Excerpt from

THE DECLINE EFFECT

The law behind diminishing returns and wildly varying outcomes in markets, politics, culture, religion, disease, and war

Dean Brooks

Overture and Chapter 1

Overture

This changes everything!

Two friends, Socrates and Francisco, are about to attend a lecture about a new theory in mathematics.

Francisco: So what is all this about?

Socrates: Did you not read the book before coming?

Francisco: No, I'm afaid I didn't. It just seemed so long. I looked at some of the graphs.

Socrates: You usually enjoy mathematics.

Francisco: Oh, certainly. But from what I could see, the text seemed to go all over the place, from IQ tests to the distribution of planetary masses to Hegel and Aristotle and the history of religion. In one chapter it's something to do with catching serial killers, and then in the next it's a graph of Wal-Mart stores. I mean, a graph of how many *Star Trek* books have been published? And a graph of Nazi party membership. It was all a bit too much.

Socrates: You usually like big books on science. Seems like you're always carrying one around.

Francisco: Yes, but in this case I just couldn't take it in. I suppose maybe I prefer my recreational mathematics to be more abstract. I usually go for fractal geometry, or Sudoku puzzles. Something uncontroversial. Founded on universal truth, and all that. A math book that spends pages and pages on Chinese dynasties and Carl von Clausewitz seems sort of weird to me.

Socrates: Hm. Maybe that's the problem with truth. The more universal it is, the more surprising it is. There's more and more at stake. I would think genuinely universal truth is the sort with the potential to upset everyone.

Francisco: You're a sour cuss, Socrates. Everyone says that about you. But never mind, you're not about to change. So in a word, what is this theory about?

Socrates: It says our ideas about probability are wrong.

Francisco: What, you mean all of them? Yes, that should upset everyone very effectively. Just your sort of thing.

Socrates: No, not all of them, but more the way they are put together. The classical model of how the world works is out of date. Newton and Pascal and the others back in the 17th century got it subtly wrong. All our discoveries since then point to a very different order.

Francisco: I presume it adds up to some sort of decline. I got that much from the title.

Socrates: Twenty-odd years ago, this fellow started collecting diminishing returns curves. You know, like when you double the number of laborers but you don't get double the amount of work from them. As if there's a friction loss somewhere.

Francisco: I see why youwould be interested. He sounds like just your sort of man. Cataloguing different kinds of frustration.

Socrates: No more sarcasm, please. Our lecturer found something very odd: Dozens and dozens of unrelated curves converged to the same shape.

Francisco: All right. I'm waiting for the punchline.

Socrates: I mean the *very* same. The same slope. Doubling input resulted in about 62 percent more output in each case. Not really 62 percent. It was 61.8034 percent. It was the golden ratio, phi. I presume that means something to you?

Francisco: Oh yes, I'm very interested in phi. It makes such pretty patterns. It's like the original fractal. So what, exactly, does that mean? I suppose if I were to double the membership of my local *Star Trek* club, I would only get about 61.8 percent more attendance at our Klingon singalongs?

Socrates: Quite likely you would. One

example I do recall is that if you double the number of automobiles on the road, you only get 61.8 percent more fatal accidents. Or drop double the number of bombs on a city and you only get 61.8 percent more casualties.

Francisco: That could be useful to know, if it really was consistent.

Socrates: More than useful. It's provocative, because of how consistent it is. There's a whole catalog of laws—from biology and economics to criminology and insurance—that turn out to be the same law. People on the cutting edge in those fields are all studying this one law, but no one seems to have realized it yet.

Francisco: Ah. So what is this law?

Socrates: That entropy always increases.

Francisco: No, wait, we already know that one. That's the second law of thermodynamics. I'm sure it was discovered by what's his name—Boltzmann? And then there was someone named Shannon? Or who was it that came up with information entropy?

Socrates: Yes, it was Boltzmann and then Shannon, but this version is different. This version of entropy was thought up by Edwin Jaynes, about fifty years ago, and it has gone underused since. It says the entropy of traffic accidents always increases. And the entropy of languages, and political systems, and armies.

Francisco: I didn't know traffic accidents had their own entropy.

Socrates: That's the point. Until very recently, no one did.

Francisco: And political systems? There's such a thing now as political entropy? Yikes.

Socrates: And religious entropy.

Francisco: Oh boy. I guess that's something the world needs to know about right away: Religious entropy!

Socrates: Are you joking now, or are you serious?

Francisco: Sometimes I can't even tell myself.

Socrates: Cultural, economic, religious, political, and military entropy, among others. They're all increasing, pretty much all the time. They help determine what the economy or the church or the government will do.

Francisco: So this new law, it makes specific predictions? What does it predict? That the world is in decline, that everything is getting worse? I hardly need point out, yet again, why that's just the sort of theory you would think was brilliant—

Socrates: Not necessarily worse. Sometimes the curve is a diminishing cost curve. Or a diminishing amount of time to accomplish something. The good and the bad quite often balance out. The same curve governs species diversity and what advertisers pay for magazine ads. It governs fetal growth rates, and the growth of the Internet, and the behavior of serial killers.

Francisco: So the point is simply that the world is governed by entropy curves?

Socrates: Yes. But that one fact changes everything. Everything!

Francisco: Well, I guess we'll see. Here we go, they're opening the doors ...

Chapter 1

Introducing the decline effect

Rethinking randomness

This book is about an historic misunderstanding in science, a mistake with roots more than 350 years in the past. It turns out that for all this time, we have been wrong—subtly but fundamentally wrong—in our thinking about randomness.

The phenomenon that science has overlooked is relatively easy to describe, and to demonstrate. In this book I have graphed more than 200 examples using public data sources. To put the phenomenon in the plainest possible language, rare events and extreme values tend to slowly become rarer. As a system grows larger, or older, or both, such events and values take up a smaller proportion of the whole. I call this tendency *the decline effect*. All that is required to see this process happening firsthand is a high-school-level grasp of how to read graphs, some knowledge of exponents and logarithms, and of course, a clue about where to look.

The process by which science came to be wrong about randomness is somewhat more difficult to explain. It is a long story, centuries in the making, sprawling across many different scientific disciplines. The implications are similarly enormous.

The error begins in classical probability, the basic rules of statistics that were first laid down by Cardano, Pascal, Fermat, and Newton in the 16th and 17th centuries. The ruling assumption behind classical probability is that events are independent. Here we are talking mainly about *random* events, and we will get into technical details later about which specific kinds.

By 'independent' we mean that if we observe a random process at any given moment, the various possible outcomes are unaffected by whatever has happened before, or by what is happening elsewhere. If I throw a six-sided die and it comes up '1,' the classical assumption of independence means that the die should have exactly the same 1-in-6 chance of coming up '1' the next time I throw it, and every time I throw it.

In reality, this doesn't happen—not quite. Not even for dice, and definitely not for more complex systems. If this was not a very good approximation to what happens, in a lot of cases, we would have realized the shortcomings of the classical view long ago. But if we look closely, there is a subtle but consistent bias pattern in virtually any set of random events that makes this assumption wrong.

To be clear from the start, when I refer to a universal bias pattern, I am not talking about some specific, familiar source of bias, such as a die that is weighted on one side, or is the wrong shape. I am also not talking about a bias introduced by the experimenter—for example, some magician's trick in how one flips a coin. For purposes of this argument, assume that if we are dealing with coins that they are fair coins, balanced to fall 50-50 either way. Assume that any dice being thrown are perfect cubes, and that if a test involves a deck of cards, it will be shuffled efficiently and honestly.

Then, according to Isaac Newton and 350 years of mathematical theory, each throw of a die, or flip of a fair coin, should present the same evenly distributed set of chances as the last. Whether you threw a '6' last time should have no influence on whether you will throw a '6' the next time.

This rule of independence has historically worked so well that it forms part of the very foundation of scientific method. I can scarcely name a mathematical, scientific, or technological endeavor that does not rely on this concept in some fashion—genetics, epidemiology, weather forecasting, stock option trading, rocket engine design, everything. Countless phenomena from aircraft accidents to sports scores, from Web traffic to subatomic particles, are all assumed (on some level) to be governed by the same rule of strict independence.

Independence is our *default* assumption. It is expected to be true, unless and until proven otherwise in a specific case.

What I and many others have found in these hundreds of observations and experiments is different. On closer inspection, we see that independence is the exception, not the rule. Even the most rigorously randomized series of everyday, naked-eye events (such as dice throws) will tend to follow these basic rules:

- Rarer events will cluster in a non-classical fashion near the beginning of a series, and will steadily become less frequent as the series grows longer.
- A variety of randomly varying measures (size, intensity, effort, cost, and so on) all tend to decline as a series grows larger, with the rarer high values steadily becoming rarer.
- If the system is sufficiently complex, then graphing this decline will lead to a curve with one of a handful of specific shapes.

Fifty years ago, two investigators independently laid the theoretical and experimental groundwork for the present study. One was George Spencer-Brown (1923-), an Oxford philosopher and mathematician. The other was Edwin

Jaynes (1922-1998), a Stanford physicist. Both are widely cited as innovators in scientific method and logic, but their names are rarely mentioned together. Each man worked on the foundations of probability, and found them profoundly unsatisfactory. They began from two quite different (but ultimately complementary) perspectives.

In 1953, Spencer-Brown published a controversial article in *Nature*, 'Statistical significance in psychical research,' addressing the mathematical implications of recent extra-sensory perception (ESP) experiments. At that time, despite growing skepticism, ESP could still qualify as an acceptable topic for a journal as prestigious as *Nature*. However, Spencer-Brown's iconoclastic thesis was guaranteed to arouse resistance not just from believers in ESP, but from skeptics as well.

The universal tendency in ESP studies, recognized since Joseph Banks Rhine's work in the 1930's, was for each subject's scores in guessing hidden values to decline irreversibly from well above chance, down to mere chance, and in many cases, continuing below chance. Spencer-Brown questioned whether this decline effect proved anything whatever about ESP. Suppose the same effect could be found elsewhere? Perhaps it signalled a failure in our theory of probability instead.

Spencer-Brown staged a series of control experiments based on published random number tables, using one group of random numbers as his target, and a second group to simulate the subject's guesses. He obtained the same pattern of initial above-chance matching, declining to well below chance. As matches are by definition rare, this was a 'decline effect' by the definition I gave earlier.

SpencerHe followed the article with a book-length treatment, *Probability and Scientific Inference*, in 1957.Spencer-Brown argued on both experimental and theoretical grounds that the existing model of probability was incoherent, and that there were perturbing forces at work in any random number generating apparatus that were poorly understood. A new theory of probability was needed. However, the effects Spencer-Brown observed were relatively small, as well as being tainted by their association (however distant and circumstantial) with ESP. Worst of all, he could not propose what the replacement for the classical theory. would be. There was a brief controversy, with further results of the same kind being generated by both Spencer-Brown and by his critics, but within a few years the issue had died away. The idea that the working of probability as such might involve spontaneous decline of rare items has received no serious discussion in decades.

Starting in 1999, I have replicated Spencer-Brown's findings on a much larger scale, both in third party data, and by direct experiment. Because the reader will surely be skeptical on this point, I need to stress that for dice, coins, cards, slot machines, lottery balls and the like, these effects are generally subtle. For small systems of fixed size, they are also quite often *transient*. The decline process is not always permanent. It depends on the size and complexity of the system. A sufficiently small and self-contained system eventually reaches equilibrium, having traversed all the states possible to it several times, and the non-classical behavior stops. This sort of transient decline is only apparent if it is clearly defined in advance, and demonstrated by careful study and comparison of cases.

For example, when throwing two six-sided dice, after a few hundred to a few thousand throws, there is a universal tendency for a slight shortage of 'doubles' results—typically 5 or 6 percent—and an even smaller bias in favor of 'sevens'. This tendency has been observed by several experimenters, both in actual casino play, and in experiment, amounting to hundreds of hours of observation in total. The amount of deviation increases logarithmically with the overall size of the random series of events. This means that the bias will be noticeably larger after 2,000 throws than after 1,000, but then to get another incremental gain of similar size, it is necessary to make about 4,000 throws—and so on.

Notice how small the effect is in percentage terms. Simply standing at a craps table for an hour, or throwing a few hundred trials at home, will not bring the observer any enlightenment. At a minimum, one must accumulate thousands of random events to demonstrate the decline. A single short test is not good enough.

This need for lengthy data series is not a unique pecularity of the decline effect. Casual observation is generally not much use in answering statistical questions, whatever they might be. On average, it would require at least a week of full-time observation at a craps table to accumulate enough trials to clearly demonstrate the decline in doubles. However, it would take an equally large number of trials to prove, say, that a set of dice was biased by hidden weights. That is just the nature of statistical reasoning.

So there is significant labor required to properly demonstrate the effect under controlled laboratory conditions. And it is evidently easy to miss, since we are only speaking of it now, in 2009, after centuries of overlooking it. Nonetheless, the effect is real, it is reproducible, and the implications matter enormously.

I am sure at this point the reader already has a swarm of questions, such as:

— Does this mean that our current methods of statistical sampling and analysis are wrong?

— Does it mean that classic casino games like craps and blackjack can be beaten?

— What causes the decline?

But the biggest question at this point—setting aside 'You've got to be kidding'—is probably this one: How could science *miss* this?

In truth, science didn't. This is where the work of Edwin Jaynes becomes critical. In the broader realm of natural science, that is, outside the laboratory, the decline effect has been documented in vast detail, and a number of cases have already been explained using Jaynes' principle of maximum entropy.

Here we need to make a distinction between the majority or mainstream consensus—the terrain immediately before us, in other words—and the more distant and obscure landscape of things that the majority consensus has so far ignored. Science has not missed the decline effect, so much as it has failed to integrate a vast array of facts and arguments. These were things that were observed and known individually, but were never understood as a coherent whole. They did not fit the classical model, and so remained as exceptions to it.

The tendency for rare events to become less frequent for no obvious reason was first observed as far back as the 19th century (and perhaps much earlier). Outside of classical games of chance, spontaneous declines are very familiar. The case is much stronger and easier to make in natural science than it is in artificial probability experiments, because the anomalies in nature are huge. Persistent downward trends of this sort have been encountered virtually everywhere—for example, in long-run manufacturing costs, animal metabolisms, traffic fatalities, battlefield performance, supercomputer network output, and participation in religious groups.

In each case, however, the decline has generally been treated as a minor, isolated curiosity, a localized law with no wider significance. We have given these various anomalies names, and filed them away. Then in many cases, a century or more has gone by without our understanding of them getting any deeper. Among the localized laws we will examine in this book are:

— Pareto's law of elite incomes
— Zipf's law of word frequencies
— Lotka's law of scientific publications
— Kleiber's law of metabolic rates
— the Clausewitz-Dupuy law of combat friction
— Moore's law of computing costs
— the 'paradox of higher demand' in religion
— the Wright-Henderson cost law
— Weibull's law of electronics failures
— the Flynn Effect in IQ scores
— Benford's law of digit frequencies
— Farr's law of epidemics
— Hubbell's neutral theory of biodiversity
— Bass's law of innovation diffusion
— Rogers' law of innovation classes
— Wilson's law of island biogeography
— Smeed's law of traffic fatalities
— the Cobb-Douglas factor elasticity curve

To find a scientific principle that had been known for centuries suddenly invalidated would be shocking. But here we have a vast range of disparate items, already recognized as exceptions to that principle. The classical assumption of independence has long since been shown to be invalid in all these cases, and mysterious variations in behavior with set size are a well-established fact.

Now we find that all these items are related. They can be explained (and a number already have been explained) by a logical extension to the original principles of probability, that is, by the work of Jaynes. This is not nearly so shocking. Such extension to cover anomalies is often the form that new scientific discoveries take. Indeed, it has happened so many times that it qualifies as a recognizable paradigm of science, a classic historical pattern.

When anomalies are encountered that resist explanation, they tend first to become *ad hoc* empirical rules, a pattern that we accept as natural without knowing its cause. For example, for centuries humans knew that lodestones suspended in water would point north, but no one could propose a reason why. This was the 'lodestone rule,' so to speak. North-pointing was the distinguishing characteristic that made them lodestones.

For even longer, humans wondered what lightning was, where it came from, what caused it. About all anyone knew was that lightning came from the sky during storms, that it was bright, that it burned. Such *ad hoc* rules are part of science's vast catalog of unfinished business, waiting decades or even centuries for something that can explain them.

Eventually a theory comes along that can do the job of bringing several empirical laws together. In the case of lodestones and lightning, the new theory was electromagnetism—plus some discoveries about the structure of ferrous and nonferrous minerals. The development of the theory of electromagnetism brought a vast number of discrete observations under a single heading. It did far more than explain why the lodestone pointed north, or what lightning was.

The thesis of this book is that randomness *in general,* the randomness of everyday objects and processes, is governed by a universal decline effect. The symptoms of this decline effect are all around us, and called by many names. The goal of this book is to integrate all these various laws under a single overarching principle—Jaynes' principle of maximum entropy. When seen in proper context, this vast list of marginal curiosities and anomalies adds up to the rewriting of scientific history.

A 'spooky' example of decline

We will jump in right away with an example of naked-eye, everyday systems behaving in what is clearly a non-classical, non-independent, and profoundly spooky way.

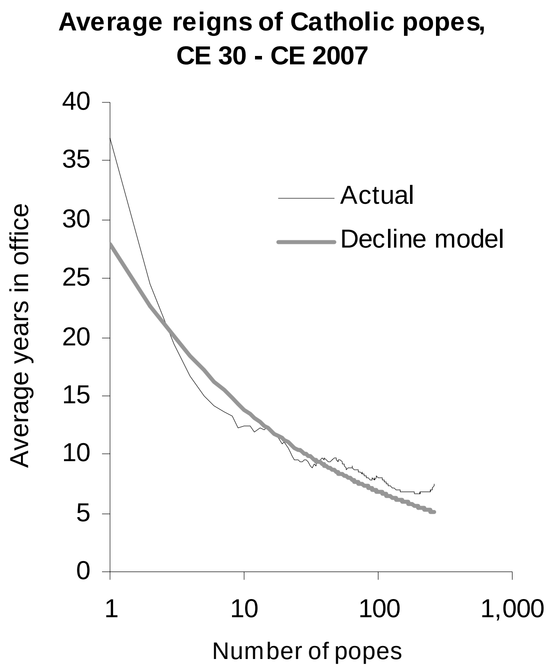
Consider a dynasty, that is, a series of rulers heading some country in history. Suppose we compute how long each king remained in power, by taking the years between his ascent of the throne, and the ascent of his successor. We will round to whole numbers of years for simplicity. For a long-lived dynasty with 40 or more kings in it, we might see a sequence like the one below, which is actually for the Catholic popes:

37 12 9 9 8 10 10 11 4 15 11 8 15 10
18 5 8 5 1 14 3 1 3 1 10 6 9 13 8 1
2 3 22 1 15 14 18 15 2 16 ...

On first inspection, there doesn't seem to be any particular logic or pattern to this sequence. It varies wildly, from a high of 37 years to a low of just one year, and that of course is the kind of behavior we would expect. We would not expect any sort of consistent relationship between the length of one reign and the length of the next. On first inspection this sequence does seem to be purely random, in the classical sense of the word.

However, on closer examination, the later numbers are somewhat smaller than the earlier ones, with the last 20 in this list averaging just 74 percent as large as the first 20. The difference is too large to be attributed to just one or two extraordinary items, or to rounding of the values, or other simple causes. This is mildly curious. Could it be that life expectancy was falling in that particular era? Could there have been secret assassinations not recorded as such? Either alternative seems unlikely. But even granting the series is historically accurate, so far all we have is a fairly trivial anomaly.

If we graph the cumulative average, we get a downward sloping line. For comparison, I have shown a theoretical curve in which the cumulative average drops by 20 percent each time the total number of office-holders doubles. Now, if we compute the cumulative average for the remaining popes throughout history, several hundred in all, the curve continues in the same fashion. The cumulative average keeps dropping. So it is not a short-term thing. Even if we track the trend over thousands of years and hundreds of rulers, it remains the same.



Now after some additional investigation, we discover something much more intriguing. The same kind of decline shows up for virtually *every* dynasty we investigate—the emperors of Japan or ancient Rome, the pharaohs of Egypt, and so on. The slope is not always a tight fit to the 20 percent curve shown here. Sometimes it is 30 percent, sometimes 10 percent. But it is always downward, and the results cluster around 20 percent. The prospects of later rulers are always worse, in a predictable way, and the trend line that governs them is global, not local. By this I mean that the rule does not relate the next reign length strictly to the last one; it relates the next reign length to the total number of reigns that have occurred so far.

When we study the records of the countries in question, we find no single explanatory principle, no practical reason why this should happen. Some wild hypotheses come to mind, but nothing that convinces. In some dynasties in history, there were problems with genetic defects associated with inbreeding. For example, Queen Victoria carried the gene for hemophilia, and passed it to royal families in Germany, Spain, and most famously, the Romanov dynasty in Russia. It did produce a decline in life expectancy for the male members of those lines, but the decline does not seem to have mattered. The most famous victim among Victoria's descendants, the Tsar's hemophilic child Alexei, was executed along with the rest of the family, a year after the regime had already been overthrown. His status as heir never had any impact on the historical record, much less his genetics. And apart from hemophilia, we have no evidence of genetic defects accumulating in dramatic fashion among hereditary dynasties. Such defects certainly do not apply to the popes!

It is the same with all the other explanations that come to mind. For example, it has become popular in recent years to try to explain history as being influenced by climate fluctuations. However, it is hard to imagine any climate fluctuation producing this specific and consistent an effect. How could the climate 'know' when a particular king comes late in a dynasty, or when a new dynasty starts? How could climate impact two different dynasties in the same region in different ways?

In fact, these curves work to substantially undermine our notions about climate affecting history, because the dynasty curves seem to just plow ahead, indifferent to any influence, including climate.

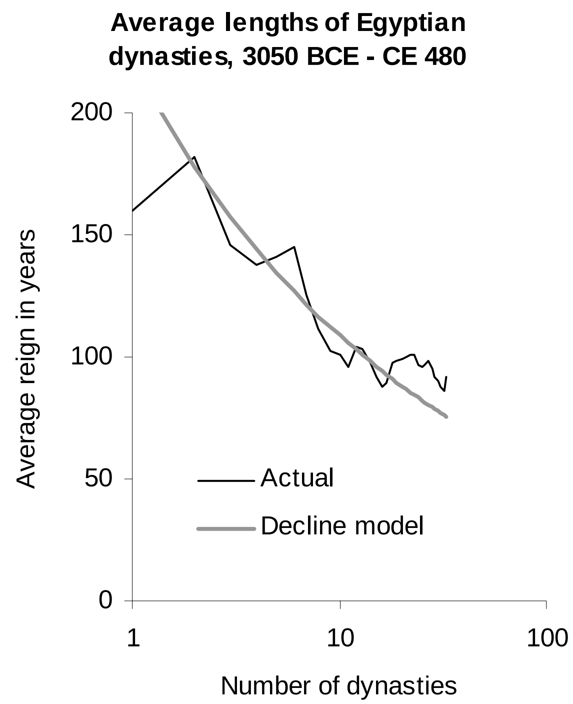
It seems plausible that long-term *social* changes, such as the increasing devolution of powers onto a parliament, would somehow shorten the reigns of later kings in a systematic way. Perhaps later members of a dynasty wound up consistently compelled to surrender power more frequently, to keep peace with the nobility or other classes. However, the same pattern shows up for the Egyptian pharaohs and other early dynasties that predate any such democratizing development. Once again the proposed explanation is plausible for a few cases, but far too narrow to cover them all.

For each dynasty, so far as we can see at this point, the decline in average reign length just happens spontaneously, with no single identifiable cause. In other words, it is an emergent phenomenon, something that is everywhere and nowhere.

Now this is strange and interesting, yes, but it still hovers somewhere on the edge of being real science. It is a curiosity, like a lodestone. Precisely because we have no good explanation, we are inclined to shrug, to put the mystery aside. We'll figure it out one day, but in the meantime, better not to worry about it.

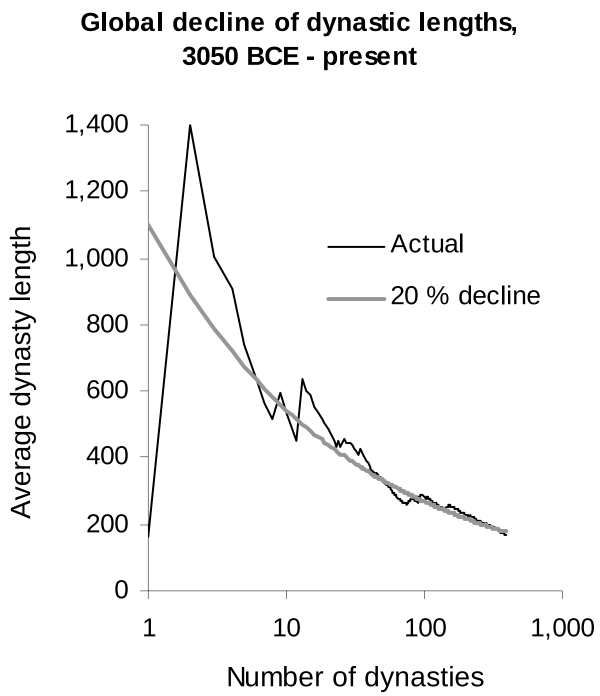
The next step signals that at least some advanced mathematics are involved. It suggests a definite direction for investigation. For a number of countries, we can plot a curve for successive *dynasties*, rather than successive *rulers*. We are now looking for a trend in a set of sets. The average slope tends to be the same yet again, a line sloping down at or near 20 percent for each country.

This reveals an interesting kind of scale invariance. The rule we have discovered is not really about rulers at all. We have now abstracted the rulers out, but the pattern persists. An individual dynasty might consist of two rulers, twenty, or two hundred. It is far more flexible in terms of potential length than the span of a single human life. If there was something peculiar about the intrinsic lifespans of kings that caused the pattern, that peculiarity is now rendered irrelevant. So for example the hemophilia argument, already dead, becomes even deader. And yet, the total number of years that a dynasty stays in power still shrinks, on average, as one dynasty succeeds another.



It's still the same puzzle, but with brand-new subject matter: How do the previous X dynasties in country Y impair the prospects of the next one in such systematic fashion? Why should it matter that France lived exactly 341 years under the Capetian kings before the Valois came along? Again, we can discover no universal rules regarding dynasties that would explain the curve. We have advanced a little, in that we see that there is some strong ordering principle at work here. Yet we remain almost as baffled as we were when we plotted individual rulers.

Now, spurred by an alarming intuition, we try plotting all the known dynasties of recorded history, in their chronological order, as one big set. This is a frankly desperate move. Assembling the set of all dynasties everywhere ought—in the orderly classical universe we think we live in—to reveal nothing at all. This disturbing pattern should disappear, returning us with relief and gratitude to the realm of the indifference principle. There should be no trend, no information, no meaningful forecast to be made. But there it is, once again, that spooky spontaneous ordering:



Up to this point, the weirdly consistent downward trend has been confined to a single jurisdiction, with either the rulers or the dynasties coming in strict succession. Whatever the causal factor might be, there is still at least some intuitive sense to it, because it is localized. The early rulers somehow take away stability from the later ones, to whom they transferred power. Or the later ones lack something the early ones had. We don't know what the missing factor is, but up to this point the phenomenon has been relatively well-behaved. It has stayed in one place, applying to just one country at a time.

It's one thing for Babylonian dynasties to decline on the same shape of curve as Chinese ones. But it would be quite another for any broader pattern to exist, reaching out between eras and empires. Babylonian dynasties of 1500 to 1000 BCE can hardly have a specific relation to Chinese dynasties occurring 2,500 years later.

Or at least so common sense tells us. But the graph suggests that they *do* have a relation, and it's a decline of the same slope. The fit is better, in fact, for this larger set than it was for any of the small ones making it up.

This implies that all of recorded history, right up to the 20th century, forms a single integrated decline pattern. The reign length of any ruler, in any country, in any era, can be fitted into this pattern. Reign lengths drop steadily as rank number—the order in which the dynasties occur—goes up. The tenth dynasty to be founded—wherever it might be on the planet—lasted a certain number of years. The hundredth dynasty lasted half that long, on average. And so on, down through more than 5,000 years of hereditary monarchy, to the present.

This global meta-pattern exists in addition to the declines that we observed for individual dynasties, and the rulers within those dynasties. The overall pattern is fractal, in the precise sense of that term—it obeys a scale-invariant geometry. As we 'zoom out' to higher levels of abstraction, the shape of the curve is unchanged.

There are some concerns here about the quality of the data for early dynasties. Have we captured all the major dynasties of history, or only a sampling of them? I address this point in Chapter 19. But setting data issues aside, it is clear now that we are now looking at the sort of mystery that leads to real scientific breakthroughs. *Any* systematic decline would be puzzling and worthy of investigation. A systematic decline that takes the same shape from three entirely different perspectives is more than puzzling. There is just no way that any sort of conventional influence or force can explain all this. Our common sense balks. We search in vain for the underlying rule.

To be sure, there were not very many plausible causes available even when the decline was confined to single countries. We were already fairly baffled. But now even our vague intuitions about the chain of causality must be discarded. Dynasties that were contemporaries in 600 BC would not necessarily have had much contact with one another. In many cases there would have been no trade, no diplomacy, nothing at all. Yet this global pattern,this invisible influence, whatever it is, leaps across oceans and continents and millenia of time as though they did not exist. So far as we can see—on such short acquaintance—the pattern is eerily nonlocal, and strictly mathematical.

For either of these things to be true is mind-boggling. The length of a hereditary dynasty is a derivative value, something that follows from fantastically complex interactions involving economics, law, culture, personalities and family rivalries, warfare, language, ethnicity, and on and on. There shouldn't be a simple rule for predicting it. We have free will, don't we? History isn't supposed to be determined ahead of time. So there really shouldn't be *any* distinctive or predictable pattern to reign lengths, even for one country—much less for the entire planet. Yet the data here are as well-attested as any in science.

We might be unsure about the precise dates for very distant eras. We might even be unsure whether certain kings existed. But we have thousands of years of solidly documented history, thousands of kings whose reigns we know to be real and definitely situated at a specific point in time. The decline pattern extends across all these eras, in each dynasty, then in each set of dynasties, and finally in the set of sets. Both the simple fact of the decline, and the form of the nonlinear, nonlocal rule governing it, are undeniable.

Ever since the Greek writer Herodotus—if not even earlier—there have been two broad contending schools regarding laws of history. One school, including Herodotus, has held that history does in fact exhibit laws and patterns, long-run trends and even repeating cycles. The other school, which includes most modern historians and such influential scientific figures as Karl Popper, insists that there is no consistent pattern, nothing in human affairs that resembles the orderliness of physical law.

I will contend in this book that the essential part of the question is now settled, that Herodotus was broadly correct. We can see in these curves—all the hundreds of examples still to come—a kind of unity that spans and governs entire nations, religions, and cultures, and ultimately all recorded history. There are cycles within cycles, each one objectively verifiable and clearly relevant in a human sense.

But the debate about the nature of history is far from over. This finding leaves us more mystified than ever, with new challenges, because while the decline pattern governs our lives in a very real and striking way, it rests on next to nothing specific *about human beings*.

The governing factor at any given point is the number of monarchies that have existed to date—nothing but mathematics. And yet this pattern really matters. I hope the reader can see that this is not some bit of numerological trivia. At the dawn of history, dynasties could easily last hundreds of years. By the early 20th century, new hereditary dynasties just being launched had an expected span of 52 years—that is, a typical dynasty would not even last as long as one lifetime. At that point, dynastic rule collapsed, the last great empires broke up, and the modern era of democracy and dictatorship began. This curve implies, incredibly, that hereditary monarchy as a human institution *unwound on a predetermined schedule*.

All manner of diverse and contradictory 'laws of history' have been proposed in the past. Whether the patterns they described were a consequence of class or culture or divine providence or the actions of great men, such laws were conspicuously and quite naturally about *human beings*, their values and achievements, their flaws and weaknesses, their divine or evolutionary inheritance of virtues.

For example, consider the long series of writers starting with Aristotle and ranging down through John Stuart Mill, Alexander Hamilton, and H. L. Mencken, who have argued that democracy tends to degenerate into tyranny. This argument makes at least one quasi-mathematical assertion, that there is something about sheer numbers, some law, that causes government to become oppressive and to run amok. When given free rein to vote its will, the social whole behaves in a way that is more (or perhaps less) than the sum of its parts. Now, the principle of maximum entropy says that the whole (all kinds of wholes) really does behave in a way that is more than merely the sum of its parts. But I doubt anyone would ever describe Aristotle's theory of democracy—or Mencken's—as a *mathematical* theory.

Despite their many differences, the intellectuals who have sided with Herodotus and those who have sided with Popper on the question of 'laws of history' would no doubt have agreed that if there really was some universal law of human affairs out there, some pattern waiting to be discovered, it would tell us a great deal about our human nature.

It would be much easier to accept, or anyway more in line with historical practice, if I offered the reader a theory in which regime lengths shrank in proportion to a trend (whether positive or negative) in agricultural productivity, or literacy rates, or religious harmony. Yet the eroding stability of monarchies down through the ages has apparently been a function not of ideas, or material factors, but of simple rank order among the monarchies themselves. The institution of monarchy suffers instead from an exquisitely simple kind of self-interference. The more monarchies humanity establishes, the less stable any of them turn out to be.

This spooky fractal curve just seems to start and end in numbers. There is still a profound insight here, to be sure, but it is surely not the one we expected.

This is the start of our journey. This pattern of everyday events governed by pure number *cannot* be, yet it is. Even more intriguing, it closely resembles the declines observed by George Spencer-Brown fifty years ago. So I will say it here, and proceed to demonstrate it through the various chapters of this book: This is not a law of history, so much as it is a law of randomness.

Understanding why will take us through a vast landscape of scientific history. We will look at literally hundreds of similar curves, some first discovered a century or more ago. We will ponder the partial explanations that have been proposed by scores of scientists, and wrestle with each man's frustrated glimpse of the truth. By the end, we will see this sprawling phenomenon of decline as forming its own kind of unity, as being a thousand applications of a principle laid down fifty years ago by a brilliant man: Edwin T. Jaynes.

Here are a few more examples of this same curve:

— If an army starts marching to some distant goal, on an expedition, its daily progress shows the same sort of pattern. One day it marches 26 miles, the next only 18, then 21, then 15, the numbers varying randomly but continually trending down in just this way.

— If a factory produces more and more examples of a particular good—cars, radios, golf clubs, air conditioners—the labor required per item drops in the same way, again varying from one instance to the next, but converging on a savings of 20 percent for each doubling in the total number of items produced.

— If we compare larger animals with smaller ones, starting at the very smallest single-celled organisms and moving up to elephants and whales, their average metabolic rates per kilogram show the same pattern. Larger animals use less energy in proportion to their body mass. Each doubling reduces energy use per kilogram by 20 percent.

— If we compare traffic accident rates in different countries, we find that the larger the total number of cars present in one country, the lower the *per capita* accident rate. The various rates cluster around a smooth curve with, once again, a slope of about 20 percent.

— If we measure religious participation—how often people go to church, how much of their income they donate, how high they rank their faith in importance—we find the same inexorable decline at work. A Christian church with 50,000 members will have a certain level of participation, and a theologically similar church with 100,000 members will find its participation levels about 20 percent lower.

— As an epidemic spreads, its infectiousness and mortality both decline in predictable fashion. This is true of diseases ranging from hemorrhagic fever (the dreaded Ebola), to antibiotic-resistant staph infections, to AIDS.

Right away we see our first challenge—to determine the proper terminology to use. We can think of the slowing of the marching army as being like a mysterious friction: the men, machines, and animals start out fresh, and then suffer breakdowns or get tired. That is, we can view the decline as a progressive increase in something bad like fatigue. But it cannot actually be fatigue in the literal sense, because the same decline in speed shows up when we increase the *size* of the army instead. Whether the army marches twice as long, or is twice as large, the same thing happens. The most specific we can be is to refer to the pattern as some kind of generalized efficiency loss.

We would logically think of the falling production cost in the factory as just the opposite—an acceleration of the process, a gain in output for a given quantity of labor. This is obviously good, an efficiency gain. Declining traffic accidents and declining mortality due to disease are also by any human standard good. Obviously, a decline in church attendance is good or bad depending on one's religious outlook—but that rule applies to more than churches. Customer loyalty and *per capita* willingness to spend tends to decline according to the same pattern, as the total customer base expands in size. Voter turnout declines as the size of the voting population increases.

Meanwhile, the trend in animal metabolisms seems neither good nor bad. The moral or practical meaning of each curve is different, but the math works the same way whether we regard the trend as positive, negative, or neutral. The common rule is mathematical, not moral.

We will also need to distinguish between what I will call local or 'anecdotal' explanations for specific curves, and the overall pattern of decline. The slowing-down and shrinking lethality of larger armies is actually called 'combat friction' and was first described by Carl von Clausewitz in 1830. The speeding-up of factories is a 'learning curve' and was first documented by Theodore Wright in 1936. The metabolism law is named after biologist Max Kleiber, who wrote about it in 1932, and the most common explanation for it has to do with heat loss and surface area. The decline in traffic fatalities was discovered by R. J. Smeed in 1949, and is attributed to something called 'risk homeostasis'.

Unfortunately, these anecdotal explanations tend to be very unsatisfying. Clausewitz did not really explain combat friction, he simply described it. The learning curve has gone through periods of popularity among manufacturers, but no generally accepted theory exists for why it converges on the 20 percent slope. The explanation usually offered for Smeed's law is that as the road gets more crowded, drivers slow down—but in the form it takes here, the law takes no account whatever of how much road surface is available in the country, or traffic speeds, or *per capita* usage rates. It depends solely on the number of cars, making the 'risk homeostasis' theory rather tenuous. As for the metabolism curve, biologists continue to debate several unsatisfactory hypotheses, but there is no consensus. And of course, there is nothing in any of these specific explanations, however useful they might be in isolation, that helps us explain why all the curves are so alike.

In short, despite more than a century of work in some cases, we do not have a generally accepted explanation for any of these laws. And so far we have no explanation at all for the brand-new regime lengths curve.

This leads to a point about the pitfalls of talking about laws of 'pure' mathematics. There is a subtle distinction here that I hope the reader will keep in mind. This law, whatever it is, takes a common form despite being applied to wildly diverse subject matter. It appears to depend on nothing but number—rank order or total set size. The fact that we cannot discern any other influence on the curve is very important. This is a basic characteristic of maximum entropy curves, one that is implicit in the theory from the start. But that is not the same at all as those influences not being present.

The decline effect is something that emerges from diverse and complex systems, a form of spontaneous order. It is precisely because the system is so rich in detail, so large and complex, that it is possible to generalize about what it will do. The influences sum up (or if you prefer, they cancel out) in a predictable way. To speak of a law of 'pure' mathematics suggests that details are actually negated or erased, that in some sense they cease to matter. We shall see that that is not the case here.

This brings me to a third key point. In all of these laws there is a strong impression of nonlocality, that is, a global linkage of the items across space and time. The spontaneous order that we see is *a property of the whole*. The whole may be social, economic, political, historical, or biological, as the case may be.

To be sure, this nonlocality or wholeness seems a good deal less mysterious in some cases than in others. We are not especially puzzled that an army travels more slowly on its second or third day of marching. We are a good deal more puzzled when having more cars on the roads leads to a lower accident rate. In the case of the marching armies, we could set aside all the talk about spontaneous order and nonlocality, and perhaps not feel that we were missing something vital. In short, some applications of the curve are truly spooky, while others seem mundane.

It is tempting to relegate the less weird-seeming cases as not really being a challenge to our classical model of statistics. But this is a bad idea. The common shape of the curve tells us we are dealing with a continuum of cases. Relatively easy and trivial cases of spontaneous order nonetheless have their place in the overall scheme. The property of nonlocality needs to be considered in relation to all the cases. It resides in the shape of the curve, rather than in the specific circumstances.

Sets, systems, and naked-eye quantum mechanics

To deal with all these laws, we need a theory of how widely distributed systems can exhibit random yet subtly coordinated behavior. That is, how separate and nominally independent parts can behave *as if they know they belong to a whole*. That is what the principle of maximum entropy imposes—a rule governing behavior from the perspective of the whole.

Science has faced this problem before, with mixed results. In quantum mechanics, or QM for short, the behavior of widely separated particles is 'entangled' across large gaps of time and space in a way that has so far been impossible to explain. Before we try and proceed with applying the principle of maximum entropy to our everyday lives, it is worth understanding what has gone wrong, and right, with that previous project.

For example, the decay of a neutral pion (a subatomic particle) produces two very similar photons that shoot off in opposite directions. According to the classical view of physics and statistics, we should be able to measure the precise state of either of the two photons sometime later, and the other photon will remain unaffected. It should make no difference whether we measure photon A first, or photon B. The measurements should be independent.

However, they are not independent. It turns out that once we measure the first photon, the expected outcome for the second subtly changes. We observe no communication between the two particles, and at the moment the measurement is taken, they may be miles apart. The numeric value that we get in the second measurement is altered for no better reason than *because it came second*. The behavior of the second photon is determined by our measurement of the first.

The very possibility that this could occur appalled many physicists. When quantum theory was first developed in the 1920's, a number of figures including Albert Einstein argued strenuously that even though the equations predicted such linkage, this could not be the final answer. Yet after eighty years of study and experiment, the principle of entanglement remains well established, and utterly mysterious. The puzzle has never been solved. We accept the math as working, in the sense of giving valid predictions, but we cannot explain what is happening in physical terms to bring these weird results about.

Quantum theory remains deeply mysterious in part because we have such limited access to the entangled particles. We cannot observe them in any way except according to the rules of QM itself. So we have no direct information—only inferences based on the QM equations—about what the particles are doing the rest of the time. Many physicists have resorted to arguing that the particles are not entirely real, that between measurements they exist only as an unresolved cloud of probabilities.

This is the origin of the famous thought experiment about Schroedinger's Cat: Suppose a cat were confined in a box, and the box equipped with some kind of lethal trap that only fires when a particular atom decays. The decay of the atom is a probabilistic event. So far as we know it is governed by nothing but mathematics. Until we open the box, we don't know if the event has taken place. So under QM theory, while the box lid remains shut, the cat remains entangled with the two alternative states of the atom and thus the cat also exists only as a mixture of two possible outcomes—half-dead, and half-alive, at the same time. Schroedinger found this idea appalling. He devised the thought experiment as a way of protesting the direction that quantum theory was taking. But Schroedinger's Cat has ironically become part of the accepted intellectual structure of quantum mechanics. Critics continue to deride the whole enterprise as intolerably 'mystical,' but as we still lack an alternative explanation, the probability-cloud approach has dominated the field for two generations.

By contrast, this linkage of dynasties across time and space is accompanied by an avalanche of free information. The slowing-down of armies, the speeding-up of assembly lines, the widely ranging metabolisms of animals, and the other hundreds of maximum entropy curves in this book all share this same character. These systems drawn from our everyday reality are all non-locally related—in what is at times a very spooky fashion—but there can be no thought here of treating any of the people and things involved as a mere cloud of probabilities. Everything in the global dynastic 'system' remains unrelentingly real, determinate, and distinct at all times. And yet the fates of the individual monarchs, or individual dynasties, are quantized and entangled in a way that is closely analogous to the pattern we observe in quantum mechanics.

This fact forces us to be creative, to attack the problem in a new way. When we construct a dynasty lengths curve, and discover this eerie coordination of events across millenia of time, we do not have the option of postulating some mystical force, operating on human beings in the same inscrutable fashion as the quantum formalism operates on particles. We already know all the relevant forces, or we think we do. There is no veil of ignorance obscuring our inquiries. The parts of all these large naked-eye quantum mechanical 'systems' are freely accessible to us—it is only their interaction, their sum, that we now realize we do not fully understand.

Apart from not being stuck behind a veil of ignorance, the event patterns we will study in this book are also incredibly diverse. As we have already observed, the same weird behavior, right down to the same slope of curve, shows up again and again in criminology, accident statistics, animal metabolisms, species diversity, economics, warfare, comparative religion, and so on. Everywhere we look, we find the same violations of the principle of independence, describable only in terms of the same rule of increasing set size. We find mysterious linkages, a relation of individual items to the whole, that in a classical world should not exist. This diversity points us to a mathematical solution, rather than a physical one.

Because of the ever-increasing specialization and compartmentalization that prevails in science—the tendency for experts to focus solely on their own field—each case has been investigated independently of the others. In many cases the investigation has stalled, leaving scientists baffled and frustrated, in much the same manner as it in quantum mechanics. Most have therefore wound up languishing in obscurity, stuck in file drawers and marked 'Unknown'. But not all. The sheer diversity of the phenomenon has proven to be an aid to our progress.

By virtue of sheer repetition, the problem has recently been solved—several times—and always in the same way. The answer, say dozens of researchers in different fields, is Jaynes' principle of maximum entropy.

Take for example the impact decline curves have had on biology. Salvador Pueyo, Fangliang He, and Tommaso Zillio are investigators in the field of species diversity—how many individuals of how many species can be expected to share a given area. They recently discovered that maximum entropy can resolve a longstanding controversy in their field. In announcing their discovery, they invoked a vivid image:

> We have shown that common shapes of [species distributions] can be predicted from extremely general assumptions. ... We expect more findings to follow, because we think we have correctly identified the prior distribution [...] which is the Rosetta Stone that allows translating concepts between statistical physics and macroecology.
>
> Even more generally, we hope to have shown that sometimes science can progress without the need of assuming that nature is less complex than it actually is. Of course, there are some simplifications in our approach, but we have moved close to a full acceptance of the complexity of nature, and simple equations have naturally emerged.

Pueyo, He, and Zillio refer to the principle of maximum entropy as a 'Rosetta Stone' that will unlock a whole range of mysteries. The original Rosetta Stone was a plaque discovered in Egypt, an ancient inscription in three languages. That one unassuming object suddenly made translation of Egyptian hieroglyphics possible, because one of the other languages was already known. The Stone opened up an entire field merely by existing.

The principle of maximum entropy has the power to do the same thing throughout the field of ecology. The enthusiasm of Pueyo and his colleagues comes through even in the dry language of a scientific paper. This, they are saying, is an elegant law, a breathtakingly simple law. But the mysteries that they have it in mind to solve are all confined to their specialization. The real scope of the phenomenon is vastly larger. The very same curve shows up in epidemiology, in animal metabolisms, in fetal growth, and in hundreds of other places. Maximum entropy is more than a Rosetta Stone for ecologists. It is a revolution for science. But ironically, because of the way science is done, it seems that few scientists have so far realized this.

The relationship between all of these weird declines and quantum theory is not just a metaphor. It is very literal and immediate. Quantum theory works on the premise that nearby particles form a system, such that the behavior of any individual particle is then influenced by the presence of the others. The terminology we use to explain the decline effect is somewhat different, but it arrives at the same end result. Objects or events belong to a set, and their values are influenced by the size of the set, as well as their rank order in it.

It follows that if we can arrive at an explanation for this kind of 'naked-eye quantum mechanics,' we will gain important insights into the mysterious behavior of sub-atomic particles. We can consolidate our thinking about probability in general, so that for the first time in nearly a century we are able to apply similar rules in both realms. Ultimately we may find that there are not really two distinct realms of causality, but one—that despite many superficial differences, the spooky consequences of entanglement in a quantum system and the weird consequences of membership in a set are really the same thing.

Edwin Jaynes believed in just such a consolidation. Over his long career, the tools of Bayesian inference and the principle of maximum entropy showed surprising predictive power in endless non-quantum problems. Jaynes believed they could reform quantum theory as well.

In a talk given shortly before his retirement, he expressed his confidence that his approach would at some point in the future do away with the 'probability cloud' model of existence.

> I am convinced, as were Einstein and Schroedinger, that the major obstacle that has prevented any real progress in our understanding of Nature since 1927, is the Copenhagen Interpretation of Quantum Theory. This theory is now 65 years old, it has long since ceased to be productive, and it is time for its retirement ...
>
> [P]resent quantum theory claims on the one hand that local microevents have no physical causes, only probability laws; but at the same time admits (from the EPR paradox) instantaneous action at a distance! Today we have in full flower the blatant, spooky contradictions that Einstein foresaw and warned us about 60 years ago, and there is no way to reason logically from them. This mysticism must be replaced by a physical interpretation that restores the possibility of thinking rationally about the world.
>
> We see the effects of this in the fact that today, a large portion of research in theoretical physics has been reduced to wheel-spinning; random fiddling with the mathematics of the old theory, without giving a thought to its physical foundations. One would think that the folly of this might have been learned from the example of Einstein; yet his repeated warnings go unheeded even as his worst fears are realized before our eyes.
>
> I believe the answer to this must be that our present formalism contains two different things. It represents in part properties of the real world, in part our information about the world; but all scrambled up so that we do not see how to disentangle them. But at least we can see that the spooky things will cease to be spooky as soon as we think of the formalism in terms of inference from incomplete information. *Then what is traveling faster than light, and backwards in time, is not a physical influence; but only a logical inference.*

The consolidation that Jaynes hoped for is in fact well underway, and many writers have already glimpsed parts of the road ahead. The reader need not take my word for any of this. Throughout this book I will cite the relevant papers, and quote the investigators. We will build up a picture of science advancing like a network of underground streams, each one remote from the others, all gradually converging and finally bursting forth from a mountainside, launching a mighty river of discovery.

The superiority of entropy as a unifying explanation

Science has taken its time in arriving at this point. One sees in the literature a great many insightful essays and books about how many power law curves there are in nature, and how similar they are to one another, and how important they are to science. The work of Benoit Mandelbrot on size laws and fractal dimensions, dating back to the 1960's, has also found a vast and appreciative audience. Richard Koch published several widely read books on the subject, including *The 80/20 Principle*. Nicholas Tassim Taleb has written several popular works as well, such as *Fooled by Randomness*. Another worthwhile title is *Ubiquity* by Mark Buchanan.

There is a quiet but widespread suspicion that science has been on the wrong track, that our understanding of system dynamics and the nonlinear character of the universe remains partial and flawed. The paradoxes of quantum mechanics stand out as an especially prominent clue that our mathematics are in need of fundamental reform.

Unfortunately, the work of actually carrying out the reform, and integrating all of these odd *ad hoc* laws is daunting. So the explanations to date have been piecemeal and metaphorical. Authors tend to take some particularly well-modeled case and then suggest, in optimistic fashion, that it can be extended to cover a large part (but never all) of the remaining anomalies. Enthusiasts talk about chaos theory, self-organized criticality, strange attractors, fractals, and so on, awed by the sheer prodigious variety of forms and mechanisms. In physics, if the discussion turns to the need for a new unifying concept or framework, the talk is likely to be about string theory or loop quantum gravity. But there is no longer much if any talk about a *global* solution, a comprehensive framework of ideas that would reunite the classical and quantum realms. No single model satisfies everyone, or even looks like it could satisfy everyone. Playing with fractal math or strange attractors can produce splendid 3-D computer graphics, but so far they have not produced a real scientific revolution. Meanwhile, despite attracting some of the most brilliant minds in science, in a generation of work string theory has yet to produce a single testable hypothesis.

Edwin Jaynes addressed this impressive but bewildering diversity of tools in his retirement talk. He believed that the popularity of such ideas was a symptom of the slow, piecemeal advance toward a single truth. He expected that truth to be best expressed by a comprehensive scheme of Bayesian analysis plus the principle of maximum entropy. Other ideas simply did not, in his view, offer the same generality and power.

> In recent years the orthodox habit of inventing intuitive devices rather than appealing to any connected theoretical principles has been extended to new problems in a way that makes it appear at first that several new fields of science have been created. Yet all of them are concerned with reasoning from incomplete information; and we believe that we have theorems establishing that probability theory as logic is the general means of dealing with all such problems. We note three examples.
>
> Fuzzy Sets are quite obviously, to anyone trained in Bayesian inference, crude approximations to Bayesian prior probabilities. They were created only because their practitioners persisted in thinking of probability in terms of a 'randomness' supposed to exist in Nature but never well defined; and so concluded that probability theory is not applicable to such problems. As soon as one recognizes probability as the general way to specify incomplete information, the reason for introducing Fuzzy Sets disappears.
>
> Likewise, much of Articial Intelligence AI is a collection of intuitive devices for reasoning from incomplete information which, like the older ones of orthodox statistics, are approximations to Bayesian methods and usable in some restricted class of problems; but which yield absurd conclusions when we try to apply them to problems outside that class. ...
>
> The great new development is Neural Nets, meaning a system of algorithms with the wonderful new property that they are, like the human brain, adaptive so that they can learn from past errors and correct themselves automatically (WOW! What a great new idea!). Indeed, we are not surprised to see that Neural Nets are actually highly useful in many applications; more so than Fuzzy Sets or AI. ...
>
> But, do we really need to point out that 1) Any procedure which is adaptive is, by definition, a means of taking into account incomplete information; 2) Bayes' theorem is precisely the mother of all adaptive procedures; the general rule for updating any state of knowledge to take account of new information; 3) When these problems are formulated in Bayesian terms, a single calculation automatically yields both the best estimate and its accuracy; 4) If nonlinearity is called for, Bayes' theorem automatically generates the exact nonlinear function called for by the problem, instead of trying to construct an approximation to it by another ad hoc device.
>
> In other words, we contend that these are not new fields at all; only false starts. If one formulates all such problems by the standard Bayesian prescription, one has automatically all their useful results in improved form. The diculties people seem to have in comprehending this are all examples of the same failure to conceptualize the relation between the abstract mathematics and the real world. As soon as we recognize that probabilities do not describe reality—only our information about reality—the gates are wide open to the optimal solution of problems of reasoning from that information.

Over the past century, as these anomalies have accumulated in various file drawers, science has not taken the time to consider them in context, and assign a rigorous, comprehensive explanation for them all. A more complete and satisfying answer has meanwhile been slowly emerging, constructed piecemeal by admirers of Jaynes, but this has happened in a much less visible fashion than the popular work done on fractals and chaos theory and string theory and so on. The rigorous answer, the answer broad enough to alter our view of every scientific discipline, is Jaynes' principle of maximum entropy.

Here, then, is the project I propose to tackle: To show how the principle of maximum entropy governs everyday life, how it radically revises the classical theory of probability, and how it can perhaps unlock the quantum world as well. The project is large enough that it requires two volumes. This volume deals with the world of observation and natural science. A second volume, entitled *Experiments in Decline*, will contain a more detailed presentation of the theory, as well as results from hundreds of experiments. It will show the reader how to reproduce the decline phenomenon in the laboratory. That volume will also deal with the potential impact on quantum mechanics.

Where my own journey began

I was not converted to Jaynes' view overnight. The theme of my original investigation was relatively simple, and did not rely on Jaynes at all in the beginning. For more than a decade, from 1987 to 1999, I researched diminishing-cost and diminishing-return curves in a variety of fields, in hopes of better understanding a huge range of everyday decline phenomena.

The goal was a single formula or principle that would compactly explain, among other things:

— Why small armies are more effective per man than large ones;

— Why manufacturing costs steadily drop as the number of items produced increases;

— Why large computer networks are less efficient than small ones;

— Why epidemics become less deadly as they spread;

— Why large animals have far slower metabolisms than small ones;

— Why there are so many more small towns than large cities;

— Why all kinds of processes from wars to telephone calls obey a logarithmic law of durations;

And so on. This was the decline effect that I was originally interested in. The idea of refuting or even so much as modifying the classical theory of probability, whether using Jaynes' model or any model, did not enter my mind in the beginning.

I was hoping to define a specific power-law decline curve that would explain, or at least compactly summarize, all these tendencies. Eventually, I did arrive at some concepts that covered an enormous diversity of cases. However, the model was not complete. I also wanted to clearly set down the circumstances where these rules applied, and where they ought not to apply. As late as 1999, I still had no clear answers to these questions.

My original somewhat vague hypothesis was that the decline was the result of complexity, and was closely related to what the German military theorist Karl von Clausewitz identified in his 1830 book *On War* as 'friction'.

Friction, according to Clausewitz, was a general tendency for the performance of armies in the field to fall well below what was theoretically achievable. Since the appearance of *On War,* the concept had gradually become well-known, in some ways almost a cliché of military philosophy.

Friction was what wrecked the well-laid plans of generals. It was the sand in the gears of the military machine. As Clausewitz put it:

> Everything in war is very simple, but the simplest thing is difficult. The difficulties accumulate and end by producing a kind of friction that is inconceivable unless one has experienced war.

In its original formulation, the concept of friction was not mathematical. Clausewitz was not a great believer in mathematical analysis. Friction was accordingly taken by most of his readers as more of a philosophical commentary on how the real world continually fell short of the ideal.

In 1987, historian Trevor Dupuy published data showing that declining performance by armies on the battlefield could be plotted as a straight-line 'log-log' relationship—for a given increase in the size of the army, there would be a consistent percentage drop in effectiveness. Dupuy believed that Clausewitz had underestimated the potential of the concept, that it could be quantified and analyzed. After reading Dupuy's work, I went on to find that the same relationship—right down to the specific slope of the graph—showed up in many other sets of data. This one finding, more than any other, defined my project in the early years.

It seemed clear to me from the start that this 'friction,' whatever name we might call it by, was essentially mathematical. Accumulating large numbers of similar entities or events in certain contexts predictably lowered the average value of each one. This pattern could be found in such a diverse range of situations that only a mathematical rule could cover them all.

In the beginning I thought of the underlying cause much as Dupuy and Clausewitz did, as a form of mutual interference between parts of the military machine, but with the added proviso that it was not limited to warfare. My belief was that this friction-due-to-complexity effect could be fit into the framework of classical probability, and that it would constitute an interesting footnote to existing law. I made five broad assumptions:

1) Decline requires a certain minimum threshold of complexity.

2) Decline is a very gradual, incremental, strictly monotonic process.

3) Decline is always scale-invariant.

4) Each case of decline is a *causal* process that proceeds from a single point of reference and spreads outward.

5) Decline is ultimately the result of physical laws, but which laws is at present unknown.

However, by 1999, each of these assumptions had been challenged by contrary data. My initial conviction that the decline effect could be fit into a strictly classical framework was wavering.

The gradual, long-term decline processes that I had begun with had led me into strange territory. I was observing results that none of the authors I had studied would have expected—patterns that hinted at a much more subtle and far-reaching law. I reached a point where it seemed necessary to consider the possibility that classical physics and probability as I had been taught them did not apply, and to explore a much wider, more radical model.

Starting in 1999, I began testing simulations and games of chance, looking for decline where theory said there would be none. Even at this late date I was still expecting classical theory to stand up. To my amazement, I found that classical probability did *not* work as theory predicted. A decline effect strikingly similar to the Clausewitz-Dupuy curve showed up in the behavior of dice, coins, cards, slot machines, and random number generators.

It was at this point that I turned seriously to the idea of entropy as an explanatory principle. I applied mathematical tools that I recalled from my undergraduate studies, and they seemed to fit the strange phenomena I was seeing. I also began reading in more depth about Boltzmann, Jaynes, and the history of thermodynamics. In due course I heard about recent successes following in Jaynes' footsteps, such as the work of Denis Evans and his colleagues in Australia. I began modeling my arguments as a variant of the principle of maximum entropy.

In effect, my 22-year investigation can be divided into two eras—the first 12 years dominated by the original, somewhat vague 'friction due to complexity' argument, and the shorter second era, which was characterized by a more systematic and radical approach, increasingly focused on Bayesian inference and the principle of maximum entropy as the cause of all these myriad decline effects.

To save the reader from sharing in the more frustrating and baffling parts of my journey, the account does not follow a strictly chronological sequence. Instead I start with some basic questions about how events are distributed in nature, and the work of early theorists like Pareto and Zipf, and then take on specific subject areas one by one.

So here is an outline of the path I took.

1. I began by asking—is it possible to find mathematical patterns in historical events?
2. I found that historical events tend to obey a bewildering variety of logarithmic or power laws, while relatively few follow a normal or bell-shaped curve.
3. On closer examination, many of these power laws turned out to be either connected, or equivalent.
4. The common factor connecting them is long-run decline—as the data set grows, individual events or transactions become smaller, with fewer extremely large or extremely small items. (The technical term here is that the distribution is *non-stationary*.)
5. The long-run decline pattern is also 'noisy,' not smooth and continuous, when studied in detail.
6. The 'noise' turns out to be a separate, localized, nearest-neighbor decline effect.
7. A very similar pattern combining long-run and short-run declines shows up in ESP research, and in other contexts in which there is no physical reason for it to occur.
8. The same pattern of long-run and short-run decline shows up in the workings of games of chance, random-number generators, and other phenomena that have up to now been believed to obey classical probability laws.
9. An explanation that covers both the long-run and nearest-neighbor declines, across the entire range of events observed, is increasing entropy.
10. From this we can conclude that the classical probability is wrong, and that Jaynes' concept of entropy governs all kinds of events.

This was more or less the historical sequence of my investigation, from 1987 through 2009, and it makes a reasonable framework for me to unfold the logical development of the ideas.

Basic anatomy of decline, random systems and sets

Throughout this book we will be looking at decline as a behavior of a random system. Our quantitative analyses and graphs will be of data sets generated by those systems.

The random systems we will study will be complex 'real world' phenomena—national traffic accident rates, district voter turnouts, variations in church donation rates with church size, the efficiency of bombing raids of varying size, animal metabolisms, and so on. Then in *Experiments in Decline*, I will focus on the simpler systems that form the foundation of classical probability, such as coins and dice, and the much weaker and subtler trends that do occur in those systems.

In truth, a random system could be virtually anything—a developing fetus, a spreading epidemic, a factory manufacturing automobiles, a forest ecosystem, or the entire set of planets circling the stars of the Milky Way. Random variation is everywhere in our world.

From the start, we can assign examples of the decline effect to one of three system types: fixed size, expanding, and time-invariant.

— *Fixed size (t-type decline)*. In this case the physical system we observe is of a fixed size, for example a single six-sided die or a single slot machine. For a system of fixed size, the only open-ended dimension is time—we throw the die or pull the slot machine lever, and generate a new outcome, and then another, and in that fashion build up a series of random values. The size of the system (denoted by N) is fixed, so it can be ignored. The set of results only grows with increasing time t. (There is a special case here to be addressed later, when the system size is fixed, but individual system elements like dice are periodically replaced. These systems are not actually t-type, but N-type or Nt-type, as described below.)

— *Expanding (Nt decline)*. A system can also increase in size. For example, if I test a whole series of dice or slot machines, either in parallel (all at once) or serially (one after the other), the total number of events has not only a time dimension but also a system size dimension. The set of results grows in proportion to N and also in proportion to t.

— *Time-invariant (N decline)*. It is possible to study system behavior from the point of view of size, while ignoring the effect of time. For example, we might throw different-sized groups of dice, or analyze the behavior of different-sized groups of people, and compare how often certain rare events occurred for different group sizes. The order in which the various events occurred is ignored in this case. The analysis does not depend in any way on t. It only considers N.

This book will consider numerous cases of N, t, and Nt-type system declines. What will quickly become apparent is that N, t, and Nt-type declines are alike in a fundamental way. This is the most important generalization one can make about the problem: Decline is ultimately a *mathematical* phenomenon, not a physical one. It is mainly dependent on *set* size, whether the set is of the N, t, or Nt type. We are fundamentally concerned with two numbers: how many possible states the system has, and how many states it has been in so far.

The t-type declines are different from N-type or Nt-type, because a small system of fixed size eventually plays out all the possible alternative states and starts to repeat itself. It necessarily finds equilibrium at some point. By contrast, a system that is expanding in size will not achieve equilibrium, no matter how long we wait. The set of possible states will continually grow, and will outrun the cumulative total of events so far. In *Experiments in Decline*, I will offer numerous examples of small, simple systems that decline for a while in t-type fashion, then reach equilibrium.

The N-type and Nt-type declines are more important for this reason. In this book, we will focus mainly on open-ended systems that just keep declining. The real world is filled with them.

The subtitle of this book is: *The law behind diminishing returns and wildly varying outcomes in markets, politics, culture, religion, disease, and war*. That is an easy title to glance at and understand. Diminishing returns constitute a familiar idea found in many fields. But this doesn't quite capture the full scope of the investigation. Sometimes instead of returns, costs diminish. Other times it is something else entirely.

The subtitle *Why rare events become rarer as system size increases*, though it would have been a little more technically precise, would not be so reader-friendly. At this point I ask the reader just to keep in mind that (a) there is a separate story to be told about the pure mathematics of non-equilibrium probability, (b) that we are not limited to considering returns, but all kinds of measurable phenomena, and (c) that these various rare phenomena become rarer over a variety of different dimensions, and not just over time. Time is perhaps the most important case from our human perspective, but we will deal with much more.

As I noted earlier, many examples of decline considered in this book are already known to science, but compartmentalized into individual, *ad hoc* 'laws' of little generality. The fact that all these laws involve a decline according to a single consistent pattern simply was not recognized as important until now. Here are some better-known examples from this first group that have been named for their discoverers:

— Nature exhibits a strong bias toward logarithmic size distributions, with far more short rivers than long, far more small islands than large ones, far more small towns than cities, and so on. (Benford's Law)

— The metabolic rates per unit mass of different living species ranging from single-celled organisms to elephants and whales decline by 20 percent with each doubling in total mass (Kleiber's Law)

— The cost and time required to produce virtually any kind of good or service tends to decline by an average of 20 percent for every doubling in total output (Wright-Henderson Law)

— Armies, navies, and air forces all decline in casualty-inflicting ability by an average of 20 percent for each doubling in numbers (Clausewitz-Dupuy combat friction law)

— In competition for scarce goods ranging from golf tournament titles to enemies shot down in aerial combat, success and failure are both self-reinforcing, such that past winners win more easily and past losers lose more frequently (the Lotka curve)

— The tendency to commit crimes is very strongly related to age, peaking around age 20 and declining exponentially thereafter (Hirschi-Gottfredson age-crime law)

Some of these are very clearly t-type declines, such as the age-crime law. The 'system' in the age-crime law is an average criminal. The set consists of how the average criminal spends his time; in this scheme, law-abiding behavior is common, and the rare events are crimes. The decline pattern (fewer crimes as the average criminal gets older) is thus a relatively simple t-type decline.

Other laws in this list are just as clearly N-type declines, such as Kleiber's metabolism law. In arriving at his law, biologist Max Kleiber averaged out the differences between individuals, and even between species. The order of his observations also did not matter. All he was looking for was a relationship between metabolism and body weight. The set size in this case, as it turns out, is the number of cells in an animal's body. The rare event occurs when an individual cell metabolizes nutrients. The more cells in the body, the less frequently this is done, and thus we have an N-type decline.

Some laws, like the Clausewitz-Dupuy combat friction law, are Nt-type. What Dupuy discovered was that if you ignore time (that is, control for it in the scientific sense) and just look at army size, combat becomes less intense as army size grows. But he also found that if you control for army size, combat becomes less intense over time. A bigger army fights more slowly than a small one, but *all* armies fight more slowly on Day 2 than on Day 1. Hence we have an Nt-type decline, which combines the two variables.

These are the three main ways in which the behavior of a system leads to the generation of a growing set. In some trickier cases it may not be immediately obvious which type of decline (t, N, or Nt) we are dealing with. But in every case of decline we can say, axiomatically, that there is a growing set of results and a decline, as the set grows, in rare values.

The same axiomatic rule is true of many more decline phenomena that have managed to go undiscovered, or at any rate have not been assigned a formal name, such as these:

— Viruses and other disease agents from yellow fever, cholera, and plague to Ebola and HIV-AIDS transmit themselves less efficiently and are less lethal as an epidemic spreads. That is, rare events of transmission or death grow rarer as the number of infected grows (N-type decline).

— Arrays of microprocessors produce lower rates of net output per processor and unit time as the number of microprocessors increases (Nt-type decline)

— Human organizations ranging from religions to political parties to restaurant franchises grow more slowly as they get larger, varying smoothly along the same logarithmic curve as epidemics (N-type decline).

Some of these examples involve declining inputs (what we would regard in our human context as costs) and others involve declining outputs (what we view as benefits). There are also in-between cases that are not easily characterized as either costs or benefits, inputs or outputs.

The universe does not favor declines in any particular category, since an input in one context is quite often an output in another. But these are all examples of a decline pattern in which rare events (often these are also large events) become even rarer as the set grows larger. As can be seen from the descriptions above, many are of the expanding-system type—the amount of decline increases as the number of soldiers or cells increases. Some are of the fixed-system type—performance for one particular object or individual changes over time.

Spencer-Brown and the failure of probability

Here I will return to the work of George Spencer-Brown. Although he does not play quite as large a role in this project as Edwin Jaynes, his ideas and his experimental findings are nonetheless essential.

In the 1940's and 1950's, before computers were generally available, anyone wanting a source of random numbers relied on published tables. A variety of mechanical 'chance machines' had been developed to generate random numbers for this purpose. One technique was to inscribe the rim of a spinning wheel with the digits 0-9 and then use a strobe light to pick out one particular value. As Spencer-Brown observed, the results from such machines were in their way just as unsatisfactory as the elusive correlations found in ESP experiments.

> Anyone who has worked with chance machines knows very well how difficult it is not to observe certain oddities in their behavior; it is only that classical probability, not having a place for them, has always prevented our talking about these oddities in terms of chance. The fact that the machine does something noticeably improbable has not been connected with the fact that it *logically* cannot do anything else.

This was more than a technical debate about the shortcomings of specific machines. The decisive argument, according to Spencer-Brown, was not that random number generation always turned out to be biased. It was that ESP experiments and 'pure' random number tests would both generate the same bias pattern—decline in rare items.

The inference to be drawn was that such declines are somehow inherent to probability. We are surprised by them, and regard them as proof of some mysterious power, only because our concept of randomness is flawed. He later wrote about this conclusion:

> This is quite plausible in light of the fact that psychical research is perhaps the only present-day science which has looked for something (not already known to exist) for sixty years and failed to find it; and if it happened that what it was looking for did not exist, we should have in effect sixty years of pure probability experiments which there is no reason to suppose should have fared, in terms of significance, better than the best (and the worst) of all the pure probability experiments down the ages. It would thus be its remarkable additions to our experimental picture of pure probability for which we owe the most thanks to modern psychical research.

This was a bold argument. In effect, Spencer-Brown argued, ESP experiments had failed to discover any psychic phenomena, but they had nonetheless led to a far-reaching consequence—they had exposed the shortcomings of classical probability. Spencer-Brown laid great emphasis on two points, both of which have lasting relevance for us.

First, our standard of randomness is inherently contextual—our level of surprise at an event varies with the length of the series we are looking at. If we start up a binary 'chance machine' and the first 16 digits it produces are 1111111111111111, we will strongly suspect the machine is biased. If on the other hand we run the machine for 100,000 digits, and the same string shows up somewhere halfway through, we judge that this is no more than normal chance expectation. The very same behavior takes on a different meaning.

The rule here is simple. A level of randomness cannot be defined—that is, we cannot decide if something is genuinely random, or not—without also specifying the length of the specific series under consideration, and the type of event we are looking for. In classical probability, because of the presumption that events are strictly independent, the context in which an event occurs winds up being thrown away. For Spencer-Brown this constitutes a huge conceptual error. There are no independent events, only events in context. A sound theory of probability must specify a context for each event. Set size therefore matters.

Second, our sense of surprise (our decision whether an event is improbable) is at its foundation self-inflicted, even arbitrary, because bias in a random series is unavoidable. There are myriad different compound events (or what Spencer-Brown called 'molecular' events) that can be constructed out of a series of simple (or 'atomic') events. It is mathematically impossible for all these different compound series to meet a fixed standard of randomness *at the same time*. Thus we are likely to wind up 'surprised' as a consequence of our own choice of what particular series to measure.

Suppose we aim to construct a random number table in which every primary element appears with equal frequency, so as to maximize uncertainty about their relative likelihood of appearance. We don't use a chance machine. We try instead to deliberately design an optimally random sequence. It is pretty obvious that if there are N possible elements, this means that the length of the table has to be an exact integer multiple of N. Most possible table lengths are thus ruled out.

Now suppose that we also want to ensure that there is no under- or over-repetition from one simple event to the next, so that we are also maximally uncertain what event comes next. This narrows the range of acceptable values still further.

If we choose N=6, there are six possible simple events (perhaps labeled 1 through 6). We can easily work out that the table can only be unbiased in repetitions if its length is some exact multiple of N x N = 36. That way, all the possible permutations—from 1 followed by 1, through 6 followed by 6—occur in equal proportion.

It follows that if we want to have maximum uncertainty about what event will come next, we have to accept much less than maximum uncertainty about the overall proportions.

These may seem like minor points, and easily overcome. For example, if the table was large enough—thousands of entries—the fact that it must be divisible by N, or by N^2, would not be very important. But from here the problem only gets worse.

Even a table with thousands of entries will unavoidably lack many specific results. These will include not just outcomes like 11111, but more subtle strings like 1313131, or 12345, that hypothetically could happen but simply didn't in the construction of this particular table. The probability in classical theory of each such string is a very small number, but a static table assigns them all a probability of zero. They are eliminated from further consideration.

Also, if we choose to maximize uncertainty by making each element show up equally, that decision ensures that prospective users will know precisely how many times each single event—and each two-element compound event—will occur.

Faced with these problems, and wanting genuine randomness (whatever that might mean) we might try a different strategy. We might let the chance machine run freely, and deliberately avoid correcting any imbalances as we generate the table, with the effect that perhaps no result will appear exactly 1/6th of the time. This prevents the card-counting type strategy. We don't know in advance how many times each outcome will occur. It also means that a result like 11111 at least *might* occur, whereas on a deliberately constructed table it would tend to get discarded.

But as makers of random number tables realized, if they chose that route, it only guaranteed that the table as a whole would take on a kind of permanent, frozen bias. *Every* subsequent user of the table would get too many 6's (for example), or would get the improbable 11111 string at some point.

A spontaneous lack of predictability when the table was *generated* would be anything but spontaneous when it was later *applied* to hundreds of successive experiments or applications.

For example, picture being stuck in a snowbound cabin with a collection of board games, but no actual dice. The only randomizer on hand is a printed table that happened to have been generated with too many '6' results. Every day you play using the biased random number table. After a few sessions of *Monopoly*, *Risk*, or backgammon, it becomes obvious that the perpetual bias in the table is leading to a bias in how you play the games.

In short, creating a series that will look perfectly random *even in retrospect* is an exercise in frustration. Our view of the data set necessarily changes as the set grows. The range of possible 'molecular' events that we will consider is open-ended, a matter of our discretion, rather than something set by nature. The idea of something being 'classically random' has meaning only as a very loose abstraction, when we focus on prospective events—and then only if we stick with 'atomic' event series and don't look too hard at the higher-level compounds they invariably lead to. All real-world series—which includes any and all actual experimental outcomes, as opposed to merely prospective ones—are inherently biased to some pattern.

> The essence of randomness has been taken to be absence of pattern. But what has not hitherto been faced is that the absence of one pattern logically demands the presence of another. It is a mathematical contradiction to say that a series has no pattern; the most we can say is that it has no pattern that anyone is likely to look for.

This notion of categorizing bias patterns as not being objectively random, but merely more or less interesting to a given observer, and therefore more or less 'likely to be looked for' is hugely important, as we will see. A similar notion forms the foundation for Edwin Jaynes' arguments about the principle of maximum entropy.

Spencer-Brown viewed the entire enterprise of probability as compromised by this paradox. In the 1950's, it was an open secret that any effort to publish a table of random values involved a degree of editing and tinkering.

> From whichever end we look at it the dilemma is the same. In order to give a practical interpretation of probability theory for scientific purposes, we have to assume the primary randomness of molecular events; but the moment we do this our random series contains limitless possibilities of predictable repetition which we cannot call 'random' in any ordinary sense of the word; when, therefore, one of these possibilities in our random series begins to be realized, we do everything we can to stop it by fiddling with the machinery, and to hush it up by suppressing its publication as such. Sir Ronald Fisher and Dr. Yates, having produced for publication some random numbers which failed to pass the test, altered them until they did. Professor Kendall and Babington Smith took the other course and suppressed 10,000 of theirs.

> It becomes obvious, then, that the concept of randomness, instead of growing more satisfactory in the consideration of longer series, tends instead to grow less so ...

Spencer-Brown believed that these criticisms demonstrated 'that the concept of probability used in statistical science is meaningless in its own terms.'

There is an obvious rejoinder available, that while these are real paradoxes, such paradoxes by themselves do not actually render our daily uses of probability theory incoherent. Even in 1957 that was perhaps overstating the case, and in some respects, we have come a long way since. For example, today we mostly use electronic pseudo-random number generators, not static tables, so every number generated is prospective and none is re-used. The problem of inherent bias in random number *tables* has more or less gone away entirely.

However, what remains generally true fifty years later is that modern random number theory has not produced anything describable as 'perfect' randomization. We still are not able to say just what 'perfect' randomization means, if anything. The concept is elusive.

Here we return to our baffling regime lengths pattern, for it is very relevant. In the experiments examined by Spencer-Brown, typically the bias pattern observed was one of long-run decline, across the length of the experiment. But he also observed that for long experiments conducted in parts, *session decline*, so common in ESP experiments, showed up in his experiments as well (italics in the original):

> The gradual decline of a particular scoring tendency throughout an experiment can be explained by a slowly changing bias in the randomizer . . . But the repeated decline in scoring *within an experimental unit*, which is so common in psychical research and which has now been found in randomized data which were prepared with no thought of psychical research in mind, is not so easy to explain. It would seem that the randomizer tends to a bias which changes periodically, and that the period is somehow pulled into step with an arbitrary experimental unit. That the tendency is something inherent in the randomizing set-up and not a result of psychokinesis now lies, I think, beyond reasonable doubt. We seem to lack only a detailed explanation.

We can think of the regime lengths pattern as a kind of probability experiment. Then the dual pattern of long-run decline combined with session decline is the same in regime lengths as it was for the ESP experimenters, and for Spencer-Brown. We are again seeing what they saw: nested levels of decline, a fractal hierarchy. Spencer-Brown was evidently on to something that was real, reproducible in the laboratory, and pervasive in nature.

During the past fifty years, probability theory has done little but harden into the very practices that Spencer-Brown condemned. In practice, today's random number generators focus solely on achieving first-order uniformity across some interval—generally between 0 and 1. The relative quality of such a generator is measured in terms of how many different first-order results it can generate before repeating itself. This is a deliberately narrow criterion. It is in fact the same narrow criterion Spencer-Brown criticized in 1957.

Depending on the method we use, various patterns of bias may well exist in the higher, 'molecular' series built by compounding single results. Some effort is made by designers and users to look for these, and if such a systematic pattern is detected, that fact carries some weight. Results demonstrating obvious patterns can be tossed out, in a manner very similar to what Spencer-Brown describes above.

Yet no one proposes to test for, much less eliminate, *all* such patterns. A designer might check for correlation between every second item, or every third. He might even go so far as to test for any correlation between every 10th item, or every 20th. But clearly one can go on postulating new patterns of correlation indefinitely. The only guarantee that modern users of random number generators get is for a series exhibiting uniformity across *the range of primary outcomes*. The 'molecular' combinations, the compound events, come with a hopeful assumption—that they will obey classical statistics as well—but no guarantee and no reliable test for compliance.

Today's modern random number generators continue to produce the same long-run declines that Spencer-Brown observed in the mechanical chance machines of 50 years ago—the same magnitudes of deviation, the same puzzling swing from above chance to below it. I have numerous experiments as well as large amounts of well-attested third party data, all to appear in *Experiments in Decline*.

Yet we are if anything less aware of the issue than we were in 1957, because modern test protocols (such as the Diehard test battery, and the modern NIST standard) are incapable of recognizing long-run trends of any kind. They divide up a long string of data into blocks of no more than 1,000 bits, then analyze each block in isolation. The long-run decline of rare items cannot be detected in this fashion. Even session declines are invisible. Science has by no means resolved Spencer-Brown's criticisms. Instead it has largely abandoned the issue.

Despite the depth of its arguments, *Probability and Scientific Inference* was still more like a shot fired across the bow of science than a conclusive proof. It left major problems that were only sketched out, or left entirely unresolved, as the author conceded:

> What must be done in the field of Probability is to replace an old metaphysical system with a new one. This means undertaking two tasks: first, to show where old assumptions were inadequate; and then, to work out better ones. Each of these tasks is a major one, and I do not think any purpose would be served in attempting them both at once.

Spencer-Brown's critique of probability led to a brief controversy, especially among enthusiasts of ESP, who understandably did not want their entire field of endeavor reduced to nothing but some kind of defect in probability. However, as no one stepped forward to offer an alternative model to explain the anomalies, Spencer-Brown's philosophical challenge wound up being largely forgotten. He later went on to write the highly regarded *Laws of Form* (1969), a philosophical work on logic. He remains famous today for the latter book.

The key question that Spencer-Brown needed to answer, and could not, was: Why decline?

He knew this was important, and made a strong effort at answering this question. He gave careful consideration to unconscious bias patterns imposed by the experimenter on mechanical chance machine. However, this kind of bias is short-lived, as it tends to wind up 'disintegrated' when it is noticed. It also cannot be applied to strictly autonomous chance machines, like pseudo-random number generators. The first computer-generated random number tables were just becoming available in 1957, and they exhibited the same kind of long-run trend. The shape of the decline, the relation of the decline to set size, and the tendency to systematically decline below chance, were also left unexplained by Spencer-Brown's argument.

The clue that allows a solution to the mystery is that the decline observed in ESP experiments and elsewhere is *logarithmic*. It has a specific, constant, power law form.

That one word brings together all the other clues in an orderly framework. A logarithmic decline immediately points to entropy as a cause, and suggests a general law is at work rather than a collection of *ad hoc* biases. It reinforces the idea that the ultimate cause is mathematical, and that it is closely tied to the size of the set, rather than some other factor.

Edwin Jaynes and Bayesian inference

The same year as *Probability and Scientific Inference* appeared, Edwin Jaynes published a brief article entitled 'Information Theory and Statistical Mechanics'. Though the article would not be widely discussed until the 1970's, and the full implications would not be apparent for many years after that, this work would eventually point the way to a new model of probability, one that could meet Spencer-Brown's challenge.

Jaynes was a versatile practical scientist, equally at home in theory and in experiment. During the Second World War he had worked on radar systems; after the war he studied electrodynamics under atomic physicist Robert Oppenheimer.

His greatest accomplishment was to show the superiority of the Bayesian approach to probability— as opposed to the more common 'frequentist' interpretation—in statistical physics. We cannot deal with the full scope of that debate, but we will touch on some essential points.

Bayesian probability begins with the investigator making an assumption about the possible states of a system. This set of prior probabilities—informally just called the 'prior'—is the foundation of the argument. As new information becomes available about the system, the investigator compares it to his prior, and arrives at an updated, *posterior* set of probabilities. These can then become the 'prior' for the next round of analysis.

So for example, a doctor may want to determine if a patient is suffering from a very common and mild illness, or a rare and dangerous one that presents similar symptoms. There is a blood test that will reveal the dangerous condition, but it is not perfect. In 98 percent of cases, if the dangerous disease is present, the test will be positive. However, in 1 percent of cases, if the disease is not present, the test will also be positive. We have a total of four possible outcomes:

- The patient has the disease, and the test is positive.
- The patient has the disease, but the test is negative.
- The patient has a head cold, but the test shows positive for the more dangerous condition.
- The patient has a head cold, and the test correctly shows his status.

What we want to know is the likelihood that the patient really has the disease, given that he tested positive. The only way to estimate that likelihood is to make a guess—a prior—about the proportion of patients taking the test who actually have the disease.

The resulting probability assessment is contextual. It depends very strongly on how much disease is present, and how large a group of people we are testing. In the midst of a severe epidemic, perhaps the likelihood of having the disease is 1 in 3. In that case, if we test 150 patients, we get this result:

— 50 actually have the disease. Of these 49 are correctly identified, and one receives a false negative result.
— 100 do not have the disease. Of these, 99 are correctly identified, and one receives a false positive result.

The odds that a positive test result means the patient is genuinely sick are therefore 49 to 1. The doctor is on solid ground in recommending immediate treatment. But if there was no epidemic, using the test would be a good deal riskier and less informative. Suppose our prior is that 1 person in 100 has the disease. We test 10,000 patients and get this result:

— 100 actually have the disease. Of these, 98 are correctly identified, and two receive a false negative.
— 9,900 do not have the disease. Of these, 99 receive a false positive result.

There are actually more false positive results than real cases! In this situation, immediate treatment might be a bad idea. The surest course of action is to run the test again for anyone with a positive result. Because the second test is applied in conditions where the likelihood of sickness is high—it has a revised 'prior'—a finding of disease can then be trusted, whereas in the first test it cannot.

This little example cannot do justice to the full potential of Bayesian analysis, but it does show the central role played by one's selection of the prior. A well-chosen prior will yield sound inferences. The wrong prior will yield only misleading nonsense.

So how does one go about selecting a suitable prior? Here Jaynes had perhaps his most important and original insight. He showed that the principle of maximum entropy could supply a plausible prior for a huge variety of systems and situations.

This was a difficult feat, not simply in terms of understanding the science, but in overcoming an unhelpful habit of mind among his fellow scientists. Jaynes' early papers were often rejected by mainstream journals, and interest in his ideas grew only very slowly atfirst.

From the very first moment that the concept of 'entropy' existed, there was a tendency to view it as a strict physical property of systems, with a single unambiguous (thermodynamic) meaning.

In fact, it is not. There is a critical distinction here. Jaynes argued that entropy is not really a physical measure of the system itself, but rather a statistical estimate of its potential behavior. In most situations—even most thermodynamic situations—we can choose from a multitude of possible entropy measures. Which one we use depends on the level of detail we want to go into, and how much information we have about the system.

Notice the resemblance here to Spencer-Brown's observation about there being no single coherent, universal standard of randomness. The two men were attacking the same problem, albeit from very different angles.

Despite Jaynes' best efforts, the unhelpful habit of mind remains quite popular. So we need to be clear about what entropy is, and Jaynes' idea about it.

Defining entropy

In the broadest sense, entropy is all about the tendency of systems to transform themselves away from local intensity, and toward a kind of local dispersion. Concentrations of energy and substance both dissipate. One classic example is hot water poured into a bathtub filled with cold water. Initially the hot water is concentrated at one end, but fairly soon the hot and cold water mix, and the heat is shared out until the entire tub is merely warm. The dispersion of heat energy equals increased entropy.

We use the term 'entropy' in at least two different ways. If we want to be exact, entropy is the unit of measurement for the amount of energy dispersion. However, it is often more casually and confusingly referred to as the *cause* of the dispersion, meaning that the tendency for entropy to increase is the reason why the dispersion happens.

Because of how earlier generations of scientists approached the subject, we often refer to increasing entropy as increasing *disorder*, rather than dispersion. This is still a tempting way to describe it, because examples of disorder are so plentiful and most are indeed examples of increased entropy. For example, a scrapyard filled with metal structures, exposed to rain and air year after year, distintegrates into heaps of rust. A mountain alongside a fjord is worn down by erosion over ages of time, until all that remains are some rounded hills and a shallow bay filled with silt and gravel. Living creatures eventually die and decay. The stars burn out. So do galaxies. Even the elements themselves are apparently unstable and transform into stabler isotopes—some quickly, some over billions of years. Unless some other, so far unknown principle intervenes to prevent it, it seems that eventually our universe will be made up almost entirely of scattered particles of cold iron, illuminated only by a steadily dimming 'fossil light'.

All of these outcomes are dictated by the second law of thermodynamics (the first law being that energy can neither be created or destroyed). The second law says that the entropy of a system will either stay the same, or increase. Increasing entropy means an increase in dispersion and uniformity. The same quantities of energy and matter are still present, but they are distributed more and more evenly with time.

If we wanted to convey the general idea of entropy to an alien race who did not know our language, it is sufficiently general and fundamental to science that we probably could get by with nothing but zeroes and ones to represent it.

Concentration would tend to look like this:

11111111100000000000000000

Meanwhile dispersion would look something like this:

10010100001010010101000010

Draw an arrow leading from the one to the other, and assign the whole diagram a symbol, and you will have succinctly referenced the concept of entropy. It is even referred to sometimes as 'time's arrow'.

This concept of entropy is fairly familiar, at least among those readers who are interested in science. The man who did the most to make it possible, Ludwig Boltzmann, is perhaps not so well known. The subtleties of his original insight are still in the process of becoming clear to physicists.

The term 'entropy,' as a measure related to heat energy, was first proposed by Rudolf Clausius in 1865. At that time, the chief interest of science in the subject came from makers of steam engines, who wanted a more precise method to determine how heat energy could be transformed into useful work.

Clausius' concept centered on understanding the distinction between reversible and irreversible processes. An example of a reversible process would be a piston slowly compressing a chamber containing a quantity of air. In this example, as the chamber volume becomes smaller, the pressure and temperature of the gas both rise, but at any point in the process the pressure and temperature remain homogeneous, that is, they have the same value throughout the chamber. Even more important, in our mental picture of the piston, it could move in the opposite direction and the pressure and temperature would fall to their previous values—meaning the process is reversible. An example of an irreversible process would be exploding a balloon with a pin. Prior to bursting, the pressure and temperature of the air in the balloon would be constant; but once the balloon bursts, the pressure of the air it contained would drop rapidly in a highly nonlinear fashion, generating pressure waves (which among other things make the popping sound we hear). To unpop a balloon is obviously not possible.

Reversible processes are generally preferable in power generation, not least because substantial energy tends to be lost in irreversible processes to side effects like pressure waves. The problem that 19th-century engineers faced was that even the most smoothly run, close-to-reversible piston-driven steam engine could not transform heat energy into mechanical work with 100 percent efficiency. At that point they understood in a general way how to build reversible machinery. They next needed to understand what other, more subtle kind of constraint intervened to make their seemingly reversible processes so inefficient. As it turned out, that meant they needed to identify a different set of changes that were taking place invisibly, on a molecular level.

Any heat engine operates between two reservoirs of heat energy, one at a high temperature (the boiler, heated by combustion) and another at a lower temperature (typically the surrounding air, or perhaps a cool lake or river). Sadi Carnot showed in the early 19th century that when the two reservoirs have the same temperature, the engine cannot do any work—and that in general, the amount of work the engine can do is proportional to the temperature difference.

More than four decades later, Rudolf Clausius showed that the theoretical maximum of usable energy that can be extracted from such a system is governed by a ratio of the two temperatures:

$$Q_1 / T_1 = Q_2 / T_2$$

and

$$Q_1 - Q_2 = W$$

where Q_1 represents the amount of energy available at the higher temperature, Q_2 represents the amount of energy lost to the surroundings at the lower temperature, and W represents the mechanical work done.

This is a very helpful rule in designing or operating real engines. For example, imagine if we introduced 1,000 units of heat energy into a boiler at 1000 degrees Kelvin, with the surrounding air at 300 Kelvin (about 27 Celsius). Thus:

$$Q_1 / T_1 = 1{,}000 / 1{,}000 = 1.000$$

Out of that total, 30 percent would be needed to make up Q_2, so that

$$Q_2 / T_2 = 300 / 300 = 1.000$$

Then 70 percent of the heat energy, or 700 units, that we put into the boiler could potentially be taken out as work:

$$Q_1 - Q_2 = 1{,}000 - 300 = 700$$

Clausius referred to the quantity Q/T as S, the entropy, from the Greek word for transformation. He described it as a measure of the limits of energy transformation—how much energy could be converted to work.

In this example, we have not specified the boundaries of our system. When we do so, we discover something else that is important. It is actually possible to reduce the entropy of one part of a system, but only at the price of increasing entropy elsewhere. In a perfectly reversible heat engine, the work done by the engine reduces the entropy of whatever system it is applied to (for example, it might drive a pump to compress air). The entropy of the engine's surroundings is increased by an equal amount, because of the waste heat the engine produces.

When a process is not perfectly reversible, the reduction in entropy in one part of the system is never as large as the increase in entropy elsewhere. Thus, in the example above, although Q_2 could not be smaller than 300 units of energy, the total amount of energy dispersed could easily be larger if the engine wasted heat energy through poor insulation, or friction losses, or some other cause. That is, if the engine cycle was not really reversible, then Q_2 would be larger. This point about reversibility implying zero entropy change will be important later.

None of these early developments touched on probability theory in any concrete way. Up to this point, entropy was strictly about the uses and sources of energy. However, the concept took on a broader scope, and potentially a very different character, when in 1877 Ludwig Boltzmann first introduced the idea of *statistical* entropy.

Statistical entropy was rooted in the ideal gas law, which explained thermodynamics in terms of the random behavior of trillions of individual molecules. The ideal gas law simply said that the pressure, volume, and temperature of a given quantity of gas are all related. They follow the formula PV = nRT, where P, V, and T represent the pressure, volume, and temperature, n is the number of molecules of gas present, and R is a constant. The formula was highly useful on its own account, but perhaps even more important for what it did to the highly speculative notion that matter was made up of atoms. With the ideal gas law, the ancient notion of atoms (going back to the Greek philosopher Democritus) became a scientific hypothesis, something that could be tested and could generate practical applications.

However, the enormous numbers of particles involved in real physical systems made exact solutions difficult. Seeing this, Boltzmann applied a kind of thought experiment. He pointed to a distinction between 'micro-states' and 'macro-states' which became essential to scientific reasoning about the behavior of atoms in general.

Imagine a closed volume containing a fixed number of gas molecules at some specific temperature and pressure. The macro-state of the gas volume (the 'system') is defined simply by the pressure, temperature, and volume, plus the number of molecules present. So long as these values all remain the same within the limits of our ability to observe them, we can regard the system as remaining in the same macro-state.

However, from the ideal gas law, we understand that from moment to moment, the individual molecules are in constant motion, colliding with one another and with the walls of the vessel containing them. Their individual states are continually changing, as a molecule perhaps travels slowly in one direction, then collides and travels more quickly in another, then collides again and takes a new velocity in a new direction, and so on.

These constantly changing states of the individual particles do not affect the macro-state of the system. This was the basic assumption behind the ideal gas law, that there are so many trillions of molecules in motion that their average properties will not change from moment to moment. There will always be about the same number of molecules moving left as moving right, and so on. Yet as Boltzmann realized, these collisions and interactions among individual molecules clearly do change *something*. They matter in some way even if they do not change average properties on the macro-state level. Each different arrangement of individual molecules, then, represents a unique micro-state. Trillions of different micro-states are possible within the same macro-state.

Boltzmann's key insight into entropy was to exploit a subtle truth about the relationship between two or more separate systems. Assume we have two volumes of gas, each at the same pressure, temperature, and volume, and thus in the same macro-state. We already know from basic principles that a pair of identical systems should have double the total entropy for one system. (We have Q units of energy, plus another Q units of energy, both at the same temperature T.) This implies that in general, when combining comparable systems, we should add the individual contributions:

$$S_{both} = S_1 + S_2$$

But if we are computing the total number of micro-states for the new combined system containing both volumes, that number does not simply double. It increases at a much faster rate, because it is computed by taking a permutation of the micro-states for each.

In a permutation, we can take any unique micro-state of System 1, combine it with any unique micro-state of System 2, and the result will be a unique micro-state of the combined system. We generally assign the total number of possible micro-states the letter W (somewhat confusingly, this is the same letter used to designate mechanical work done). So to find the total number of micro-states in a combined system, we must multiply the two micro-state totals:

$$W_{both} = W_1 \times W_2$$

I will illustrate with a very simple example. Suppose System 1 consists of one molecule and it can only exist in 7 different states—the molecule is either moving, in one of six directions, or it is standing still. Suppose the same is true of System 2. Then given that S1 and S2 are independent, S1 could be stationary while S2 takes on any of 7 states; or it could be moving left while S2 takes on those same states; or it could be moving right; and so on. The total permutations number 7 x 7 = 49. W for the combined system is therefore 49.

For systems with many molecules, each with many possible states, the number W explodes in size. We might find at a given moment that molecule 1 is traveling to the left at 5 meters per second, molecule 2 is traveling right at 8 meters per second, molecule 3 is traveling upward at an angle of 17 degrees from the vertical at 6.5 meters per second, molecule 4 is spinning clockwise in place, and so on. We can imagine cataloging precise states for all the trillions of individual molecules.

Thus there are uncountable numbers of equivalent micro-states even for systems made up of very small numbers of molecules. A cubic meter of air contains more than 10^{25} molecules (that's 10 multiplied by itself 25 times), and there are many different speeds and directions that each molecule could travel or rotate or vibrate in. Even if there were only two available states for each molecule, using the permutation method we find that the total number of available micro-states for one cubic meter of air would be 2 to the power 10^{25}—an enormous number.

Boltzmann realized that the only way that the rule of *adding* entropies and the rule of *multiplying* micro-states could both hold true at the same time was if the entropy for a given system were to be proportionate to the logarithm of the number of micro-states. Then multiplying W1 by W2 is the same as adding their logarithms:

$$S_1 = k \times \ln(W_1)$$
$$S_2 = k \times \ln(W_2)$$
$$S_1 + S_2 = k \times \ln(W_1 \times W_2)$$
$$= k \times \ln(W_1) + k \times \ln(W_2)$$

I will not explain here how a logarithm works—see any math guide. The constant 'k' was determined experimentally and is in units of Joules per degree Kelvin, that is, units of heat energy divided by temperature. Its value is $1.3806505(24) \times 10^{-23}$ and it is known today as the Boltzmann constant. The Boltzmann constant thereby relates the older Clausius concept of entropy to the statistical version.

This was more than a minor mathematical trick. It was a tremendous intellectual achievement, because it expressed a system's potential for action purely in terms of the number of distinct states the system could take on. It transformed a concept founded on two physical measurements (heat energy combined with temperature) to one founded solely on measurements of probability. Boltzmann's insight was in relating the microcosmic world to the macrocosmic one simply by counting permutations.

The critical breakthrough in Boltzmann's system lay in recognizing that macro-states representing concentration were relatively rare (that is, they consisted of fewer distinct micro-states), while macro-states representing dispersion were more common. This is just the same observation we make when we create a permutation triangle and notice how few cases there are of 100 percent success, compared with the huge number of cases resulting in partial success.

The tendency for systems to move from concentration to dispersion could then be understood in probability terms as simply being *the tendency to move from a rare, unlikely state to a more common, probable one.*

In other words, going back to our example, there are invariably fewer states resembling this:

11111111100000000000000000

than there are resembling this:

10010100001010010101000010 0

This means that in any given physical system, the gradually increasing dispersion of heat energy (or whatever quantity we wish to measure) is simply a matter of chance and time.

Notice that in the late 19th century, there could be no question of Boltzmann, or anyone else, proving this through direct experimental measurement—that is, by observing and counting up individual micro-states. In those years the existence and nature of atoms was itself still mainly a matter of speculation. The electron would not be discovered until 1897, the approximate size of atoms would not be estimated until Einstein's paper on Brownian motion in 1905—and the atomic nucleus would not be found until Lord Rutherford's shocking laboratory experience in 1907. Lacking real data, the 'atomists' were often compelled to argue in terms of crude hypotheticals.

Today we have definite, experimental entropy figures for many simple cases of atomic micro-states, but at that time, statistical entropy was founded on a philosophical argument, not empirical evidence. In particular, no one had ever enumerated all the micro-states for a practical system, or observed how much time the system spent in each one. This is still true, especially for anything big enough to see with the naked eye. There are simply too many possible states to count properly.

Boltzmann's law proceeds from the principle of extension, that whatever is true of one system in isolation must also be true when considering several systems at the same time. The rule is accordingly quite general, once proven. That is, the rule isn't necessarily just about atoms and their level of heat energy. What else it might apply to isn't immediately clear, but there is nothing in Boltzmann's argument that makes other uses of the rule impossible. Remember, Boltzmann knew almost nothing about the nature of atoms, so his argument couldn't be more specific. The twofold challenge for Boltzmann was to prove this rule for thermodynamic cases, and also to determine the limits within which it could legitimately be applied.

There is a longstanding concern among experts on entropy about intellectual sloppiness in taking the concept from its thermodynamic context (in which entropy refers strictly to dispersal of energy) and using it metaphorically, or in some more general way. Frank Lambert, writing in 2000 in the *Journal of Chemical Education*, put it succinctly:

> There is no more widespread error in chemistry and physics texts than the identification of a thermodynamic entropy increase with a change in the pattern of a group of macro objects. The classic example is that of playing cards. Shuffling a new deck is widely said to result in an increase in entropy in the cards.
>
> This erroneous impression is often extended to all kinds of things when they are changed from humanly designated order to what is commonly considered disorder: a group of marbles to scattered marbles, racked billiard balls to a broken rack, neat groups of papers on a desk to the more usual disarray. In fact, there is no thermodynamic entropy change in the objects in the 'after' state compared to the 'before'.

This goes right to the heart of our problem, because throughout this book I will argue—following Edwin Jaynes—that there really is a meaningful change in entropy in these cases, just not of the thermodynamic kind.

Let us stipulate that what Lambert has written is true in a literal sense. When we throw a die, toss a coin, or shuffle cards, each time the macro objects come to rest, the classical *thermodynamic* entropy of the objects remains the same. The only measurable increase in thermodynamic entropy in such cases would be in the muscles of the thrower, due to the need for biochemical energy to make the throws or do the shuffling. There might also be some very small amount of heat that starts out as muscle energy, and will wind up shared between the table and the die, once friction brings the die to a stop. But these are all negligible quantities and there is nothing to suggest that they influence the value of the die that comes up. (It is only a perverse figure of speech that makes streaks of lucky throws 'hot' and failure to repeat a success 'cold'.)

Nonetheless there is a kind of entropy that we can observe increasing in games of chance, or in traffic fatalities, or in measures of political stability across 5,000 years of recorded history, that is not thermodynamic in nature. Granted, even naming this as a form of entropy involves controversy. It was first done back in 1948 by Claude Shannon, who needed to describe a new measure he had invented for communications systems. He chose to label it as 'information entropy', but was by no means entirely sure that he should:

> My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. John von Neumann had a better idea, he told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function goes by that name in statistical mechanics. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.'

In his 2000 essay, Lambert argued that only thermodynamic entropy really merits being called entropy, and that information 'entropy,' though a successful scientific innovation and widely used in communications theory, was an ill-chosen term that has resulted in decades of confusion. If we focus solely on entropy's historical origin as a measurement of available energy, many physicists would no doubt agree.

But not Edwin Jaynes. What Jaynes accomplished, decades after Boltzmann, was to vastly broaden the uses of Boltzmann's concept of enumerating possible states. Jaynes' idea of entropy requires above all a certain flexibility in thinking. What 'entropy' still means today, to most physicists, is a very specific rule applied to a limited set of cases—as we said above, there is a tendency to think in terms of 'the' entropy for a system. Thermodynamic entropy is measured in units of temperature and energy. The measurement cannot be done in any other way.

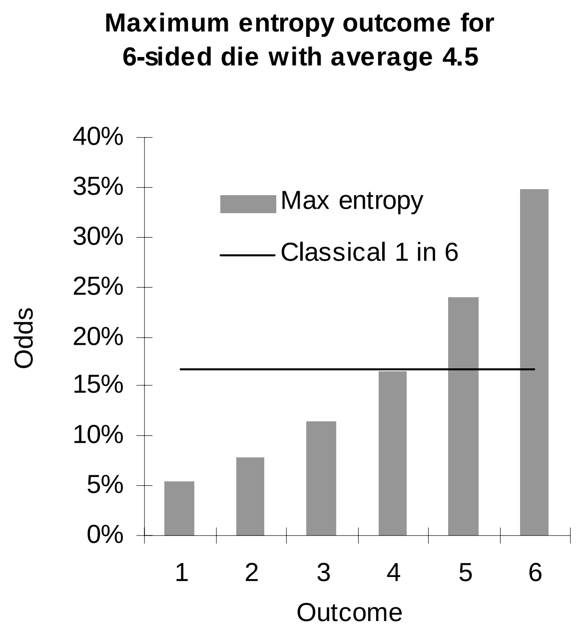
Jaynes' principle of maximum entropy is far broader in application, in two senses. First, Jaynes argues that there are in fact multiple versions even of the supposedly sacrosanct measure of thermodynamic entropy. As Jaynes observes here, the controversy raged for decades among the pioneers in the field:

> Clausius and early Planck sought to establish the second law as an absolute law of physics, true of necessity in every case. Late Planck recognized that this is impossible, and advanced to the viewpoint of Gibbs. As recounted by Kline, Boltzmann changed his mind many times.
>
> There are two cogent reasons why one must abandon the view of Clausius and early Planck. First is the recognition that the entropy to which [the two laws of thermodynamics] refer is not a 'real physical property' of a physical system; it has also an anthropomorphic quality because it is a property of the Thermodynamic system that you or I create by the measurements we choose to make on the system.
>
> A given physical system (say, a quartz crystal) corresponds to many different thermodynamic systems, depending on which macrovariables you or I choose to observe and/or control. They have different entropy functions, depending on different macrovariables ... It is therefore meaningless to speak of the 'entropy of the crystal' as if it were a physical property like its energy.
>
> This is no more paradoxical than the fact the music produced by a piano is not an 'objective physical property' of the piano; it has also an anthromorphic quality, because a given piano may produce an unlimited variety of music, depending on which keys you or I choose to press down.

Second, it is clear from Jaynes' description above that we can choose to focus on other macrovariables than energy and temperature. The only thing that is essential in this kind of analysis, is that we enumerate all the possible states, and then group them according to a) how frequently they occur, as well as b) how they look to us from a particular perspective, or using a particular kind of measurement.

Edwin Jaynes' great achievement was to devise a method of calculation that could predict changing properties of complex physical systems, based solely on one or more measurable values that characterized that system's behavior. This innovation expands the range of applicability of entropy to virtually anything that can be measured. When combined with Bayesian inference, this idea becomes an extraordinarily broad method for reasoning about scientific problems—not just a system of probability, but a system of scientific logic.

Jaynes' earliest and still most famous illustration of this idea was a biased six-sided die. A standard six-sided die can take on six values, 1-6, relating to the number of pips showing. These can be said to be the different states of the die. If the die is unbiased, then our expectation—our Bayesian prior—is that it will alternate among these six states in equal proportions. There is no gradient in the distribution of entropy: the distribution curve is flat. Alternation in equal proportions maximizes our uncertainty and the disorder of the system. It is the maximum-entropy pattern.



But Jaynes showed that if the die is somehow biased, then the maximum entropy outcome is different, and a new prior must be devised. Suppose instead of an average score of 3.5 pips (the mean obtained by summing 1 through 6 in equal proportions), the die averaged 4.5. Jaynes showed that in that case, the maximum entropy principle would dictate a skewed, logarithmic distribution of results. A result of 6 would be most common, followed by 5, 4, 3, and so on down to 1, as shown in the graph above.

In making this prediction, Jaynes observed, we are not *modeling* the behavior of the biased die. We have received no information about how the die is thrown, or any of the unique physical properties that cause it to be biased in this case. We just know the six possible states that the die can end each throw in, plus one additional datum: that the average must end up being 4.5. Apart from this very limited set of facts, we believe, not necessarily from observation but simply because it is so common in nature, that the die must also obey the maximum entropy principle.

The prediction that results is austerely mathematical in nature, just like the regime-length curves I showed earlier. It is based on counting the possible outcomes in terms of pips, and then applying the maximum entropy rule to them. There is no consideration here of energy or temperature or binary bits, much less the intricate dynamics of actual dice rolling and bouncing. The prediction does not refer to any other properties of the die, only this one range of possible values from 1 to 6, and the expected average.

Knowing only the various states that the system can take on, plus the expected long-run average of those states, we have enough information to construct a plausible Bayesian prior of how often each state will occur. All the myriad complexities of the system, still hidden from our view, are rendered irrelevant. We do not need to know them.

Jaynes found to his growing surprise, from 1957 onward, that this minimalistic method turns out to be quite sufficient for many complex physical problems. Notice that it also works perfectly if the die is not biased—that is, if it behaves in a classical way. There is no contradiction here between the two ways of computing outcomes. When classical behavior really does occur, Jaynes' system is perfectly capable of reflecting it.

Multiple predictions for the same system are not only possible, but necessary. For example, if we threw two dice, we would have a choice of at least two distinct distributions to theorize about. We could estimate the distribution of single die results, without any regard for how they are combined. Then we would also have the choice of estimating the behavior of the pairs. We might find, for example, that the distribution of single die results from 1 through 6 approaches a classical flat line, with all numbers having equal likelihood, and no significant deviation—but that the pairs show a systematic bias away from doubles results. (In fact, as I will show in *Experiments in Decline*, we do find this. A steadily increasing shortage of pairs is in no way hampered by the requirement that there be long-run equality between the number of 1's and the number of 6's.)

From the point of view of George Spencer-Brown, the separate prediction for pairs as distinct from single dice constitutes a different standard of expectation regarding randomness. The fact that we get different answers depending on how we choose to measure the system reflects the underlying incoherence of the classical theory of probability. What we find depends on what we look for, and at what level. There is no one single 'right' scale upon which to observe.

By this Spencer-Brown meant more than that we *can* look at these two different perspectives. He meant that we *must* look at them both, simply to understand what is happening. In this example, if we were to look solely at the distribution of single die outcomes, and then extrapolate without further direct observation to the other levels of 'molecular' events, our extrapolation would yield the wrong answer. The same thing happens when we examine the distribution of 'atomic' values emerging from a random number generator. The skewed and changing pattern on the 'molecular' level cannot be predicted based on 'atomic' behavior.

Jaynes would say something very similar, with only the vocabulary changing. The relative frequencies of different single die results would constitute one macrovariable, and the changing frequency of occurrence for doubles would be another. The idea underlying Jaynes' work, as I have said, is that probability distributions are not really models of behavior, so much as models of our *knowledge* of behavior. Start from zero and ask yourself: What do we really know about single dice? What do we know about pairs? Classical probability makes an assumption about the relation between the two, but what (if anything) actually validates that assumption?

This is the subtle point that opens so many doors, but that can also be very confusing. We are revising our view of probability so that *different frameworks of measurement give different outcomes, even for the simplest naked-eye systems.* (The relevance here to the puzzles of quantum mechanics should be obvious.)

In the classical or 'frequentist' view of probability, the 'atomic' events are treated as genuinely independent. In adhering to classical premises we make a metaphysical assumption, a strong statement about causality in the real world, and then use that assumption to derive our arguments about consequences. Arguments about the behavior of two dice are to be kept strictly deductive, based on what is actually a highly questionable assumption about single dice. *If* each die is truly independent, then systems consisting of more than one die, or successive throws of the same die, will generate permutations of a strictly predictable kind.

In Jaynes' model, or Spencer-Brown's, we do not make assumptions of this kind. What we know from one method of measurement (looking at each die singly) does not justify our making assumptions about what we will see when we make a different sort of measurement (collecting the dice into pairs). Different levels of behavior need not be related in a strict, explicit, known fashion, that is, a deductive model. They are to be dealt with by their own distinct Bayesian priors. At this point we don't need to know *why* the outcomes from 1 through 6 come up evenly, yet doubles grow increasingly rarer. We don't even know, when we first approach the problem, *that* this happens. But we do automatically expect each perspective, each macrovariable or measurement scheme, to yield its own distinct result.

We can see at a glance in this simple example that different behaviors by the system on its different levels are mathematically compatible. The same is true, of course, of the multiple levels of the regime lengths curve. We see a certain kind of rule that governs the spans of individual rulers in a dynasty, and then we see the same kind of rule governing the spans of successive dynasties in a particular country—or the set of all dynasties across the globe. Each level of the system must be approached separately: on this point Spencer-Brown and Jaynes agree.

In dealing with the dice or the hereditary regime lengths, we are reasoning from incomplete information, but that is what the principle of maximum entropy is uniquely able to allow us to do. Our goal is not to achieve perfect understanding of the system dynamics as they actually exist. We do not know the details, any more than we know the details of an entangled quantum mechanical system. The true dynamics involved may be far beyond our capacity to model. All we need in order to reach a conclusion, and all we are justified in assuming, is a plausible Bayesian prior for each distinct measurement scheme or level in the analysis.

We can summarize all this in a few points:

— Bayesian analysis has distinct advantages over other forms of reasoning about probability, but it requires careful selection of the prior.

— The principle of maximum entropy mandates that our prior be multi-level. We must establish separate expectations for 'atomic' and 'molecular' levels of observation.

— While the principle of maximum entropy offers an astoundingly good fit to natural behavior, it also implies in many cases (because of its austere mathematical basis) that there is a spooky causal relation that exists between widely separated parts of the system, one that transcends time and space.

Once we adopt a Bayesian perspective, we can focus on what is possible to us to know in a given situation, and what the appropriate context is for each estimate we make. Then a far richer view of probability emerges. It becomes possible to imagine that dice might behave one way when viewed as single units, and a very different way when viewed in pairs. It becomes normal and natural for regimes separated by thousands of years to obey a common length rule, to seem closely related to one another despite the vast distances involved. The observations made by Spencer-Brown and by ESP enthusiasts, which have been languishing in file drawers for decades, suddenly seem a great deal more plausible as well. Quantum mechanics no longer stands as an eerie exception to the common rules of science. The rules are again the same everywhere. A vast range of phenomena that were invisible or inexplicable from a classical perspective are rendered routine from a Bayesian one.

The universe has a systemic bias toward maximum entropy that out of ignorance we tend to call nonrandom, but that is really just a different *kind* of randomness. Once we understand that randomness exists on many levels, that it depends on our framework for measurement, that it tends toward this particular kind of order, our entire picture of the universe changes. We can see the possibilities literally all around us.

Because of the way entropy is computed, the maximum entropy principle frequently yields a logarithmic or power-law distribution curve. We have known for centuries that nature is filled with logarithmic curves of this general type. Now this kind of curve becomes the form of our prior expectation for a huge range of practical tasks.

We might be plotting the prices of used books on the Internet, the sizes of gas giant planets, the acreage devoted to hay fields in Norman England, career home runs of baseball players, traffic on a network of Web servers, casualties from World War Two bombing raids, and so on. All these phenomena follow power-law curves. Their behavior, however spooky and unexpected, is consistent with the method of Bayesian inference and the maximum entropy principle.

Continuing Jaynes' revolution

Jaynes found very little support among his fellow physicists in the early days. He joked that it took, on average, 20 years before any paper he wrote began to have an impact on scientific practice. His early papers were routinely turned down by mainstream journals.

In his retirement speech, Jaynes tried to assess how much progress had been made since he began his program of investigation.

> Indeed, today I am far more self-confident about probability theory than was possible forty years ago when I started lecturing on probability and information theory. The reason is that the essential evidence is now in; we are sustained no longer by faith and hope, but by proven theorems and accomplished facts on the level of new useful numerical results. There is no doubt that the formulations of probability theory and statistical mechanics as extensions of logic are here to stay; they will be the universally accepted basis of those fields 100 years hence. Too many things are coming out right to allow any other outcome. ...
>
> Functionally, probability theory as extended logic includes as special cases all the results of the conventional "randomvariable" theory, and it extends the applications to useful solution of many problems previously considered to be outside the realm of probability theory. The now much strengthened theoretical foundation and continued pragmatic success and the failure of critics to uncover any defects in it or offer any usable alternatives justify this confidence in it.
>
> When applied to problems of parameter estimation or hypothesis testing, probability theory as logic is generally called Bayesian inference, on historical grounds explained elsewhere, and it is accomplishing a major house-cleaning in the field of statistics. ...
>
> Orthodox methods of inference had another distressing property: on the one hand, to get conclusions from data they offer no way to take into account our prior information about the parameters of interest; yet on the other hand they require us to make additional assumptions about the frequency distribution of errors that are arbitrary; that is, not justified by any of our information. In contrast, it is now a proven theorem that, when we apply strict Bayesian principles with initial probabilities assigned by maximum entropy, our subsequent inferences depend only on the data and the circumstances (maximum entropy constraints) about which we had prior information; there is no room for gratuitous assumptions because circumstances about which we have no information automatically cancel out and contribute nothing to our final conclusions.

As the remark above of '100 years hence' suggests, Jaynes did not regard the project as anywhere near complete. He described himself as 'trying desperately' to complete a textbook, *Probability Theory: The Logic of Science*. It would only be issued posthumously, in 2004, by Larry Bretthorst, one of his students.

Jaynes anticipated that the coming generation would need to keep fighting the battle against scientific inertia. He offered some advice to those who would come after him:

> Looking back over the past forty years, I can see that the greatest mistake I made was to listen to the advice of people who were opposed to my efforts. Just at the peak of my powers I lost several irreplaceable years because I allowed myself to become discouraged by the constant stream of criticism from the Establishment, that descended upon everything I did. I have never—except in the past few years—had the slightest encouragement from others to pursue my work; the drive to do it had to come entirely from within me. The result was that my contributions to probability theory were delayed by about a decade, and my potential contributions to electrodynamics —whatever they might have been—are probably lost forever. ...
>
> In any field, the Establishment is not seeking the truth, because it is composed of those who, having found part of it yesterday, believe that they are in possession of all of it today.
>
> Progress requires the introduction, not just of new mathematics which is always tolerated by the Establishment; but new conceptual ideas which are necessarily different from those held by the Establishment (for, if the ideas of the Establishment were sufficient to lead to further progress, that progress would have been made).
>
> Therefore, to anyone who has new ideas of a currently unconventional kind, I want to give this advice, in the strongest possible terms: Do not allow yourself to be discouraged or deflected from your course by negative criticisms ...
>
> Take comfort in the historical record, which shows that no creative person has ever been able to escape this; the more fundamental the new idea, the more bitter the controversy it will stir up.

Jaynes' proposal has attracted growing interest from scientists over the years, and a number of specific applications have resulted, but to date it has not inspired wholesale, revolutionary change. Indeed, I would say that many of the really exciting possibilities were not fully appreciated even by Jaynes himself.

Jaynes was conservative in his arguments. He believed that 'for scientists to expend their serious professional time and effort on idle speculation can only delay any real progress'. He confined himself to publishing only results he was sure of, and this largely meant results in physics, chemistry, and probability theory. He wrote relatively little about the application of the maximum entropy principle to social phenomena, or biology, or other fields. For example, he had some ideas about entropy and economic equilibrium, but he regarded himself as an amateur in the field, and his notions as no more than 'half-baked'. He did not publish them during his lifetime. They appear as a brief aside in his posthumous textbook.

The leap I am proposing to make with these two books is one that Jaynes foresaw in general terms, but could not actually make. It is to declare victory in the revolution he started, to commit to the view that most, if not all, of the phenomena we see around us are can be predicted and explained by the maximum entropy principle. Whether they are 'caused by' entropy, rather than explained by it, is a tricky question. Strictly speaking, when we set up a maximum entropy curve using incomplete information, we are not specifying the ultimate *cause* of the events. We are only summarizing the limited information we have about them. All the same, I will from time to time refer to entropy as 'causing' decline. This is a verbal shorthand that feels natural and has some value, but it should be used cautiously.

To make this argument requires a broader foundation than Jaynes' original work, more even than the work done by the dozens of others since. A number of writers have noticed how common power-law or logarithmic distributions are in nature, and have declined to make the leap—to actually assign entropy as the cause—because that fact by itself is not conclusive. There have been some unanswered questions to solve.

Denis Evans and system evolution in time

One recent discovery makes my proposed leap far more compelling. Thanks to the work of another pioneer, Denis Evans, we are not stuck at the level of observing that the universe sort of *looks like* it follows Jaynes' principle. We can actually show that it does experimentally. We can track the process step by step.

A serious shortcoming of work using Jaynes' original method of computing distributions is that it has tended to be applied mainly to static, limiting cases—that is, it predicts where the system ought to *wind up* in the long run, but it does not necessarily tell us anything about the process of getting there. If we use the terminology introduced a few pages earlier, we see that Jaynes' original maximum entropy curves (and those of his early followers), while found in great numbers in nature, do not venture beyond the N-type. To develop a general formula for t-type or Nt-type systems poses a different problem altogether.

In 1993, a team at the Australian National University, led by Denis Evans, developed the fluctuation theorem, a brilliant extension of Jaynes' principle. Jaynes' principle says that the entropy of a randomly varying system will tend to increase over time, approaching an equilibrium state of maximum entropy. Evans and his collaborators worked out a method to estimate how much the incremental increase would be at any given point in time, for any given system.

As Evans puts it, the fluctuation theorem is a proof (or perhaps a limiting case) of the second law of thermodynamics. It defines how much time is required, for a system of a given size, for short-term reversals of entropy—fluctuations—to die away. It establishes a kind of boundary at which the second law truly becomes a law, and not just a tendency.

The fluctuation theorem was proven experimentally for the first time in 2002, using a laser on a blob of plastic floating in water. I will explain the theory and the experiment in more detail in *Experiments in Decline*.

What matters most at this point is that Evans and his collaborators implicitly opened up a second enormous class of everyday phenomena that we can test for Jaynes-type behavior—complex systems *evolving* over time. Here again, the expected form of the evolution tends to be a logarithmic or power-law curve. Entropy increases in logarithmic proportion to the size of a set or system. It follows that the values we choose to measure (using Jaynes' method) should show a logarithmic trend.

As we did with the N-type cases, we find that nature is filled with such curves. This is the decline effect in its most blatant, straightforward form. Here is where we find the juvenile criminal whose appetite for crime declines with age, and the army with a declining capacity to inflict casualties.

In practical terms, then, science already 'knows' about the decline effect. A mass of evidence has already been accumulated in diverse fields. Suitable theorems for explaining the evidence have been developed. In many fields, statisticians and other professionals already allow on a day to day basis for decline over time. They already work on the assumption of a maximum entropy distribution. The problem is that they do so on a piecemeal basis from one problem to the next, without recognizing the overall picture.

Something like this has happened before

It is admittedly rare for a widely held theory or law to be proven wrong. It is even rarer for a theory to be proven wrong in a way that forces major changes in daily practice.

Scientific advances tend to be drawn-out affairs, a gradual process in which a thesis is tentatively advanced, weighed, modified, tested, and eventually adopted across decades or even centuries. Often the more far-reaching the problem, the longer science takes to fully deal with it.

For example, a rough value for the speed of light, of 224,000 kilometers per second, was proposed by Danish astronomer Olaf Romer in 1676. The debate over whether light is a particle or a wave dates from work by Newton in that same period. Numerous attempts were made to refine the speed estimate; yet it was not until two centuries later that Albert Einstein (along with Poincare and several others) saw the implications of light having a fixed velocity and developed his theory of special relativity.

Indeed, while we now have a firm number for the speed of light, the debate over wave-particle duality has never really ended, since the present consensus is that somehow, mysteriously, light combines aspects of both.

While scientific advances are mostly slow-motion affairs, there are some precedents for the kind of abrupt reversal I am proposing. For example, we have the well-known story of the Copernican cosmos—the idea that the Earth revolves around the Sun.

In the course of the Copernican revolution, science did indeed radically change direction, in a relatively short time. What is interesting about the Copernican case is that the abrupt change related to a *detail* of the picture, not the heliocentric theory as such. There was a long lapse between the time it was first proposed that the Earth circled the Sun, and the point where the model finally (and fairly abruptly) broke through into wide acceptance.

The idea of heliocentrism dates back as far as Aristarchus of Samos in the 3rd century BCE, and may have been known even earlier. But for many centuries the heliocentric universe was considered an absurd, wildly speculative notion. How could the solid Earth be moving?

Many educated people in ancient and medieval times were aware of the theory, but most preferred to side with the cosmologies of Aristotle and Ptolemy—not to mention Christian teaching and their own common sense. It was the simplest answer by far. Whether one considered it a matter of divine revelation or of scientific fact, the Earth seemed to sensible people everywhere to be the unmoving center of Creation.

When Nicholas Copernicus created new astronomical calculation tables in 1541 that assumed the planets including Earth all circled the Sun, the Catholic Church promptly put his book on the Index of prohibited works. This did not stop Copernicus' tables from being used. Ironically, however, opinion did not change even among all the astronomers who adopted the tables.

There was a convention in those times, a quiet compromise to allow the work to proceed. Astronomical tables by themselves were not necessarily considered to have a definite physical meaning. This spared practical astronomers from charges of heresy if they used better tools. (There is an echo of this attitude in the approach of most contemporary scientists to the paradoxes of quantum mechanics. Quantum formalism works to give the right answer, even though no one is entirely sure what physical reality it represents.)

Copernicus' method made for slightly more accurate predictions, but it still rested on the clumsy Ptolemaic system of perfectly circular orbits, and smaller circular epicycles within those circular orbits. As his colleagues were not obliged to defend a literal heliocentric cosmos on his behalf, and his system's predictive power remained limited by its Ptolemaic foundation, Copernicus' work did not change many minds.

Enter Galileo Galilei. Galileo's chief inspiration and contribution to the problem, nearly a century later, was his development of a 30-power telescope, which allowed him to observe the moons of Jupiter, sunspots, and the details of the lunar surface. All these observations of a cosmos rich in detail (and quite unlike the simple, perfect, heavenly spheres of the Ptolemaic system) strengthened the case for heliocentrism, if only indirectly. Galileo eventually proposed in his *Dialogue Concerning the Two Chief World Systems* (1632), that the mystery of what causes the tides, and the mystery of the Earth's motion, were one and the same—that the circular motion of the Earth disturbs the oceans like water in a shaken vase.

For endorsing Copernicus, and for supporting heliocentrism with various compelling if inconclusive observations and arguments, Galileo was tried, forced to recant, and placed under house arrest for the rest of his life. He paid a heavy price for his convictions. The Catholic Church did not remove its ban on his work until 1835.

The trial of Galileo has come to be viewed as a paradigm case. Cultural historians can point to long-term consequences of the battle over heliocentrism not only in religion, but in art, politics, mathematics, language, and many other places.

Yet ironically, Galileo's work did not actually win the battle for Copernicus. His tidal explanation was deficient, even though it was intuitively brilliant, for it failed to point to any role for the Moon. Moreover, even his strongest arguments for heliocentrism were regarded as inconclusive not only by the Inquisition, but to some extent by Galileo himself. Observations of the planets and discovery of the moons of Jupiter made heliocentrism plausible, but it was not a strong enough foundation to compel real change.

At the time of his trial, Galileo had been a proponent of the Copernican system for more than 20 years, and the Copernican tables had been available for nearly a century, yet most astronomers were still not entirely convinced. Galileo was the most famous advocate of the theory in his time, and had inspired many colleagues to undertake their own investigations, but his contribution was not decisive.

What finally convinced the astronomers—and thereby ended close to 2,000 years of dominance by the Aristotelian Earth-centered cosmos—was the mathematical work of

Johannes Kepler.

Inspired in part by Galileo's first work on the subject in 1610, Kepler developed the period-distance law in 1619, and then in 1627 published tables based on the orbits of the planets being ellipses, not circles.

Kepler's mathematical approach offered such superior precision in predicting the movements of the planets (as well as comets and moons) that within a few years it had been adopted by the majority of astronomers, religious objections notwithstanding.

Once Isaac Newton's theory of gravity was published later in the 17th century, the case became overwhelming and the many objections that had seemed so plausible for centuries ceased to matter.

The Copernican revolution thus took place in three phases. The first was a centuries-long accumulation of controversial hypotheses and poorly explained anomalies. The second was the abrupt realization that the Earth does go around the Sun—*in an ellipse*.

Each phase had its key figure. In the first phase, it was Galileo. Galileo's arguments may not have been decisive, but his commitment to science, to gathering tangible evidence, was unsurpassed. It was Galileo's famous muttered remark, after he knelt before the Inquisition and formally renounced his theory, that captured the essence of the struggle: *E pur si mouve*—even so, it moves.

The hero of the second phase was Kepler. We need to keep in mind how long Kepler worked on the problem—decades—and that Kepler resorted to his elliptical formulas only after many years of trying to salvage the sacred Aristotelian-Ptolemaic model with its nested perfect solids setting the orbital radii of the planets.

The third phase was exploitation of the new ellipse model by Newton. Now that the way was open, a host of practical applications became apparent. The law of gravitation came *after* Kepler's discovery of the elliptical orbit, and the process most likely could not have happened any other way.

How does this story apply to our present puzzle? Classical probability forms part of the very foundation of science. For the past 350 years it has gone from triumph to triumph, in fields as diverse as genetics, thermodynamics, cryptography, insurance, operations research, and on and on. Given the clear historical superiority of the classical model, anyone hoping to make a serious attack on it must aim, following the example of Galileo, Kepler, and Newton, to provide three things:

— A wealth of evidence that was either not previously available or has not been examined with sufficient care, presented to raise the plausibility of the argument and inspire conviction that an error has been made (Galileo's contribution)

— A more precise forecast of actual outcomes than the existing approach delivers (Kepler's contribution)

— A physical principle or general law that covers all the cases and is consistent with other science (Newton's contribution).

This would be a great deal to ask, for anyone to perform the roles of Galileo, Kepler, and Newton all at once. There is a chasm here that few have ever considered crossing, much less attempted—not just to refute classical probability, but at the same time to resolve the conflicts that would necessarily arise concerning the rest of science.

However, there is no point in trying to minimize the challenge involved here, as if it could be thought of as another routine, incremental discovery. Better to acknowledge that as in the old saying, extraordinary claims do require extraordinary proof.

Fortunately for me, the roles of Galileo, Kepler, and Newton have already largely been filled in this instance by others, leaving me with the much more manageable task of explaining (or at least suggesting) how it all fits together. An army of distinguished scientists, including in particular George Spencer-Brown, Edwin Jaynes and Denis Evans, have all prepared the way.

This book lays out a number of far-reaching claims. Some are original to me. The majority are modifications or elaborations on work by others. What connects them all is the theme: I believe these phenomena are all caused by, or shaped by, entropy. Viewed in the light of Jaynes' maximum entropy principle, they revert from spooky and inexplicable to routine.

— An updating and elaboration on the Wright-Henderson cost reduction law for manufacturing and services.

— An explanation and vastly improved mathematical model for 'combat friction,' first described by Carl von Clausewitz in *On War* (1830).

— An alternative to the Ewens distribution, used in Hubbell's unified neutral theory of biodiversity.

— A radical new theory concerning the growth rates of human movements, multi-celled organisms, and epidemics—that they all follow a universal 'fractal' efficiency function.

I ask the reader to always keep in mind that this book is not intended to be an omnibus volume of otherwise unrelated hypotheses and models. While these are all valuable and interesting scientific topics in themselves, the standard for inclusion is whether they shed any light on the underlying nature of randomness, and the appropriate method for analyzing events.

Each chapter should advance us toward our ultimate end, to answer the two-part question: Is classical probability wrong, and if so, why? Or to put the question a different way: Does entropy govern the distribution of everyday events, and if so, how?

Wegener and his two curves

Just to show that extraordinary proofs do occur—and that they can rest more on determination and attention to detail than on superhuman powers of intellect—here is one more recent example where actual daily practice in science and technology did shift abruptly, in a revolutionary manner, when sufficient proof was supplied. I refer to the theory of plate tectonics and continental drift.

Alfred Wegener proposed in 1912 that the similarity in shape between the east coast of South America and the west coast of Africa meant that the two continents had once been joined together, and had drifted apart over 200 million years. As with Galileo and heliocentrism, he was not the first to suggest this. Mapmaker Abraham Ortelius had made the same observation as early as 1596, and so had others over the years.

Like Galileo, Wegener faced legitimate skepticism. Despite supporting evidence from local geology, fossils, and traces of glaciation, Wegener could not overcome the chief obstacle to his theory, the puzzle of how continents could move without disintegrating under the stresses. He believed the cause must be a combination of centrifugal force and tides—that the original supercontinent (which he called Pangaea) must have been pulled apart by rotational stresses.

Wegener catalogued the case for continental drift in successive editions of his book *The Origin of Continents and Oceans,* which first appeared in 1915 and continued in print until his death. In it, he appealed to every argument he could think of, such as calculating how far large aquatic dinosaurs could swim, or matching up the distribution of coal beds in South Africa with those in South America.

At one point Wegener even tried calculating from star sightings made in ancient navigators' logs how much further Iceland was from Europe in the 20th century than it had been in the 10th. The story was in many ways a replay of the story of heliocentrism. Despite Wegener's tireless effort—and while some of his ideas turned out to be wrong, much of his work was inspired and sound—common sense simply rebelled at the idea of solid continents moving in such a fashion. In 1926, Wegener attended an American symposium on his ideas in which virtually every geologist in attendance denounced it.

Wegener's story has an element of tragedy. He did not see his theory vindicated. Instead he died in 1930 while on an expedition to Greenland, apparently of a heart attack, still collecting evidence. For several decades his ideas were forgotten. But then between 1957 and 1961, came the first explorations of the Atlantic Mid-Ocean Ridge and discovery of the mechanism of seafloor spreading. The notion of continental drift abruptly swung from being a very old and dubious speculation that contradicted basic physics, to just one undeniable consequence of the new scientific framework of plate tectonics. Using deep-ocean submersibles and seismic tools, scientists could record the process actually occurring, even filming lava erupting from the seafloor to form new oceanic crust. The remaining objections were very quickly overcome.

With that discovery, several scientific disciplines had their foundations radically altered in a very short time—our understanding of earthquakes, the fossil record, the Earth's magnetic field, mineral exploration, paleobiology, evolution, were all changed. Wegener's explanation of how mountains are formed is now the standard model. Pangaea, the continent he postulated as having existed 300 million years in the past, is regarded as a real place now.

Ironically, in contrast to the Galileo affair, the cultural impact of this discovery was small. The reality of continental drift mattered intensely to geologists, scientists, and engineers, but it aroused little or no public passion, no crisis in religion. Alfred Wegener had endured decades of professional isolation and in the end decisively overturned generations of orthodox, respected science. To geologists he is something of a hero, a pioneer and standard-bearer. Yet he is scarcely known to anyone outside of the profession. His chief legacy now is not fame, but the difference he made. Today, the essentials of continental drift and plate tectonics are taught to schoolchildren, and few can remember when these were not commonplace truths.

There will probably never be another Newton, and I am certainly not one. But a scientist of more modest talent can hope instead to triumph through sheer persistence, in the manner of Wegener.

Over the years I have felt a special kind of kinship with Alfred Wegener. To be sure, I have had a far easier road than he did. His work required arduous journeys by dogsled across the roof of the world. Mine has required years of collecting statistical data, and many hours of throwing dice—tedious but safe. His theory was derided as 'utter rot' and scientifically insane by his colleagues. So far the harshest criticism I have encountered has been a few raised eyebrows and skeptical smiles. But what is similar is this: We were both inspired to a lifetime's work by the similar shape of two curves.

In Wegener's case, it was the outline of the coasts of North and South America, which fitted so closely with the coasts of Europe and Africa that he staked his career on their sharing a common origin.

In my case, it was two graphs that I constructed in 1990 from data in *Understanding War* by Trevor Dupuy. In that classic work, Dupuy wrestled at length with the nature of 'combat friction'. As I observed earlier, it is a known fact that large armies inflict fewer casualties per man per day than small ones, and that armies inflict fewer casualties in percentage terms on Day 2 than on Day 1 of a battle. Dupuy added data showing that large armies march more slowly than small ones, and that armies march more slowly as a campaign goes on, advancing only about half as much on Day 10 as on Day 1.

Tinkering with the numbers, I noticed something striking about Dupuy's data sets. They all adhered quite closely to a common pattern. The performance declines in time seemed to follow the same scale-invariant logarithmic curve as the performance declines that were due to increasing army size.

Like Wegener, the more I studied these curves, the more convinced I became that the similarity was not chance. There was some law, some principle, that led to consistent results in utterly different realms of activity—a law that treated numbers and time as if they were interchangeable.

Over the next decade, I collected dozens of examples, not just of similar-looking curves, but of literally the same curve, with the same exponent. The more I looked, the more I found the decline principle (as I came to think of it) lurking everywhere.

Eventually I turned to experimenting with classical probability, and so formulated the argument you see here. All the 200-plus graphs in this book really are the result of one principle or law.

Well, that is what I intend to prove, and how I came to be engaged in this project. So now what is the evidence that our view of the world is wrong?