipient, whether the message is transmitted by radio or by wax tablet or by lights in the steeple of the Old North Church. This realization would revolutionize the field of physics. But first, information and redundancy revolutionized the field of cryptography and changed the course of world history.

Modern cryptographers think of their craft in terms of information and redundancy. The goal of a cryptographer, after all, is to generate a stream of symbols that is meaningful to the intended recipient—in a sense, the cryptographer is creating an artificial language. Unlike ordinary human languages, which are intended to share information freely, the cryptographer's cipher is intended to be meaningless to an eavesdropper. The information in the original message is still there in the encrypted version; however, it is hidden to those who don't know how to decipher the message. A good cipher shields information from those not authorized to understand it. A bad cipher lets the information leak through. Often, when a cipher fails, it fails because of clumsy redundancy.

You already know this if you are an amateur codebreaker. On the comics pages of many newspapers, you will find a little puzzle known as a cryptogram. It's usually a famous quotation that has been encrypted in a very straightforward way: each letter is substituted for another letter of the alphabet, yielding a string of nonsense. For example, you might see something like FUDK DK V

NTPVFDOTPM KDIAPT GSHDJX KGUTIT. DF KUSYPH JSF FVWT IYGU FDIT FS ZNTVW DF. With a little practice, you can quickly decipher this sort of puzzle and extract the information it contains.

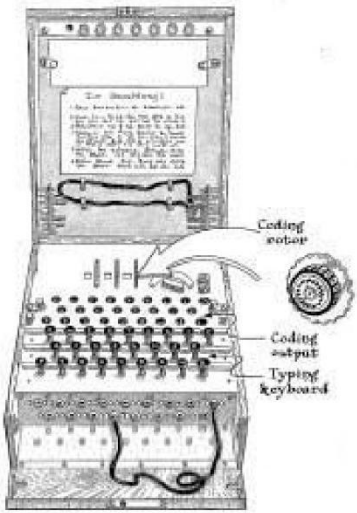
There are several ways to decode a cryptogram, and they all exploit the unwritten rules of English. Even though the information is garbled, these rules allow you to figure out what the message is. One of the rules is that if you have a letter standing alone, it's either an *A* or an *I;* no other single letter forms a valid word. Therefore, in the above cryptogram, the symbol V must represent either the letter *A* or the letter *I.* Another rule is that *E* tends to be the most frequent letter in the English language, so in the above sentence, the most frequent symbol—T—probably represents the letter *E.* Some other letters, such as *S,* and combinations of letters, such as *TH,* are relatively common, so they are almost certain to appear in a given message, while others such as *X* or *KL* are quite rare and may well be missing from a typical cryptogram. Stare at the cryptogram and play around with it for a little while and you will soon be able to decipher the message. The rules of English allow you to extract the information from the message even though it's been hidden. In other words, these rules give the message redundancy and allow you to break the cipher.[5]

Redundancy, the collection of those patterns and rules, is the enemy of a secure code; it helps information leak through, and codemakers go through a great deal of effort to try to hide the redundancy in a message. This is the only way that a codemaker can have hope that a new cipher *might* be secure. Understanding

this relationship among redundancy, information, and security is a cornerstone of cryptography, but before information theory was born, nobody really had a deep understanding of what lay beneath that relationship. Nobody understood the nature of information or redundancy. Nobody had a formal method to define them or measure them or manipulate them. As a result, even the most sophisticated coding schemes of the early twentieth century tended to be insecure. Even those that were thought to be uncrackable.

In February 1918, the German inventor Arthur Scherbius filed a patent for an "unbreakable" cipher machine that would soon become infamous the world over: Enigma.

Enigma was an ingenious way of encrypting a message. It was so complex that most contemporary cryptographers and mathematicians thought that it was hopeless to even attempt to break it.



An Enigma machine

Scherbius's machine looked somewhat like a sexed-up typewriter. However, a keystroke wouldn't make a mark on a paper; it would cause a light on the machine to turn on. If you pressed the letter key "A," for example, the light for the letter "F" might illuminate; the letter *A* was encrypted as an *F*. But if you pressed "A" again, it might show up as an "S" or an "O" or a "P"; each time you hit the letter "A," it would wind up encoded in a different way. This is because the heart of Scherbius's machine was a series of mechanical rotors. Every time you pressed a key, the rotors would turn, clicking forward one step. When the rotors changed position, the encryption changed, too. Every time you pressed a key it was encrypted in a different way. It was as if the Enigma machine changed ciphers with every keystroke.

Most models of Enigma used three rotors (though some had four), each of which would click forward twenty-six times before returning to its original orientation. These rotors could be wired in a number of different ways and placed in each of the three (or four) rotor slots. There were also wires and plugs that could be changed and some other features that could be altered, too. All told, a standard three-rotor Enigma machine could be configured in more than 300 million billion googol ways. If you were handed an encrypted Enigma message, you would have to figure out which one of those 3×10^{114} configurations the encrypter's machine was in when he started typing the message.

Brute force is out of the question; there's no way you could try each one of those 3×10^{114} configurations by hand. If every atom in the universe were an Enigma machine, and each one were trying a million billion combinations per second from the beginning of the universe until now, they still would only have been able to try 1 percent of all the possible configurations. No wonder Enigma had a reputation for being uncrackable. Luckily for Western civilization, it wasn't.

One of the best-kept secrets of the war was a small cadre of codebreakers at a Victorian estate: Bletchley Park in Buckinghamshire, England. Winston Churchill would later call the group the geese that laid golden eggs but never cackled. And Alan Turing was the most famous goose of them all.

Born in 1912 in London, Turing was to become one of the founders of the discipline of computer science—the field that deals

stages of World War II, Germany's U-boat fleet nearly strangled the island fortress of Great Britain. Prime Minister Winston Churchill later wrote that "the only thing that ever really frightened me during the war was the U-boat peril." In the second half of 1940, the Nazi navy's "happy time," U-boats sent about half a million tons of shipping per month to the bottom of the Atlantic, nearly bringing Great Britain to her knees. The Enigma codebreakers changed that trend. Since U-boat communications were encrypted with the naval version of Enigma, the Bletchley Park codebreakers helped British antisubmarine forces hunt down the U-boats that had caused their nation so much trouble, and helped win the war.[7]

The cryptanalysis of Enigma was the last great codebreaking effort before scientists learned how to define information, manipulate it, and analyze it. The Bletchley Park codebreakers, without really knowing it, were exploiting the irreducible, palpable nature of information. They were using redundancies, computer algorithms, and mathematical manipulations to burn through the cipher and extract the information that had to lie underneath it. In a sense, the cracking of Enigma was the shining star that heralded the birth of both computer science and information theory—and Turing's ideas would be an important part of both.

Sadly, Turing himself would not play a major role in the newborn science of information theory. In 1952, Turing, a homosexual, pleaded guilty to charges of "gross indecency" for his dalliance with a nineteen-year-old boy. To avoid imprisonment, he consented to undergo "treatment"—a set of hormone injections

that were supposed to end his sexual proclivities. They didn't, and his "moral turpitude" was a stain that he never recovered from. Two years later, the tortured Turing apparently killed himself with cyanide.

Turing's tragedy came at the very moment when physicists and computer scientists would learn to deal with the entity of information, at a time when scientists would see that this hard-to-define concept of information holds the key to understanding the nature of the physical world. It was not the only suicide that cast its shadow on the science of information. In fact, tragedy lingered at the very roots of information theory, around the early work in physics that set the groundwork for the revolution to come.

DEMONS

> A hostile Demon are you, that I well perceive, And fear your work is ever turning good to ill.
>
> —Johann Wolfgang von Goethe,
> *Faust*

On the afternoon of September 5, 1906, Ludwig Boltzmann found a small cord and wrapped one end around a crossbar in the wooden casement of a window. As his wife and daughter paddled happily about the bay of the resort town of Duino, in what was then Austria-Hungary, Boltzmann fashioned a crude noose with the other end of the cord and hanged himself. His daughter found the body.

Inscribed on Boltzmann's grave is a very simple equation: $S = k \log W$. This expression would revolutionize two seemingly unrelated fields of physics. The first, thermodynamics, deals with the laws that govern heat, energy, and work—and is the source of the most powerful law of physics. Boltzmann would not survive to see the second, information theory, come to life.

At first glance, thermodynamics and information theory might

seem as if they have nothing in common. One deals with the extremely concrete ideas that any nineteenth-century engineer could appreciate. Heat. Energy. Work. These are the things that make factories run, steam engines chug, and foundries glow. Information, on the other hand, is seemingly evanescent and abstract; you can't put information in a vat and have it melt steel, or stick it in a mill and make it spin wool. Nevertheless, the roots of information theory lie in thermodynamics. And both disciplines are rife with demons.

In the late eighteenth century, Europe was a continent full of demons, and France was no exception. The French Revolution in 1789 deposed Louis XVI and eventually toppled the head from his shoulders, and in the despotic fervor in the years afterward, a great many citizens followed their monarch to the grave. Among them was the great French scientist Antoine-Laurent Lavoisier.

Lavoisier was partially responsible for the birth of the discipline now known as chemistry. His experiments showed that chemical reactions neither destroyed mass nor created it—when you burn something, for example, the mass of the products always equals the mass of the reactants—a principle now known as the conservation of mass. He also proved that the process of combustion was due to a substance in air: oxygen. In his *Elementary Treatise of Chemistry,* which was published the same year as the French Revolution, he set the groundwork for the new scientific field of chemistry, in part by listing a set of "elements,"

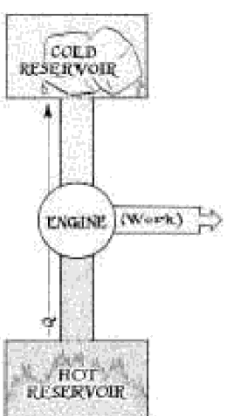
thermodynamics.

Not all the revolutions in Europe were political. Just as kings were overthrown, so were ancient lifestyles and old ideas. The science of thermodynamics was born in a revolution that swept away the last vestiges of the feudal system: the Industrial Revolution. All over Europe, inventors and entrepreneurs were trying to automate labor-intensive tasks and create machines that were stronger and faster than humans and pack animals. The cotton gin, the power loom, the locomotive—all of these inventions didn't need wages and allowed industrialists to make ever-greater profits. But at the same time, these inventions needed power to make them work.

Before industrialization, human power, animals, and the flow of water were sufficient power sources for the machines of the day. But the machines of the Industrial Revolution required much more power than the machines of old, so the "engine" was born.[1] The most famous was patented in 1769 by the Scottish inventor James Watt: a sophisticated version of the steam engine.

In principle, the steam engine is very simple. First, you need a fire. This fire causes water to boil into steam, which takes up more room than the equivalent amount of water—it expands. The expansion of the steam does work: it moves a piston which then, in turn, can move a wheel or lift a rock or pump water. The steam then either flies away into the sky or moves into a cool chamber exposed to the air and then condenses, flowing back toward the fire to begin the cycle again.

Even more abstractly, the steam engine sits in between a high-temperature object (the fire) and a cold-temperature object (the air). It allows heat to flow from the high-temperature reservoir to the cold-temperature one through the motion of the steam. At the end of the cycle, the hot object is a little cooler (you have to keep stoking the fire to keep it going), and the cool object is a little warmer (the steam has heated the surrounding air a bit). But in allowing that heat to flow, the engine extracts some of the energy and performs useful work.[2] And so long as there is a temperature difference between the hot reservoir and the cold reservoir, an ideal engine like this—a heat engine—will continue to putter away.



A heat engine (Q = heat)

Benjamin Thompson, the English physicist James Joule, and other scientists would show that there is a relationship between work and heat—that both work and heat are ways of transferring *energy.* There is energy stored in a lump of coal or a drop of gasoline. By burning them, you can release that energy and

transfer it into the engine itself. The engine then uses some of that energy in doing useful work—lifting a concrete block a few meters, for example. But some of that energy is released into the environment. And unless you keep adding energy to the hot reservoir to keep it warm (or continuously remove energy from the cool reservoir to keep it cool—more on this shortly), the two reservoirs will soon reach the same temperature and the engine will sputter to a halt.

Obviously, engineers would like to use as much of that energy as possible in doing useful work, and waste as little of that energy as they can by minimizing the heat released into the environment. In other words, they want to make their engines as *efficient* as possible. This became a major effort; in the early 1800s, one of the big problems was figuring out how to make steam engines ever more efficient. It was a child of the French Revolution who discovered the ultimate limit to the power of an engine.

Sadi Carnot was born in Paris in 1796, two years after Lavoisier lost his head to the guillotine. His father, Lazare, was a general and a member of the pre-Napoleonic French government. The young Carnot became, like Benjamin Thompson, a military engineer, but his interests soon turned to the problem of steam engines. And he was more scientifically minded than Thompson: he wanted to find out the general principles that limited the engines of the engineers.

In the 1820s, scientists still had little understanding about the interconnections between heat, work, and energy in an engine, so

Carnot began calculating, setting up careful analyses to figure out how the ideas interrelated. For example, in 1822 he tried to determine how much work can be done by a given amount of steam. But Carnot is most famous for figuring out how much work a steam engine *can't* do.

Carnot's brilliant idea was to examine an engine that is, in theory, totally reversible. Each of the steps in this (imaginary) engine's cycle can, the instant after completion, be reversed without any loss. For example, a quick and violent compression of a cylinder full of gas is reversible; if allowed to do so, the gas could expand to its original volume, pressure, and temperature, completely reversing the compression. It turns out that the efficiency of a reversible, Carnot, engine depends only on the temperatures of the heat reservoirs. Nothing else matters. For example, a Carnot engine that uses just-evaporated steam at 100 degrees Celsius and ejects the steam into the air on a freezing day at 0 degrees Celsius can only be about 27 percent efficient. Only about 27 percent of the energy stored in the steam can be turned into useful work; the rest flows as heat into the air.

That doesn't seem like a very efficient process. Three-quarters of the energy is wasted by a Carnot engine operating between 0 and 100 degrees Celsius. But it turns out that this is the most efficient heat engine you can get. Here's where the reversibility comes in.

A heat engine straddles a hot reservoir and a cold reservoir. By cycling through several steps, the engine allows heat to flow from

the hot reservoir to the cold reservoir and, in the process, extracts useful work, say, by turning a crank. In a Carnot engine, though, every step is reversible. In fact, you can run the entire cycle backward. You can take a Carnot engine and put work into it. Turn the crank. This makes the cycle run in reverse. The engine pumps heat from the cold reservoir to the hot reservoir: the hot side gets hotter and the cool side gets cooler. A heat engine, in reverse, is a heat pump: put work in and you cool down a cold reservoir and heat up a warm reservoir.

Refrigerators and air conditioners are heat pumps like this. In refrigerators, the cold reservoir is inside the fridge, and after you add work with an electric motor, the pump takes heat from inside the fridge and releases it into the hot reservoir, the room-temperature air in your kitchen. With air conditioners, the cold reservoir is the room you are cooling; the hot reservoir is the summer's day outdoors, which is why you always must have a component of your air-conditioning system sticking out of the room you're cooling.

Now, imagine a Carnot heat engine and a Carnot heat pump straddling the same reservoirs. The Carnot engine allows a certain amount of heat, Q, to flow from the hot reservoir to the cold reservoir. In the process, it produces a certain amount of useful work. The Carnot heat pump consumes that work and, in so doing, pumps Q heat from the cold reservoir to the hot reservoir. Hook the heat engine and the pump together and they exactly cancel out each other. Looking at the engine-pump system as a whole, no net heat flows from reservoir to reservoir, and no net work is

spinning motion to drive a crank or do other useful work.

James Prescott Joule, son of a Manchester brewer, was experimenting with electric motors when he realized that current passing through a motor heats up the motor itself. But a motor that's doing useful work will generate less heat than a motor that's jammed and not spinning. Do more work, make less heat. Do less work, make more heat. Like Benjamin Thompson, Joule had found a connection between physical work—lifting rocks, turning drill bits—and the generation of heat. But unlike Thompson, Joule was a careful experimenter and set to work measuring precisely how much heat and how much work were generated under different conditions.

Joule did a large number of experiments with different systems, not only electric motors but also physical systems like waterwheels, and figured out how work is converted to heat is converted to electricity, and back again. For example, he dropped a weight and used the physical motion of the weight to turn a generator and create an electric current in a wire, showing the relationship between physical work and electrical energy. In his most famous experiment, he used the motion of a paddle wheel to warm up a container full of water, demonstrating, once and for all, that work can be converted into heat. Because they are interconvertible, work, heat, and electrical energy—in fact, all forms of energy—can be measured in the same units.

Just as the fundamental unit of time is the second and the fundamental unit of distance is the meter, the fundamental unit of

energy is the joule. One joule will allow you to lift a one-kilogram rock by about one-tenth of a meter; it will heat a gram of water by about a quarter of a degree Celsius; it will light a one-hundred-watt electric lightbulb for one-hundredth of a second.

By his basement experiments, James Joule showed that work and heat were means of transferring energy from one body to another. If you lift a one-kilogram weight by one-tenth of a meter, the weight has one joule more energy than when you started; similarly, if you heat up a gram of water by a quarter of a degree, the water has one more joule of energy than when you began. He also showed that if you are really clever, you can convert from one form of energy to another; dropping a one-kilogram weight by one-tenth of a meter can, in theory, heat up a gram of water by a quarter of a degree.(In reality, you can never convert the full amount, as will become abundantly clear shortly.) But in all of these experiments, Joule realized that you never get more energy out of a system than you put in. A kilogram weight dropping a tenth of a meter will never, ever heat up a gram of water by *more* than a quarter of a degree. The energy doesn't appear out of nowhere. In the experiments, Joule was converting energy from one form to another, but he never was able to *create* energy.

Joule—and a number of other contemporary scientists—had discovered what is now known as the *first law of thermodynamics.*[3] Energy cannot be created. In fact, it cannot be destroyed, either. It can change forms; it can be transferred in the form of work or heat; it can dissipate; it can speed out of the room that you are experimenting in. But energy never simply pops into existence or

is annihilated into nothingness.

This is an extremely powerful law. It tells you that the amount of energy in the universe is a fixed constant, that all the energy we will ever be able to use is already here, stored somewhere in another form. Whenever we use energy—to heat something or to do physical work—we are simply converting preexisting energy (the chemical energy stored in coal)[4] into a different form that is more useful to us. A steam engine, for example, cannot create energy; it is extracting energy from its fuel. It's one of the most fundamental rules in physics: energy can be neither created nor destroyed. But there was an even more powerful law to come.

In the 1860s, the German physicist Rudolf Clausius noticed a subtle pattern to what energy transformations do to their environment. A heat engine relies upon a temperature difference—a hot reservoir and a cold reservoir—to work; it allows heat to flow from the hot side to the cold side and extracts work in the process. When the engine is done running, the hot side has cooled down and the cold side has heated up; the two reservoirs are closer in temperature than they were when the engine started running. The two sides started off very different, and the engine brought them closer to equilibrium with each other. In a sense, the equilibrium of the universe as a whole increases when you run an engine.

Can you get the two reservoirs *further away* from equilibrium instead of closer? Sure. All you need is a heat pump straddling the two sides. Add energy in the form of work, and the hot side gets hotter and the cold side gets colder; the two are brought ever

further out of equilibrium. But Clausius realized that there was a hitch. How can you do the work to run the heat pump? Perhaps you run another engine—but that engine increases the equilibrium of the universe as it runs, canceling (in fact, *more* than canceling) the decrease in equilibrium caused by the heat pump. The equilibrium of the universe increases, despite your best efforts.

What if you don't use an engine? What if you turn a crank by hand? Well, in actuality your muscles are acting as an engine, too. They are exploiting the chemical energy stored in molecules in your bloodstream, breaking them apart, and releasing the energy into the environment in the form of work. This increases the "equilibriumness" of the universe just as severely as a heat engine does.

In fact, there's no way to get around the ever-growing equilibriumness of the universe. Whenever someone uses an engine or does thermodynamic work, the process automatically brings the universe closer to equilibrium. You can't counteract the increase in equilibriumness with a heat pump or another device, because the work needed to run the device would have to come from an engine or a muscle or some other source that cancels out the heat pump's efforts.[5]

This is the *second law of thermodynamics.* It is impossible to reduce the equilibriumness of the universe; in fact, every time you do work, you drive the universe closer to equilibrium. Where the first law says you can't win—you can't create energy out of nothing—the second law says you can't break even. Anytime you do useful

work, you are irreversibly increasing the equilibriumness of the universe. The second law is also why there is no such thing as a superengine that works better than a Carnot engine. A superengine hooked up to a Carnot heat pump is doing useful work without changing its environment; in fact, you can isolate that engine-pump system in a box, yet it will still be able to keep doing useful work indefinitely. The equilibriumness of the universe wouldn't change at all, even though your machine can do useful work. But the second law of thermodynamics says that an engine or other device must feed on the nonequilibriumness of the universe—and you can't create work out of nothing, thanks to the first law of thermodynamics. Therefore, the superengine cannot exist. It would lead to a device that does work indefinitely without reducing the equilibrium of the universe. It would lead to a perpetual motion machine.

Inventors and hucksters have been trying to build perpetual motion machines for centuries, and even today there are lots of people who will try to sell you one. (Since "perpetual motion machine" is a sure way to scare off investors, the current term of the art is *above-unity device.*) Some of the designs are based upon magnetic fields; others are based upon various "quantum" technologies. The U.S. Patent Office has been so deluged with applications for perpetual motion machines that the office has a special rule for such devices: an inventor must submit a working model with his application. (Nevertheless, some slip through the cracks and actually attain a patent.) But the second law of thermodynamics—now considered to be the most unassailable law

behaved in this way.

Modern physicists know that a gas, like helium, is made of tiny particles—atoms. These atoms are constantly in motion, flying about the container at different speeds. When an atom collides with the container, it ricochets away like a racquetball off a wall. But the collision gives the container wall a tiny knock. One collision has little effect on the walls of the container, but zillions upon zillions of these tiny ricochets collectively take their toll. They exert a strong pushing force on the container walls, forcing them outward. This is the source of a gas's pressure.

If you squash the container, then the same number of atoms are in a smaller space, and because the container is more crowded, more of these racquetball atoms slam against the walls per second. The number of ricochets goes up, the collective force that the ricochets exert increases, and the pressure rises. This is what causes Boyle's law: because decreases in volume increase the frequency of collisions and vice versa, volume and pressure are inversely proportional. Similarly, physicists now know that a gas's temperature is a measure of how much energy the atoms have. This, in turn, is related to how fast the atoms are skittering around. The hotter the gas is, the more energy it has and the faster the atoms move, on average. (This is the true nature of temperature. It is a measure of how energetically, and by extension how fast, an atom is moving. A hot atom of helium is moving faster than a cold atom of helium; conversely, a fast-moving atom is hotter than a slow-moving atom of the same species.) The more energetic an object is and the faster it moves,

whether it be a racquetball or an atom or an SUV barreling down the highway, the more kick it imparts to the object it slams into. So, increase a gas's temperature and the atoms move faster and ricochet off the walls harder, giving the walls of the container an even firmer push—the pressure of the gas increases. If the walls of the container are allowed to move, the container will expand to equalize the pressure. This is what causes Charles's law.

Atomic theory ties pressure, temperature, volume, and energy—all the concerns of thermodynamics and the steam engines of the Industrial Revolution—into a nice, neat package. However, what is obvious to modern scientists was difficult to accept for nineteenth-century physicists. After all, nobody had the means to detect an individual atom; as late as the early twentieth century, some eminent scientists refused to believe that atoms existed at all. But in the mid-nineteenth century, physicists were beginning to realize that atomic theory—the idea that matter was composed of constantly moving, billiard-ball-like particles—did an excellent job of explaining the properties of gases and other types of matter. In 1859, Rudolf Clausius published a paper that set the groundwork for what became known as the *kinetic theory of gases*—but he ran into trouble. He couldn't get the numbers to work out quite right.

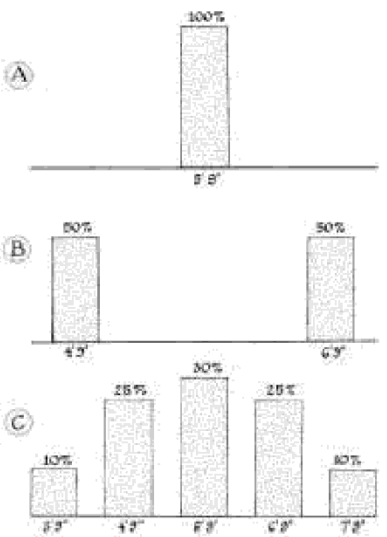
The trouble was with temperature. Clausius knew that temperature was a measure of the energy of the atoms in the gas: the hotter the gas was, the more energy the atoms had and the faster they were moving. Indeed, if you know how hot a gas is and how heavy the atoms are, then you can easily work out the speed of an average atom. Clausius did this and then worked out what

would happen if you had a container full of tiny atomic billiard balls all moving at this particular speed. While the results were encouraging, Clausius's analysis wasn't quite correct; the relationship of pressure, temperature, volume, and energy wasn't quite what was observed in nature.

In 1866, the Scottish physicist James Clerk Maxwell figured out the flaw in Clausius's argument. While Clausius assumed that all the atoms in the gas were moving at the same speed, Maxwell realized that when the billiard balls collided with the walls and with each other, they exchanged energy. Some would wind up moving faster than average and others would wind up moving slower than average. Maxwell realized that by assuming that the speeds of the molecules had a particular *distribution,* he could fix the inaccuracies in Clausius's theory.

A distribution is an expression that appears often in a branch of mathematics, probability theory, that deals with uncertainty. It is a measure of how common something is. Imagine someone asks you how tall the average American adult male is. That's not too tough a question. You can say that the average height is around 5 feet 9 inches tall. But what happens if the person asks you to describe how tall American males are, in general? You can't just give the average, because that doesn't provide very much information. An average height of 5'9" could mean that every male is exactly 5'9" tall, or it could mean that there are two groups in the population: 50 percent are 4'9" and 50 percent are 6'9". Or maybe 10 percent are 3'9", 25 percent are 4'9", 30 percent are 5'9", 25 percent are 6'9", and 10 percent are 7'9". The *average* height in

each of these cases is 5'9", but a roomful of men from one of these populations would look very, very different from a roomful of another because the *distribution* of their heights is different. In a distribution where every male is 5'9" tall, there is zero probability that if you pull a male randomly off the street, he will be taller than 6 feet. He must be 5 feet 9 inches exactly. But in the five-group distribution cited above, there's a 35 percent chance that if you pull someone off the street, he will be more than 6 feet tall (25 percent at 6'9" plus 10 percent at 7'9").



Different height distributions

Of course, those examples don't really represent the true distribution of heights. In reality, the distribution of heights is very close to what is known as a *bell curve* distribution. In a bell curve, the "extreme" events are much, much rarer than "average"

the time it will wind up in the left half. It's a pretty pathetic way to kill time—it's only slightly better than watching reality TV—but this simple setup is all we will need to understand the concept of entropy.

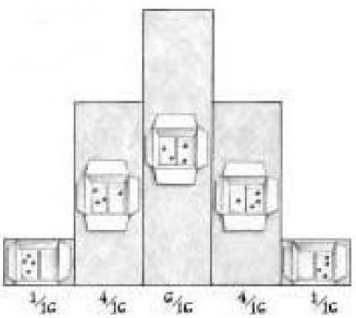
Let's start out with two different marbles. Plunk, plunk. And now let's look inside the box and see what happened.

When I peer into the box, I'm greeted by one of four possible outcomes. Case 1: The first marble I threw landed on the left side of the box, as did the second marble. Case 2: The first marble landed on the left, but the second landed on the right. Case 3: The first marble wound up on the right and the second settled on the left. Case 4: Both marbles landed on the right side of the box. Each one of these possibilities is equally probable; that is, there's a 25 percent chance for each case.

However, things change slightly if the marbles look exactly the same. In this case, you can't tell which one you tossed first, so when you look in the box, there are only three possibilities: both marbles are on the right; both marbles are on the left; or there's one on each side. In other words, case 2 and case 3 become indistinguishable (or *degenerate,* in physics-speak). This degeneracy means that the possibilities are no longer equally probable. As before, there's a 25 percent chance that both marbles are on the left side of the box and a 25 percent chance that both marbles are on the right side of the box. But the third possibility—that there's one marble on the right and one marble on the left—happens 50 percent of the time, because there are two ways it can come about.

Marbles on the left	Marbles on the right	Probability
4	0	1/16
3	1	4/16 = 1/4
2	2	6/16 = 3/8
1	3	4/16 = 1/4
0	4	1/16

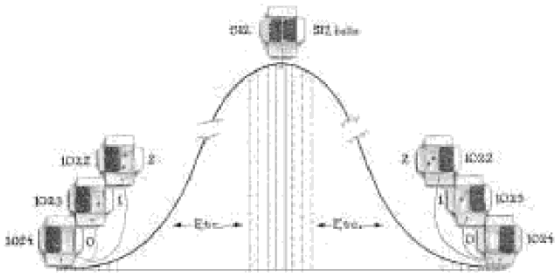
Probabilities for four indistinguishable marbles in a box



Four indistinguishable marbles in a box

In fact, the more marbles we toss into the box, the clearer the bell curve becomes. No matter how many marbles we throw, on average half the marbles will fall into the left half of the box and half will fall into the right half, and this is always the most probable outcome of any given trial. The most extreme events are when all the marbles are on the right half, or all on the left half, and these extremes are much, much less probable than the average, or *mean,* event. All the other events lie in between the mean and the extreme and become dramatically less probable as they move from the mean to the extreme. And the more marbles we toss into the box, the less probable the extreme events become. For instance, let's throw a nice large sample of 1024 marbles into

the box. On average, 512 will wind up on the left-hand side and 512 will wind up on the right. An extreme case, such as 1024 on the left and 0 on the right, is unimaginably improbable.



1024 indistinguishable marbles in a box

How improbable? Toss 1024 marbles randomly into a box. Look inside. Take out the marbles and toss them in again. Look inside. Take the marbles out and toss them in. Look again. Repeat and repeat and repeat. If you did this once a second from the beginning of the universe to now, the odds of ever seeing a 1024-on-one-side case are about 10^{290} to 1 against. Indeed, if every atom in the universe were one of these 1024-marble boxes, filling up randomly with marbles over and over again every second from the very beginning of the universe, not one of those boxes would ever have a 1024-on-one-side trial.(It's not even close. There are only about 10^{80} atoms in the visible universe.) While it is not *impossible* to randomly get all 1024 marbles on one side, it is so improbable that it's *functionally* impossible. It won't happen in this universe.

So what? Why waste our time playing around with boxes and marbles? Because it leads directly to a simple definition of entropy.

entropy will be high, and the atoms will be, more or less, uniformly distributed throughout the container. In fact, no matter what property of the atoms we look at, the highest-entropy state corresponds to a uniform distribution of that property. For example, whenever we peek into the container, the high-temperature, fast-moving atoms will tend to be evenly distributed throughout it; so will the low-temperature, slow-moving atoms. It is extremely unlikely that all the high-temperature atoms will cluster on the left side of the container and all the low-temperature atoms will clump together on the right. Instead, the gas will almost certainly be a uniform temperature throughout. That's the highest-entropy state, and it's a virtual certainty that the helium will be in that state when we peek in. Under conditions where an isolated container of gas is allowed to randomize—where it is allowed to evolve to equilibrium—we are almost guaranteed never to see the left side cold and the right side hot.

But what about a drafty room? It's cold near the window, while it's toasty near the radiator. At first glance, this might seem to contradict the concept of entropy. However, this is a system that is not isolated; the radiator keeps pumping warm air into the container, while the window lets it escape. If we were to seal off the window and turn off the radiator, the room would quickly reach an equilibrium where every place is the same temperature. Similarly, we can inject a bunch of helium atoms on one side of a container, knocking it out of equilibrium, but left alone, the container will rapidly revert from its low-entropy state (with lots of atoms on one side and few on the other) to its high-entropy

state (a roughly equal number on each side). It's as if the system is attracted to its high-entropy state—and, in a sense, it is. Just as a ball "wants" to roll down a hill, a box full of gas "wants" to maximize its entropy. You can do work—put in energy—to reverse the trend toward high entropy in a system such as using an air conditioner or a heat pump to keep one side of the container hot and the other side cold, but left to its own devices, a container full of gas will revert to its maximum-entropy state, with hot and cold atoms evenly distributed throughout.[9]

Atoms' "desire" to maximize their entropy leads to an irreversible change in a container full of gas. If you start off with all the atoms in one corner of a box, then after a little while the atoms will spread out, maximizing their entropy. Since it's so improbable that all the atoms will move back to the corner they came from, the gas is essentially permanently in a state of high entropy: once it reaches equilibrium, the gas will always be in a high-probability state and will never revert to the low-probability state whence it came. Similarly, if you start off with a bunch of hot atoms on the right side of the box and cool atoms on the left side of the box, then after a while the hot atoms and the cool atoms will jostle around randomly and will move into the most probable configurations: the hot atoms and cold atoms will be equally distributed on the right and left sides of the box. And once the box is in equilibrium, you can watch it for centuries and you will never see the evenly distributed hot and cold atoms suddenly separate themselves and wind up on opposite ends of the box. If left to itself —if you don't use a heat pump or add any energy to the system—

a state of chaos when you chill down a bottle of beer in your fridge.

Boltzmann's statistical and probabilistic view of the motion of atoms in matter was incredibly powerful. By looking at a gas as a collection of randomly moving particles, he was able to explain the physical principles that drove engines, that were responsible for heat flow, for temperature, for work—and for entropy. Most of all, for entropy. Through simple probability and statistics, Boltzmann's work led to the understanding that systems naturally "try" to increase their entropy, and that the universe as a whole is constantly, irreversibly, getting more entropic. But hidden in his logic was a time bomb.

The probabilistic nature of Boltzmann's work made it appear as if it undermined the absolute truth of the very laws it explained. The second law of thermodynamics was based upon the fact that gases *probably* wind up in their most probable configurations. Sounds redundant...but probably isn't absolutely. Once in a while, perhaps, the gas randomly winds up in an improbable configuration—it can happen. That means that the entropy of the system, without any energy being added, can spontaneously decrease. The second law, to all appearances, is suddenly violated. Worse yet, James Clerk Maxwell, the man who embraced the statistical nature of gases and came up with the distribution of speeds of the atoms in a gas, devised a clever method that seemed to separate hot atoms from cold atoms without any work at all—an even more profound violation of the second law.

Boltzmann proved that the second law must be true. But at the