INFORMATION, COMPLEXITY, AND LIFE:

being Chapter One of
A GENERAL THEORY OF VALUE

The theory of value offered in this book revolves around three propositions: first, that positive value is attributed to that which preserves or creates *more life*; second, that "lifefulness" is characterized by a particular quantity and combination of *complexity* and *organization*; and third, that in the case of human societies and minds, achieving this optimal quantity and combination of complexity and organization depends on the quality and flow of *information* among people, and between people and their less-animate environment—plants, animals, buildings, places, things.

The purpose of this first chapter is to lay the foundations for making these propositions plausible as well as useful. After all, if value is going to be a measure of evolving, changing lifefulness, one must first find ways to characterize "lifefulness" in suitably abstract and objective terms. To do this, I propose, we need to understand "complexity" and "organization" in a particular way, and here *information theory* is a great help.

This chapter begins with an account of basic information theory, showing how the concepts of uncertainty and complexity are related. It then addresses how uncertainty or complexity, relative to maximum possible uncertainty or complexity, tells us about a system's degree of organization. I discuss some applications of these concepts to games like tic-tac-toe and chess. A thought experiment involving prediction proves, I think, enlightening. We then move closer to the core of the book's argument *vis-à-vis* biological life with a discussion of *evolution* in complexity terms. I talk about complexity at different spatiotemporal *scales*, using examples from music and natural language and DNA, and conclude by touching on some of the implications of our account for human culture.

It will be Chapter Two before we see how the material offered in this chapter translates into a viable definition of value—and it will be in later chapters yet, of course, that we see how the theory of value developed in Chapter Two plays itself out in our everyday lives.

I. What is Information?

Most of us would say that information (about *x*) is what leads to an increase in our knowledge (of *x*). It would not matter critically what *form* that information took—whether it came to us in a book read, a story heard, a picture seen, or an actual event witnessed—because we could readily appreciate that the information itself is largely independent of the medium and symbol system used to convey it. Contra McLuhan, the choice of medium may be *a* message, but it is not usually *the* message.

We also know that having access to information is usually a good thing—unless it is *mis*information; but we understand this misinformation to be information nonetheless.

In recent times we have begun to speak more and more easily of creating information, of losing, recovering, processing, transmitting, copying, buying, owning, and selling it...the list goes on. Whole industries—entertainment, computing, banking, advertising, design, management—are founded on its manufacture and distribution. We live in the Information Age... And so we are soon led to wonder just what sort of stuff this "information" is. Is it something we *need*, like air or water, something *with* which we can have security and influence, and without which we are helpless? Where does it come from? And how is it that, on the one hand, information seems logically to *require* a mind with the intelligence and motivation to see it *as* information at all—much less as having meaning or value—while on the other hand it seems to be a quality or property of things themselves, quite apart from us? After all, a page of the Encyclopedia Britannica would seem, quite definitively and objectively, to contain more information than a page of comic book, a clock more information than a clay pot, a leaf more than a pebble, a young Einstein's brain more than a young dog's. Our intuition is here telling us that the amount of information "in" a thing—the amount of information it would take for us to understand it completely—has more to do with its *complexity* than its mass or size. Informationally speaking, small-and-complex is "bigger" than big-and-simple.

In light of these intuitions, the boldness of the formulation given to the concept of information by Claude Shannon and Warren Weaver in 1949 was breathtaking.[i] It gave rise quite literally to the science of cybernetics and to the computer and telecommunications revolution. As an idea, however—as a concept—it swept from electrical engineering through physics and theoretical biology to psychology and even art history. For what Shannon and Weaver did (and, very soon, what others such as John von Neumann, Norbert Wiener, W. Ross Ashby, Leon Brillouin, Ervin Shrödinger, and Satosi Watanabe did) was show that "information" was not, or rather not only, what one reads in the newspapers, hears on the radio, sees in a picture, or gives over in conversation.[ii] It was also something deeper, something that connected consciousness itself to the thermodynamic processes of nature. The intuition that "information" or "knowledge" or "intelligence" or "evolvedness"—call it what you will at this point—might be a universal quantity/quality binding man to nature was taken from the realm of pure philosophy and given a precise mathematical form—indeed, a form useful to this day.

Consider: after Shannon and Weaver, we say that the unit of informational stuff is the *bit,* the same bits, bytes (= 8 bits), kilobytes, and megabytes we hear about every day in connection with computers. Etymologically a contraction of the electrical engineering term "binary digit," a *bit* is an atom of a very particular kind. It represents the amount of information minimally involved, created, or required to make a decision as to which of two equally-probable, mutually exclusive events will, or did, transpire. Whether a machine or a mind does the "deciding" is of some interest of course, but it is of no consequence to the measure of informational magnitude of the choice itself, which has a "size" or content, in this case, of one bit. Unlike an electron or proton or other subatomic particle, a bit has no intrinsic physical qualities such as momentum or charge or spin. Nor is it made up of smaller things. It merely marks the collapse of two equi-probable alternative possibilities into one actuality.

Now this explanation of the nature of a *bit* is rather abstract, rather metaphysical. For one, we are forced to consider in what sense "a possibility" can exist other than in a hypothesis-framing mind.[iii] Are bits to be entirely subjective, mental? The answer is, No: bits can be conceived of as residing in or travelling between minds and machines and organisms and molecules indifferent to the mass-energy type of their host or purveyor. Bits can be emitted and absorbed, channeled and stored. When we make a telephone call or use a computer, for example, it is bits that travel through the wires and chips, here *as* bursts of electrons, there *as* magnetized atoms, here as groups of photons, there as the settings of mechanical switches. When chemical "messengers" travel within a living cell, or when DNA divides and replicates itself, or when neurons in the brain send trains of electrical pulses down nerve fibers to other neurons...in every case what is preserved and transferred and transformed is information in finite increments, in multiples of bits.[iv]

Be all this as it may, Shannon and Weaver's formalism has lent itself to many practical interpretations and insights. It may not tell us what information *really is* in some final sense, but it does tell us a lot about what information does and how it can be measured. (Readers familiar with information theory will find that they can move very quickly through what follows. They may wish to make note, however, of the particular way I define and identify entropy, uncertainty, and complexity.)

Information theory: Shannon and Weaver

Imagine a situation in which we know that there are only two possible outcomes. Flipping a coin is a good example: "heads" or "tails" are the only outcomes. The coin, we shall say, is unbiased, and the flipper competent.

Now, with every flip of the coin, *something is decided*. Before the coin lands we are uncertain as to whether it will land heads up or tails up. After the coin lands we are certain of the way in which it landed. Heads-up or tails-up...the matter is decided. While the coin is spinning in the air we say that the probability of heads (or tails) landing face up is "50:50" or 0.5. After the coin comes to rest, the probability of heads or tails facing up (as the case may be) is "100%" or 1.0, and becomes not so much a "probability" as a certain fact. This simplest, most elemental transition from ignorance to knowledge, from "50:50" uncertainty to 100% certainty, emits one bit of information into the observer's mind.

Shannon and Weaver gave mathematical form to this reasoning by designing a special numerical measure, *U*, for "*U*ncertainty." This measure works so that "U_{before}"—one's degree of uncertainty while the coin is spinning in the air—is 1 bit; and "U_{after}"—one's degree of uncertainty after it has in fact landed—is zero. (Why zero? Because unless the coin rolls out of sight, our uncertainty is eliminated). The information, ***I***, gained per flip is then defined to be equal to U_{before} minus U_{after}, which is equal to 1 bit.[v]

The uncertainty measure *U* is designed so that the *highest* value of U_{before} for a spinning coin—or for any system that produces only binary output, like "left-right-left-left-right...," or "yes-no-no-no-yes-yes...", is 1 bit per event. When U_{before} is equal to 1, that means we are completely uncertain of what the system's next state or output will be. In this situation we say that the system is behaving "randomly," or is producing a random binary sequence. When one is perfectly certain of the outcome, U_{after} is zero, as we have noted. But U_{before} can be zero too—if you know exactly what's going to happen. (Readers interested in the mathematics behind *U's* behavior can consult Footnote 5.)

Just as all good scientific formulae do, Shannon and Weaver's uncertainty measure tracks our intuition up to a point and then surprises us with its generality.

For example, let us imagine that the coin is *biased,* that we know it is biased, and that we know by how much it is biased. We would expect that any good numerical measure of our uncertainty about the coin's future behavior would have a value less than 1, because we know something about the coin's tendency to land heads-up. Consider a coin that we know lands heads-up 80% of the time over the long run. Is U_{before} for this coin less than 1 bit? Yes. For a coin known to be biased 80:20, U_{before} is equal to 0.71 bits per flip, which is also the greatest number of bits that can be "produced" by the coin with each flip. For a coin known to be biased 90:10, U_{before} is equal to 0.47 bits, which is the greatest number of bits that can be produced by *this* coin with each flip. And if, somehow, we had known beforehand with *complete* certainty which way the coin would land, the flip itself could produce *zero bits* of information for us. "Heads" or "tails," as the case may be, would be a foregone conclusion. General conclusion? The more we know beforehand, the less information we get.

Pretty tidy.[vi]

Figure 1.1 charts the value of U_{before} against coin bias. We see that U_{before} is greatest when heads or tails are equally likely, which is the same as saying, when we have no reason to predict one outcome over the other, which is the same as saying, when we are most ignorant or uncertain about the outcome of the flip.

Figure 1.1 How uncertainty varies with coin bias (or probability distribution in the two-possibility case).

Now let us consider a unbiased six-sided die. Instead of two possible outcomes, as with the coin, there are now six. While the die is being shaken, or while it is tumbling across the table, our uncertainty as to which number will come up is greater than our uncertainty was with respect to the spinning coin. With the fair die, U_{before} is equal to 2.585 bits. When the die stops rolling, the probability of *whatever number actually comes up* is equal to 1.0, and U_{after} is equal to zero bits. The most information, ***I***, that a fair die can produce per roll is equal to U_{before} minus $U_{after,}$which is 2.585 bits.[vii] An unbiased, rolling die has 1.585 bits more information "in it" than does an unbiased, spinning coin.[viii]

One more gambling example before we move on.

For an unbiased roulette wheel, with its 18 red and 18 black and two green ("0" and "00") numbered compartments, the probability of the ball coming to rest in a *red* (or *black*) compartment is 18/38, and the probability it coming to rest in a *green* compartment is 2/38. The probability of a specific *number* coming up (00, 0, 1, 2, ...36) is 1/38. Correspondingly, while the wheel is spinning, there are two categories of uncertainty: one about colors (what color will come up? red, black, or green?), and one about numbers (what number will come up? 00?, 0?, 1?, 2?...36?). Given "beforeness," Shannon and Weaver's formula for degree of uncertainty gives us U_{color} equal to 1.25 bits, and a $U_{numbers}$ of 5.25 bits.[ix] When the ball comes to rest, both uncertainties collapse. Both number and color are now singular facts: determined, known. U_{colors} is zero, and $U_{numbers}$ is zero too. When gamblers decide whether to bet on numbers or colors, they are deciding upon which of the two levels of risk they are willing to take. The rewards for taking the higher risk—betting on specific numbers—are proportionately higher.

There is one more interesting feature of Shannon and Weaver's formula. We already know that when all the possible alternative states or outputs of a system are equally likely to occur, then our uncertainty, *U*, is at a maximum. Interesting is the fact that this maximum is equal to the logarithm of the number of possibilities. More specifically, U_{max} is equal to $\log_2 N$ bits (where *N* is the number of possibilities and 2 is the base of the logarithm). This feature of the formula is both a computational short cut and a confirmation that when all possible outcomes are equally likely, the degree of our uncertainty depends only on the number of possibilities.[x] It also helps us see that any assignment of *un*equal probabilities to a system's *N* possible future outputs—this unequalness obtaining for whatever reason—will evoke less uncertainty in us than that the uncertainty it could theoretically, potentially, evoke.

Figure 1.2 How uncertainty varies with probability distribution in the many-possibility case.

So far, hopefully, so good. Now for some wrinkles.

We have assumed that U_{after} is always equal to zero. But this is not always the case. U_{after} might reflect a revised set of probabilities and be non-zero. This could happen when we learn that a die we *thought* was fair is, in fact, biased, or when we observe that the number "22" on a certain roulette wheel never comes up. In other words, the qualifiers "before" and "after" can attach to any event that changes how we want to assign probabilities to the possibilities. In these cases, the information gained, $\boldsymbol{I}$, is still equal to U_{before} minus U_{after}, but U_{after} is greater than zero and $\boldsymbol{I}$ is therefore less than U_{before}. One's initial degree of uncertainty is not erased; it is merely lowered.

In certain circumstances, however, it is possible for information-gained, $\boldsymbol{I}$, to be a negative number. This happens whenever U_{after} is greater than U_{before}. Although uncertainty is always zero or some positive number, by definition, one can easily have or receive "negative information." This is not information *gained* but information *lost*. For example, say that we believe a die is biased and then discover that it is *not*.—that it is perfectly fair. Then U_{after} is equal to U_{max}, U_{before} is less than U_{after}, and $\boldsymbol{I}$ is less than 0. We have received *negative information*. With this information we are worse off in the sense that our uncertainty about which number will come up on the next roll of the die has increased, but better off because now we are not *misinformed* about the nature of the die. So, we should not think of negative information as a bad thing: it may bring us to a truer representation of the situation. Indeed, *whenever we learn of new possibilities or options we receive negative information, and whenever we revise our judgment of the relative likelihoods of alternative events so as to make them more equal, we receive (or create) negative information, too*. Negative information makes the world potentially more complex, and this is often liberating. But at other times it is confusing or dissappointing. To go back to the actually-fair die: if we continue to bet in ways that depend on our previous (and mistaken) beliefs about its unfair bias, we will do no worse than if we randomly guessed at numbers—but also no better.

Negative information and *mis*information are thus quite different. Negative information about a system can no longer be produced once maximum uncertainty, U_{max}, is reached. U_{max} is like a ceiling. But *mis*information can be produced and re-produced indefinitely, and it can go undetected by changes in the measures U_{before} , U_{after}, and $\boldsymbol{I}$. This is because many different probability distributions on the same N possibilities can yield the same value of U. Two dice, for example, could be biased to the same *degree* and evoke the same degree of uncertainty in us,

U_{before}, but be biased differently with respect to the actual numbers favored and disfavored. Or, by some hidden process, the bias of a single die might slowly be changing *which* numbers come up with greater frequency than others...but without changing its overall degree of bias, which is all that *U* really reflects.

II. Complexity

We turn now to the question of *complexity*. In so doing, we translate our discussion of uncertainty into terms that make it easier for us to think about things-in-the-world having a complexity of their own, an *actual* complexity. This also allows us to take the position that we are *responding* to such actual complexity in some way when we become more or less certain about, or get more or less information from observing them. This is not to eliminate the subjectivity of uncertainty; it is only to take the point of view that our uncertainty generally has an objective cause, and to give it limits. Any knowledge, or any theory, that predicted the future state(s) of a given system perfectly would be knowledge whose own actual complexity matched that of the system exactly. Until more knowledge or a better theory proves otherwise, this complexity is effectively irreducible and objective: it belongs to the thing. And when we measure—or, more realistically, when we *estimate*—the actual complexity of a given system, we are saying something about the limits of what we (think we) *can* know about it.

The connection between *U*, uncertainty, and *complexity,* denoted *C,* is in principle quite clear: with a fixed level of prior knowledge, the more complex a thing actually is, the greater the degree of uncertainty we are likely to experience when first attempting to understand it better.[xi] Over the next few pages we explore the ramifications of this statement.

Webster's Third New International Dictionary gives us an opening definition of *complexity*:

> ...having many varied interrelated parts, patterns, or elements, and consequently hard to understand fully…marked by an involvement of many parts, aspects, details, notions, and necessitating earnest study or examination to understand or cope with...

Shannon and Weaver's formula for uncertainty addresses these factors rather well.[xii] To appreciate this, consider the two ways we have learned that the quantity *U* might increase. The first is by an increase in the total number of alternative possibilities, *N*, and the second is by a levelling, or equalization, of their probabilities. Now *N* can be thought of as the number of possible states the whole system can be in; and the various probabilities of those *N* states can be thought of picturing our confidence, or lack of confidence, as to which of these state the system-as-a-whole will be in next, which entails knowing or guessing how every part of the system will act. When *N* is very large and the probability of each and every one of the *N* possibilities is tiny but still positive, it becomes impossible to distinguish a very complex object (or process or phenomenon) from a randomly behaving one. For if to "understand" something is to be able to anticipate its next total state, or to be able to extrapolate the shape or behavior of a remote part based on a local part, and if to do this is to have in mind a *rule* that will predict or generate the phenomenon (either this, or knowledge of a set of *constraints* that the system obeys), then we can see how randomness can defeat us in the same way that extreme complexity can. *To us, an exceedingly complex thing might as well be a randomly behaving one*. This statement is worth emphasizing only because the belief that randomness is a form of simplicity is so widespread and so wrong. Paint spattered on the floor, "snow" on the television screen, noise from an air conditioner...these are easy to produce. But as easy as they are for us to produce, and as boring as they are to experience, these phenomena are very complex indeed. *Too* complex is their problem.

Take our spinning coins, dice, or roulette wheels. They are, on their face, very simple devices. Simple in construction (as compared, say, to a bacterium), they put forth an imbecilic stream of "decisions" as to face, number, or color, each decision totally unrelated to the one that preceded it. These sequences have no rhyme or reason, no discernible pattern. We call them "random." But they are each, at another scale, extremely sensitive to the world around them and therefore highly complex in behavior. After all, as the ball skitters around the whirling roulette wheel, who knows what myriad molecular fluctuations in conjunction with what tremor in the hand of the croupier, spin of the earth, and momentary disposition of the stars, really cause the ball to drop into 24 Red? Who knows what Coriolis forces and room drafts affect the spinning coin, or what microscopic bump on the table finally stops the die's tumbling? Our devices are not free to choose their next face or number or color "at will," or randomly. In fact, they are rigidly subject to external influences of the most microscopic, delicate, and multitudinous sort. They just happen to be influences we can't see.[xiii]

Given that we are paying attention, our having great uncertainty about a system thus coincides with its actually having great complexity—its every part being very loosely linked to every other and moving or changing with maximum individual freedom. By the same token, coming to have less-than-maximum uncertainty about a system coincides with our coming to understand how it is in fact ordered or *organized,* that is, how its parts do *not* do everything they *could* theoretically, systemically do if they were not constrained by each other or by some outside agency we understand better.[xiv]

Certainly, if a system is organized in some way and yet we have maximum uncertainty about it, then it must mean that the relevant information is missing *in us*. Conversely: if a system is actually extremely complex, and yet we are *not* perplexed by it, then either (1) we are holding a false and simplifying hypothesis about it that empirical evidence ought ultimately to undermine or (2) we are satisfied to know only general, and perhaps trivial things about the system: for example, that *it exists*, that it is "large," or "bluish." For example, while we can measure very precisely the pressure and temperature of a body of gas, these two numbers tell us very little about each of the trillions of individual molecules that are zooming about and bouncing off each other in every direction and at different speeds and that make up a *state*, at any given moment, of the whole body of gas. "Pressure" and "temperature" are macro-scale measurements (not unlike *U*) of billions of molecules at once, arrived at through a process of averaging. They are statistics. This does not make measurements of pressure or temperature (or uncertainty) intrinsically inaccurate or wrong. They usefully measure something about wholes, by design. (When we discuss markets and prices in Chapters Eight through Ten, this idea becomes central.)

Let us make the relationship between uncertainty, *U*, and complexity, *C*, more explicit. They are equivalent—except that uncertainty describes the state of the observer, and complexity describes the state of the system observed. If we define the *degree of organizaton* of the system, denoted *R*, as the *difference* between the system's *potential complexity* ,C_{pot}, and its (observed, known, effectively irreducible) *actual complexity*, C_{act}, then we can draw a strong parallel between *R* and our earlier definition of the amount of information gained, ***I***, when "U_{after}" is not retrospective but, rather, a *revised* "U_{before}" relevant to the next state or outcome.

To see this, assume that the system under inspection is behaving, we believe, "randomly." All we know for sure about it is the value of *N*, the number of alternative states it *could* next be in. Then U_{before} is equal to C_{pot}, and both are equal to $\log_2 N$.

Now imagine that we study the system's behavior for a while and find that its actual complexity, C_{act}, is lower than its potential complexity C_{pot} (as it would be if it showed any "preference" for some states over others). Our next U_{before} would be revised downward to match, and ***I*** would be equal to *R*; that is, *the amount of information gained would be a measure of the system's degree of organization*. This does not mean that there is nothing more to be learned about the system. On the contrary, until U_{before} and C_{act} reach zero, *something* mysterious remains. But if more *were* to be learned, we would immediately ascribe a new and yet-lower C_{act} to the system, and consider its *R* increased by the same amount.[xv]

Now, because potential complexity, C_{pot}, is always greater than or equal to actual complexity, C_{act}, *R* (unlike ***I***) is always a positive quantity or zero. But, like ***I*** , *R* may change over time. When *R* decreases, it means that the system has lost organization for some reason. It has become more *dis*organized. Perhaps it has absorbed negative information. Again, this need not be a "negative" development in a moral or evolutionary-fitness sense: decreasing *R* can sometimes be a "good thing," as can increasing *R*. In fact—to look ahead—I am going to argue that perceived and real changes in organization relative to complexity, up *and* down, are precisely where all judgments of value—moral, epistemic, aesthetic, and economic—can be found to originate. But this is later.

Note that a system's degree of organization, *R*, may increase not only when actual complexity decreases but when potential complexity increases, since it is the spread between them that defines *R*. This means that a system may seem—or actually become—more organized without actually changing its behavior, simply by increasing *N*, its total number of alternative possible behaviors. It acquires options it doesn't use, states it never visits. By the same token, a system may seem—or actually become—more disorganized without actually changing its behavior, simply by decreasing *N*. This peculiarity of *R* as a measure of organization would seem to be problem, but it is one of the most useful features of the theory of value we are developing with its necessarily large component of "psychology." By way of preview, consider the following thought-experiment:

Paul is asked to sit down in front of an apparatus with a small screen. The experimenter tells him that the screen will display four discrete colors, one per second, for two minutes. Paul's task is to figure out if there is an order to their appearance, and if so, *what* it is. He is to prove his understanding of the system's behavior by successfully predicting, several times, what the next color will be.

Later, another man, Quentin, is presented with the same apparatus and problem; but *he* is told that the screen is capable of producing *twelve* discrete colors. Quentin is to prove his knowledge also by predicting several times what the next color will be.

Now in truth, Quentin's screen still displays only four colors, just like Paul's. (Indeed, it's the identical apparatus.) Moreover, unbeknownst to either man, the experimenter has arranged the order of the four colors *to be random*. That is: each color has a probability of 0.25 of occuring. It follows that C_{act} and C_{pot} are both equal to 2 bits, and R is equal to zero. U_{before} is also equal to 2 bits.

The experiment runs.

After many failed hypotheses about the order in which they appear, Paul finally comes to the conclusion that the color sequence is either random or beyond him in complexity. Quentin, on the other hand, figures out that there are only four colors, and then also concludes that these four are randomly displayed. Here's what to notice: C_{pot} for Paul is equal to 2 bits, whereas C_{pot} for Quentin is equal to 3.59 bits. To arrive at the same conclusion as Paul, Quentin had first to realize that *there were eight colors that were not showing up at all*. This means that for Paul, the apparatus's degree of organization, *R,* was equal to zero, whereas for Quentin, the apparatus's *R* was 1.59 bits. Quentin had to dispel the misinformation given him about the number of potential colors, so he "profits" in information gained by the 1.59 bits that represents his reduction in uncertainty.[xvi]

<u>Figure 1.3 Paul and Quentin "face the World."</u>

Furthermore, if Quentin believes in the experimenter's truthfulness and doubts his own skills, he might continue to entertain the possibility that any one of the eight "missing" colors might show up at any moment. If he does, then he believes the experimental apparatus continues to be *organized*, to the extent of 1.5 bits. Paul, on the other hand, is disillusioned. He has nothing to show for his efforts to find organization: he started by giving each color an equal chance, and he ended the same way. Paul's only solace? He finds out that he was right.

In this experiment, it can be argued that Quentin comes out ahead, psychologically. Inasmuch as the screen's behavior in this experiment constitutes or symbolizes a "world," Quentin arrives at a world-picture whose potential complexity remains high and whose actual complexity is discovered to be relatively low—and thus a picture of a world with both possibility *and* order in it, actual complexity *and* organization. Paul on the other hand finds a "world" that is small and as complex as it can be—a random and unpredictable one. Quentin, deluded though he may be (after all, he was lied to by the experimenter), lives in a subjectively richer world than does Paul. He at least can speak of the eight colors that never show, but *might* if the world were less ordered.

But all is not lost for Paul. He, too, can delude himself into discerning a spread between C_{pot} and C_{act}. He might imagine—he might strongly *believe*—that there *is* an order to the succession of colors, just an order that he doesn't quite understand yet or that the machine itself is "trying" to follow but is inexplicably diverted from displaying to him. Like a gambler who won't quit or a scientist with a theory too sweet not to be true, Paul can believe the *actual* actual complexity to be lower than *apparent* actual complexity; and, as long as he does so, he too lives in a world of order and possibility. The absence of evidence to support Paul's belief that there is some order to the succession of four colors is no worse than the absence of evidence to support Quentin's belief that the other eight colors will eventually show.

In summary: Paul finds organization in the world because he believes that a theory *does* exist which proves that the world's complexity or randomness is illusory. Quentin finds organization in the world because he believes the world is *potentially* more complex than it actually is, and must therefore have been designed, or have evolved, in such a way so as to display the order it does.

Strictly speaking, all that this thought experiment can claim to show is how expectations and beliefs about world-behaviors can affect the measures of "information," "complexity," and "organization" as defined. But we can draw more meaning from it than this. It illustrates how the concepts of complexity and organization, even in their most straightforward exposition, might begin to help us understand some of our deepest philosophical questions: whether there is Design in the world or not, and whether we can learn anything about that design—or about creativity, truth, freedom, and many other things besides.

Clearly, we need to add another layer of complexity to our understanding of "complexity."

The complexity of fields

So far we have used examples that are all temporal sequences: "head,tails, tails, heads,...." in the case of the coin; "3, 4, 1, 1, 6, 3, 5, 1, 6,..." in the case of a die, "blue, red, red, blue, green, yellow, blue..." in the case of Paul and Quentin's screen, and so forth. We need briefly now to generalize the complexity argument to spatially extended systems; that is, we need to go from *strings* of discrete symbols, states, etc., to whole ensembles or *fields* of them.

Simplest case. Imagine four "cells" brought together to form a field, thus:

Figure 1.4 A field of four cells.

Now let each cell, at any given time, be in one of two states, "white" or "grey." At every tick of a clock, let each cell "decide" to change or not change color. Our question: what is the potential complexity of the field as a whole?

Extrapolating from the cases we have studied, we can guess that complexity would be a maximum when, after studying it to the best of our ability, we must conclude that with every tick of the clock every cell is just as likely to become or stay white as it is likely to become or stay grey. The *state* of the whole field (at a given time) is a particular color distribution on all four cells. How many possible field-states are there in total? The answer is 16, as per Figure 1.5, which presents an exhaustive listing. C_{pot} for this little four-cell system is then $\log_2 16$, which is equal to 4 bits (per tick of the clock).[xvii] Because each of our cells can be in only one of two states, each cell represents or carries at most one bit of information. Indeed, the potential complexity of any field of 1-bit cells or elements is just the total number of elements. If each of our cells could be in one of *four* states—say white, grey, black, or red—then each could carry 2 bits and the four together, as a group, could carry 8 bits per tick of the clock. And so on.

Figure 1.5 The sixteen states of a 4-cell ensemble, each cell capable of occupying one of two states.

And what of the *actual* complexity of the field, C_{act}? By the same calculations that we used with uncertainty, any information that helps us choose or predict future cell-states successfully, either singly or in groups (because of how cells seem to affect each other), will reduce C_{act} relative to C_{pot}, and cause R to be greater than zero. If we cannot divine such information, then, to all intents and purposes, R is equal to zero.

Consider the child's game of *tic-tac-toe* (in Britain, "naughts and crosses"). Tic-tac-toe uses a 9-cell field, where each cell is permitted to be in one of not two but *three* possible states: blank, X, or O. For each player, however, considering only their *own* next move, each cell can be in one of *two* states: for the X- player, empty or X; for the O-player, empty or O. An empty field, at the start of the game, thus represents 9 bits of potential complexity to the player who goes first. With every "move," one less cell is available for playing on. This is because cell states are irreversible. Potential complexity, C_{pot}, thus goes down with every move: 9, 8, 7, 6, 5...until the game is over, i.e., one player has succeeded in arranging 3 of his symbols in a row, or it concludes in a draw. Now to think of actual complexity. Each player is to some extent uncertain as to which of the remaining eligible cells he or she should mark next. If a player knows the rules of the game but lacks the intelligence or experience to play it well, he or she will assign roughly equal preferability to each eligible cell. To this player, each cell is about as attractive as the next, and so is the probability that he or she will select it to mark. The player with experience, however, will know with some certainty which of the remaining cells is *best* to mark. C_{act} for the experienced player is lower than C_{act} for the less experienced one. Since C_{pot} is objectively given, and because R is equal to C_{pot} minus C_{act}, this second player, we can say, is more *organized*. And he or she wins.

In Figure 1.6, I have analyzed one typical game of tic-tac-toe. Here the X-player is stronger than the O-player. The reader is invited to appraise the plausibility of the shown probability assignments.[xviii]

Figure 1.6 Move by move complexity of the field facing each player of a game of tic-tac-toe.

As an example of much more complex and more organized game, take chess. Each and every one of the 64 squares of a chess board can theoretically be in any one of 31 "states."[xix] But, starting at any one field-state—i.e. one whole-board position—it is not true that *any one* of the other approximately $8 \infty 10^{32}$ theoretically-possible board positions is as likely to follow as any other.[xx] Indeed, the vast majority of possible board positions could *not* occur next, given the rules that govern initial starting positions and thence, where, how many, and which chess pieces may move next, and given the tiny subset of all the still-possible arrangements that would be useful to improving either player's position. As complex as it is, the game of chess is still far from being as complex as a game *could* be that had the same number of pieces and the same size of board.[xxi] The rules of chess, in other words, because the forbid so many moves and disadvise so many others, are highly organizing.

Like tic-tac-toe, the degree of complexity and the degree of organization in a chess game changes throughout the game. Unlike tic-tac-toe, however, which is a "death game" inasmuch as it steadily constricts each players' freedom until one of them is squeezed out, chess is a "life game," opening up and developing before it tapers and finally closes down. Let us look at this trajectory more closely in complexity terms. In the opening game of chess (which constitutes the first ten moves or so), there are a large number of pieces on the board but relatively few possible places for them to move. All the pieces arranged in a compact, regimented, and highly organized pattern. From this initial regimented position, many of the open squares are unreachable because passage to them is blocked or a piece of the same color is already there. In the middle game, however—10 to 20 moves in—things open up. There are usually fewer pieces in play, but, since they are scattered over the board, there are a larger number of legal moves available to each piece, and therefore a *far* larger number of possible next board positions. Both potential complexity and actual complexity rise to high levels for both players. Finally, during the end game, there are very few pieces on the board, but this is compensated for by the fact that there are many more open squares to move to.[xxii] The middle game, being the most complex, both potentially and actually, makes the most demands on the player. This is why chess novices rehearse opening-game and end-game strategies first: they are easier to recognize and learn. A chart showing all this in idealized form might look like Figure 1.7.

Figure 1.7 The changing complexity of potential (i.e. legal and seen) and actual (i.e. legal, seen, and played) board positions, over an average 40-move chess game.[xxiii]

We have spent some time discussing tic-tac-toe and chess in terms of changing complexity because the pattern illustrated in Figure 1.7 characterizes a wide range of human activities, from solving mathematical problems (which tend to start with succinct statements, flower into long expressions with many alternative possibilities for manipulation, and then collapse into new and equally-succinct statements) to playing American football (where the ball is constrainedly entered into each "play" with the players arrayed in a static and highly-organized pattern, the play quickly develops in fluidity and complexity and openness-to-accident, and then converges again to a static conclusion with the ball temporarily "dead"). Courtroom trials, most novels and short stories, most movements in classical concertos and symphonies, songs in jazz... all of these start and end in states that are simpler and more organized than their middles. In *Art and Experience*, John Dewey took this pattern to be characteristic of all art, of all things beautiful, inasmuch as the pattern abstractly recapitulates the arc of a life—birth, growth and development (maturation), and resolution or death.[xxiv] In the Chapter Two we will look into this claim and this analysis more fully. It is redolent with our feelings about *value*.

III. Complexity, Organization, and "the Thrust of Life"

To tell the story of life's evolution in terms of complexity, organization, and information would be an enormous undertaking, involving physics and cosmology, chemistry and biochemistry, all of the biological sciences, the social and cognitive sciences, computer science, economic history, the history of art and of technology, and much more. No one writer could render it all except in the broadest of strokes, and only a few have taken up the challenge.[xxv]

If, further, one wants to find absolutely firm, scientific grounds for the view that the "thrust of life" is a thrust towards increasing organismic and social complexity—which is a common-enough belief, and one to which I essentially subscribe—certain obstacles stand in the way.

One obstacle is that not much is understood about "complexity." The sciences of complexity have blossomed only in the last thirty years or so with studies in artificial intelligence (AI), cellular automata (CA), and artificial life (AL) systems, and with explorations of "self-organization," "chaos," and "catastrophes" in certain branches of mathematics, theoretical biology, chemistry, and computation. In part because of their different disciplinary backgrounds, and in part because the question is so deep, researchers tend to define complexity in different ways. Indeed, defining complexity is still very much *the* problem facing researchers in the field, and the literature is rife with competing measures, models, and proposals.

What does it mean, then, to study complexity itself? To complexity researchers, it means that the system or phenomenon studied matters not very much: each is merely illustrative. They might study chemical reactions in laboratory flasks or thunderstorm formation on the prairies. They might track the rise and fall of species in ecosystems or the rise and fall of prices on the stock market. They might explore ways to evolve computer programs *in* computers or ways to prove difficult theorems *with* computers. They might study how organisms become adept at handling more difficult (read: more complex) tasks. They might model the interaction of large numbers of "agents" in "economies." It turns out that parallels abound. Graphs drawn in one study, slightly re-labeled, seem copied from another, and no one is surprised. Complexity itself, such scientists suspect, is an ineradicable feature of nature, like space or time, like magnetism, gravity, or energy, and has its own laws, its own ways.[xxvi,xxvii]

Since the *idea* of complexity is quite old, one might wonder why the science of it is so young—younger by far than chemistry and biology, say.

One reason is the only-recent availability of sufficient, affordable computing power to carry out the modeling required. Another is the only-recent pressure on humankind to understand the workings of the weather and the global biological environment. But a major reason is simply the maturation, through debate, of the theoretical insights first provided by such thinkers in the 1940s and 50s as Claude Shannon, Norbert Wiener, John von Neumann, Erwin Shrödinger, Gregory Bateson, Ludwig von Bertalanffy, W. Ross Ashby, James A. Miller, Herbert Simon, Marvin Minsky, Lila Gatlin, and many others who are less well known today. These individuals were all *systems theorists*, a term which at the time was quite radical because it asserted that everything *was* "a system," with inputs and outputs, with control mechanisms and feedback loops, and so on. The degree of generalization attempted was impressive, especially since it came not from philosophers (who were expected to come up with such things), but from mathematicians, biologists, and engineers. Here was a science of everything, a great unifier!

It did not turn out that way, as least not with the moniker "systems science" or "systems theory" still attached.[xxviii] Norbert Wiener preferred to call what he did "cybernetics"—the science of communication and control—a term that also became outmoded. But certainly, by 1960, it became much easier to think of information and complexity as scientifically approachable subjects. All the pioneers mentioned above were well aware (and perhaps none more than Bateson) of the implications of their information-based theories for life-processes, with von Neumann, Simon, and Minsky going directly to its implications for human decision making and human social interaction. Even when their focus was not upon evolution per se, each accepted the tenet that "evolution" and "growth-in-complexity" went hand in hand—indeed, that they somehow implied one another.

***The question of direction*.**

Complexity is not well understood. As we have seen, this presents something of an obstacle to asserting, without qualification, that evolution naturally favors complexity. But there is also contention amongst biologists and paleontologists, first as to whether evolution *has* a direction; second, *if* it has a direction, whether it is towards greater complexity; and third, if it *is* somehow towards greater complexity, whether this increase is really essential or just a side-effect of more important processes,

More questions arise from these. If evolution follows a path, does that path lead towards some goal or destination? And if the path leads to a destination, is it a destination we humans could see from afar, like a distant mountain peak? Or is it a something only God can see—a place or condition or state-of-the-planet-and-its-life-forms that we cannot even imagine? And if we accept (just for argument's sake, perhaps) that evolution traces out a path, might it yet be a path towards no particular goal or destination at all? Might it, rather, just be a movement up (or down) a slope that is only locally perceptible, a path leading nowhere in particular except towards more and more (or less and less) of whatever the "slope" represents—such as "biomass" or "spirituality"—ad infinitum?

Then again, perhaps *all* of these suggestions are nothing but bad poetry, the fruit of hope. Perhaps the process of evolution has no overall direction or destination, and shows no tendency towards any one condition over any other, not even for the shortest time period. Evolution might randomly "investigate" all of life's possible configurations, with whole species just as likely to wink out here in one month as proliferate there for a million years. Or perhaps, saving some vestige of directionality, the only tendency that evolution consistently exhibits is a tendency towards diversity itself: towards having light *and* heavy elements, simple *and* complex life forms, together, "all creatures great *and* small," strong and weak (but not *too* weak), long-lived *and* evanescent, tortoises and mayflies, all spreading, diffusing, and sniffing out new possibilities for reproduction and self-replication at the same time.[xxix]

Among those who see an inherent direction to evolution—myself among them—there seems to be some agreement as to what that direction is, and it has to do with the "diversity" just mentioned. Here are four related descriptions.[xxx] All, I think significantly, strike us as somehow positive in moral value. Take this as portent.

• *Over the long run, evolution proceeds from producing only small and simple organisms (usually in large numbers) to producing—producing* also—*larger and more complex organisms (usually in smaller numbers).* Thus: from pre-DNA self-replicating molecules to DNA-based self-replicating macro-molecules around 4 billion years ago, to the earliest living cells such as blue-green algae and bacteria around 3 billion years ago, to multi-celled plants and animals from 1 to 2 billion years ago, to vertebrates using sexual reproduction about 1 billion years ago, then to amphibians and reptiles (400 million years ago), to birds and warm-blooded mammals (100 million years ago), and lastly, to tool-using, language-using social animals such as chimps and ourselves (around 4 million years ago). A familiar sequence, this, but it is the story only of evolution's leading edge, complexity-wise. At each step along this way an increase in brain-size relative to body weight betokened an increase in the complexity and organization of the creature's responses to the environment—where "the environment," we should remember, might consist in good part of *other creatures* in the immediate vicinity, including large numbers of its own kind.[xxxi] It is less a story of succession and replacement than it is of splitting off, subdivision, and coexistence, so that today tens of thousands of species far less complex than ourselves share the material and energic resources of the earth (and of each other).[xxxii]

• *The evolution of species "proceeds from fewer, less specialized forms to a greater number of more specialized one ...exploit(ing) increasingly diverse physical habitats."*[xxxiii] As Robert Ayres remarks (the preceding sentence is his), this is conventional wisdom in evolutionary biology. Specialization through speciation—i.e. through the emergence of new species with more particular skills and tastes—reduces competition with other species for food, shelter, etc., and is therefore *organizing* in the sense that members of a more specialized species have less to worry about in defeating or negotiating with others who are interested in exactly the same general "goods." But specialization is *complexifying* too, in that in order to locate and wrangle from the world its unique supply of goods, the specializing species must develop sensitivities to many finer, previously-unimportant properties of an old environment. They must learn to pay attention to the unfamiliar contingencies of a new one, or in general work harder and smarter at staying alive and raising young. Speciation/specialization is all the more complexifying if the parent species' behaviors are not forgotten but remain part of the new species' repertoire.

• *Evolution coincides with increasing* efficiency *in capturing energy* (read: food, heat, light) *from the environment and putting it to metabolic use*. This trend is very broad, and there are a number of exceptions. Interesting to note is that this form of efficiency, called "assimilation efficiency," has a counterpart, a kind of efficiency at which earlier-evolved, lower, and simpler life forms excel over later-evolved ones. It is called "net production efficiency" and is a measure of the (energy) efficiency with which the mass of food consumed by a species is converted *not* into the day-to-day activities of individuals but into the *total biomass* of that species, this through enlarging individuals and reproducing them faster. At *this* kind of efficiency, plants excel over insects, little fish excel over big fish, herbivores excel over carnivores, and man—no creature's food—excels over none. Rabbits, for example, can turn carrots into more rabbits with an efficiency that is a marvel, but not half so fast as locusts can turn wheat into more and fatter locusts.[xxxiv] The greater assimilation-efficiency direction of evolution, if it does not always hold between species, does seem to hold within them, however. Modern varieties seem quite uniformly to be better—i.e., more efficient—than their ancient forbears at doing what they need to do in this regard. (The whole concept of efficiency, of course, goes to the heart of economics as a discipline, and we will have reason to examine it closely in later chapters. As we can already see, however, there is more than one kind of efficiency.)

• *The surface of the earth as a whole, as a locus of life, seems to be resisting the increase of entropy* (read: complexity) *in the cosmos as a whole, as decreed by the Second Law of Thermodynamics*.[xxxv] This seems to be happening in spite of the fact all member-creatures of the biosphere produce and export entropy through body-heat radiation, excretion, and general exuberance. Living things must be adding total "informational mass" to the planet (in the form of their durable bodies, architectural structures, behavioral habits, genes, and other information-accumulating, organization-preserving practices) *faster* than they "manufacture" entropy by ingesting nicely-structured molecules (food) and excreting them in worse condition, faster than they wreck and despoil the environment, radiate body heat in every direction, or shake free from rigid adherence to seasonal and diurnal rhythms.

On the other hand, it may not be true that the biosphere resists the cosmic increase in entropy in quite this either-or way (i.e., such that entropy can only increase *or* decrease in time). As biologists Daniel Brooks and E. O. Wiley argue in *Evolution as Entropy*, there are two relevant concepts of entropy (read: complexity) here: *potential* entropy, which may be increasing on earth just as it does cosmically, and *actual* entropy, which may not be increasing, or may not be increasing as fast. After all, the earth is not a totally closed system, and the Second Law applies only to totally closed systems. Our planet is an open system, meaning that energy comes in and leaves it all the time. Moreover, that incoming energy is not, and never has been, characterless in wavelength, or uniform in direction, or random in temporal pattern. For four billion years it has had distinct texture, geometry, and regularity—i.e. it has had considerable organization, *R*.

Consider: the earth has been fed for aeons by the regular contrapuntal rhythms of solar and lunar/tidal energies (and has radiated whatever of this energy was not absorbed into outer space). Add to these two sources of organization (as well as energy) the disciplining effect of *gravity*. This gave all evolving creatures the privileged directions of "up," "down," and "level" to which to adapt. Add the orienting effect of the cyclically changing inclination of earth's axis with respect to the sun. This gave all evolving creatures the seasons and locally decipherable latitude. Add the orienting effect of the earth's sphericality and rotation on an axis. This gave all creatures the distinctively different but regular "cardinal directions" of north, south, east, and west. Add the complementary organizing effect of geomagnetism, of prevailing and seasonal wind directions, ocean currents, of light, temperature, and salinity gradients, and so forth. Add to these again the faster rhythms of the lapping, flapping, and thrumming inherent in the motions of water and wind over elastic materials, and add the orderliness inherent in the swirls, eddies, bubbles, and pockets that develop in all convective, turbulent fluid flows. A veritable concerto of mechanical percussions and vibrations provided life its prelude and continues to provide its structural system. It is as though life’s "drum track" were layed down first. All of these sources of spatial and temporal organization (and others besides, as we shall learn) act as local retardants to the global increase in entropy. All serve to keep *actual* entropy increase on earth from keeping pace with *potential* entropy increase.

Thus, even as potential and actual entropies each grow in broad accord with the Second Law of Thermodynamics, there is, according to Brooks and Wiley, an *ever-widening gap* between them. And this ever-widening gap is sufficient to account for all of the "organization," "structure," and "information" we see increasing over evolutionary time in apparent contradiction to the Second Law.[xxxvi]

Figure 1.8 How increasing potential (or maximum) and actual entropy generates increasing organization (shaded) at evolution's leading edge.

Complexity and scale

We must return briefly to the workings of information theory. We will then be able to come back to Figure 1.8 and the work of Brooks and Wiley with more in hand.

We considered the degrees of complexity, C_{act} and C_{pot}, and the degree of organization, R, of linear, one-dimensional sequences such as coin flips and die rolls, and then of sequences of states of two-dimensional fields (with implications, I hope, for three-dimensional ones). Here we consider the important idea that different degrees of complexity and organization can co-exist in a single system—*at different scales*. To get a flavor of what this means:

Earlier we said that a clock contains more information than a pot. This is only true however at scales larger than several million molecules wide. If both the clock and the pot are made of the same material (say, brass), then at the molecular scale and smaller they contain the same amount of information. Indeed, under a powerful microscope they could not be distinguished from each other: brass is brass. The pot-nature of the pot and the clock-nature of the clock only begins to kick in, as it were, at the scale at which their identical molecules are organized differently.

This is the basic idea. Here are some examples

Language. Starting with strings again, look at this string of 16 letters and 2 spaces: gnsir tisoc ihtred**.** Imagine that this 18-symbol string is part of a decipherable *code* of some sort, one that we know beforehand uses only the 26 upper-case letters of the English alphabet and a space. How would we measure the information content of this string?

It is easy to compute the value of C_{pot}. We imagine that the string was generated by a 27-slotted roulette wheel, spun 18 times. Since any symbol is as likely to follow a given one as any other, C_{pot} is equal to $\log_2 27^{18}$, which is the same as $18\log_2 27$, which is 18 times 4.75, or 85.5 bits. With every spin of the roulette wheel, with every new letter or space, 4.75 bits was added to the complexity of the string. Given a finite, known-to-us-beforehand vocabulary of symbols, the total length of the string is directly proportional to the upper limit of the amount of information it could possibly convey to us.

Now the number "27^{18}" is an exceedingly large number, close to 6×10^{25}, and it is highly unlikely that our brains generate and check these trillions of possible strings to see which one(s) make sense, just as it is unlikely that the chess player's brain runs exhaustively through trillions of possible future board positions to choose the best next move. Certainly, *consciously* we do nothing of the sort, and there is no *neurological* evidence that we do so subconsciously either. How, then, is it done? How do we read written language so effortlessly (and play chess) at all? Because there is microstructure in the string: objective and manageable *organization* at different scales. Starting with our theoretical 6×10^{25} possible permutations of 18 letters and a space, and assuming that they were in a natural language, we can be sure that

(1) a tiny number of these 6×10^{25} strings would pass the test of being parse-able (divisible) into *pronounceable syllables* in that language,
(2) a very small number of the strings that survived the test of pronounceability would pass the test of containing actual *words* in that language,
(3) a very small number of these words in turn would pass the test of being *grammatically* linked,
(4) a very small number that would pass the grammatical-linkage test would make any *sense*,[xxxvii] and finally
(5) a very small proportion of those that would pass the test of making sense would be *relevant* to the *situation* at hand.

It is clear that natural languages impose drastic constraints on what sequences of letters are "legal." If our string of letters *is* a pronounceable, grammatical, sensible, and relevant statement in some language, then these constraints together amount to the string's total level of organization, R_{tot}.[xxxviii]

Now to the question of scale directly.

We observe that natural languages do not really consist in strings of letters (and spaces and punctuation marks) from an alphabet, although they may be *written* this way. These are just the language's atoms. Rather, they consist most fundamentally of permutations of a small family of pronounceable elements called *phonemes*. When written out as "syllables," in Romance languages for example, these phoneme-groups or syllables are apt to be two to four Roman letters in length and have at least one, and rarely more than two, vowels, of which there are five altogether. In other words, when we read a string of letters we suspect to be ordinary language, we do not consider the letters individually. The smallest unit sought is the pronounceable, familiar, syllabic phoneme, 3 letters long on average, and in some certain order.

Now, there are 26^3 or 17,576 ways to arrange 3 letters. Of these, about 2000 are humanly pronounceable as syllables, but less than 1000 of them actually appear in English. Which is to say, there are roughly 16000 combinations that are "impossible," i.e., that are simply not used in the language. This is English's organization, *R*, at the scale of syllables. We might estimate $R_{\text{syllables}}$ is equal to $(C_{\text{pot}} - C_{\text{act}})_{\text{syllables}}$, which is approximately equal to $\log_2 17{,}576$ minus $\log_2 1000$, which comes out to 4.2 bits per syllable.

The same sort of estimates can be made for English's degree of organization at the scale of *words,* which are groupings of one to four (or so) syllables or phonemes. We can ask: how many of all the possible combinations of one to four "English-legal" phonemes correspond to actual words in the English language? I shall not venture a calculation. The answer is: very, very few.

And once again we can ask: how many *N*-length sequences of English words, seen either as groups or as succeeding each other with certain (conditional) probabilities, are *grammatical*, or make *sense*, or are *relevant*?[xxxix] Each answer would give us a different estimate for *R*. There would be a unique *R* for letters, another *R* for phonemes/syllables, another for words, and so on.[xl] In 1950, Shannon himself went through the exercise of approximating English sentences by randomly generating letter strings with just such successively larger groupings and probabilistic constraints, with striking results: strings of random letters step by step turned into a close approximation of real, if silly, English prose.[xli]

What all this shows, stated quite generally, is that interpreting a string (or field) of symbols is not necessarily to be done one symbol at a time. Rather, it can be done by inspecting ever larger *groups* of adjacent symbols, and groupings of those groups, with memory and expectations unique to each group-length, or "scale." Over and above a string of symbols, then, one can have—and often *must* have—a *reading scheme*. And what is a "reading scheme?" It is a method, a strategy—itself organized—for finding and extracting the *varying* degrees of organization that exist at different scales in a string of symbols, in a sequence events, or in a field of changing states. Many of these reading schemes are learned; many are hard-wired into our brains; many—perhaps most—await our discovery. Moreover, reading schemes need not be linear or unidirectional, or rigidly fixed to one exact scale. They may use mixed strategies, searching for and detecting the degree of organization at different scales or "wavelengths" of an incoming stream or field information the way one might tune a short-wave radio for the clearest and most interesting stations, or survey a painting from its overall organization down to its most nested details, and back up again...[xlii]

All this in order to do what? to maximize what? If we are to believe Dan Sperber and Dierdre Wilson, it is in order to maximize *relevance* per se—this last and largest-scaled manifestation of language's organization.[xliii] I think they are right; but only if "relevance" is seen as something above and beyond the successful functioning of reading schemes themselves. For we must ask further, relevant to what? To our values? Needs? Interests? Clearly relevance does not exist in a vacuum: what is relevant changes with our beliefs and wishes and the situation at hand. "Relevance" and "value" are such closely allied notions, however, that it will take much of the rest of this book to explore the subject. At this point it is important only to make note of the typical variability of degree-of-organization, as measured by R, at different scales (and with different reading schemes) of the "same message." And I offer this as a preliminary hypothesis: *that the greatest relevance in a given message always resides at that scale* at *which—and under that reading scheme* with *which—the mathematical product of actual complexity,* C_{act}*, and organization,* R*, is at a maximum for the reader/observer*. Chapter Two will bring us closer to understanding why I venture this hypothesis.[xliv]

Music. A similar analysis could be carried out for music. A *musical scale* is made of a finite number of discreet sonic frequencies, or *pitches*, selected from all audible pitches, and called a *key*. Organization already: possibilities excluded. The two temporal units of music are the *note* and the *rest*. A *melody* is a sequence of notes and rests that can be parsed into groups of varying group-lengths, small to large. This parsing gives us "bars" or "measures," "phrases," "passages," "themes," "songs," "sections," "movements," etc., each of which can have its own level of complexity and organization. Notes sounded simultaneously, notes give us *chords*.[xlv] We consider harmonious and interesting only a select subset of all possible chords, and these chords generally succeed and supplant one another in certain, *genre*-constrained ways: in the West, "romantic classical," "blues," "ragtime," and other genres all have their own typical chord progressions. Certain composers tend to favor certain chordal patterns within these genres.

In addition, *within* each note, there is further room for variation having to do with the note's "substance" or sonic character: its timbre, loudness, sharpness, reverberance, and other qualities. In order to be organized and to convey information, such qualities, like the notes themselves, must demonstrate a finite number of possible characters, so that a "vocabulary" of sound-types (and notes) can become known to the listener—a territory of potential complexity within which the decision-making system that is composition or performance or appreciation can make its exclusions and surprise and satisfy us with its organization and complexity. Emphasis on loudness alone, for example, can form temporal patterns at several grouping-sizes concurrently. These constitute a major part of music's *rhythmic* organization, temporal groupings that delineate, result from, and sometimes depart from the groupings elicited by other sonic qualities such as note-proximity (in pitch and time) or the placing of rests.[xlvi] Indeed rythmic order can be impressed upon notes that are without determinable pitch because they are nothing more than packets of white noise (i.e. random frequencies).

But note the all-but-universal agreement among composers—jazz, rock or classical, Eastern or Western—that a musical pieces should open with simple "statements"—melodically, harmonically, rythmically, insrumentationally—and then proceed to increase in complexity over time as the composer can be sure that the listener having learned the fundaments of the piece, will be able to "follow"—see the organization in—the more complex stimulus.

And as for reading schemes that are not entirely linear or mono-filamented, these too are many in music. Consider the power of an orchestra-conductor as he brings a 30-instrument musical score to sonic life: how he reads and "misreads" the sheet music, molding the performance in emphasis, color, anticipation, pace, intention, effect...this in the light of what he knows about the piece as as a whole, about the composer's *oeuvre* and times, about his musicians' capabilities, his audience's taste, what other conductors have done, and so forth.

Clearly, music is an information-producing system of wondrous—some would say inexhaustible—richness. Music can reach levels of complexity and organization that take us rapidly to the edge of cognitive overload, to the limits of what we can coherently follow, think, feel, dance to, or produce. And yet it is spawned from a handful of variables, fewer even than phonemes. How much more wondrous it is, then, that all of biological life seems to be written in a music with far fewer notes than in any musical scale, in a language with fewer letters than any language, without intonations or inflections: namely, in the so-called *genetic code* of a single molecule: DNA.

DNA. Because DNA is discussed in later chapters for what it teaches us about positively valued evolution, we need to understand the way DNA works a little more thoroughly than would be necessary just to illustrate (again) how different degrees of complexity and organization can exist at different scales in a code. The workings of DNA also shed light on what I called "the thrust of life" in the previous section, and it brings us back to Brooks and Wiley as promised. And so I ask the reader to bear with me.

DNA, or deoxyribonucleic acid, is often called the genetic code. It is the "instruction book" for all known forms of life, from bacteria to man, and a copy of it is found in every cell of every creature.

Physically, DNA is a large super-molecule resembling a twisted zipper whose tines are pairings of four molecules called *nucleotides*—adenine (A), guanine (G), cytosine (C), and thymine (T)—positioned sequentially along its length. A, G, C, and T constitute DNA's four-letter "alphabet." These "letters" are ordered into enormously long sequences, some three billion in a row in humans. The DNA molecule however, unlike a single line of music or text, is a three-dimensional, coiled, folded, and motile object. Less a passive record than a moving score or cookbook, it guides the production of twenty different kinds of amino acids along with ways to combine them into proteins, the proteins that make up all living cells and living bodies. More than this, DNA is able to replicate itself. It does this by "unzipping" so that each side of the zipper can assemble a new and perfect partner for itself from a soup of random A, G, C, and T molecules waiting, like unattached dancers, to join their open-handed partners in the line...A to T, G to C, T to A, C to G.[xlvii]

Now if we looked at DNA as we might look at a text or musical score, we could represent the nucleotide[xlviii] zipper as something like this

.

Both sides of the zipper contain the same amount of information. Indeed they contain the very same information but for the fact that not all triple-letter sequences—called *codons* (which are like syllables, note)—code for the 20 different amino acids that constitute proteins. Some codons, such as TGA, TAA, and TAG, function like spaces in texts, or like rests in music, breaking up the codon string into "words" that start and stop protein-building processes. At a larger scale, groups of such words are separated by *introns* into "phrases" and "sentences." These groups in turn constitute *genes*, each around ten thousand codons in length. And it is genes that specify the thousands of traits, features, and processes that living creatures display.

Interestingly, even though every species uses a different complement of codons, arranged in a unique order, the complexity (C_{act}) of DNA at the scale of codons is very nearly the same across all species. This means that the major difference between the DNA molecules that characterize different species, information-*amount*-wise, is their total *length* rather than their intrinsic complexity: for *drosophila* about a hundred thousand base pairs, for humans about three billion. We shall revisit these last facts again in a moment, and then again in Chapters Two and Three.

So far, so good. Life seems to be encoded in patterns of "symbols" not too dissimilar to text, music, or even computer code, but with the critical, remarkable, and unique feature that this symbol string—DNA, life's code—in the right chemical environment *is able to replicate itself.* DNA is a blueprint *and* a copying machine, and thus a blueprint *for* copying machines, i.e. for itself.

Now it turns out that in progressively higher animals, a smaller and smaller fraction of the total DNA molecule actually describes or specifies amino-acid-to-protein assembly functions. Rather, as one ascends the evolutionary ladder of organismic complexity, a greater proportion of the total genetic code specifies what can only be called *reading schemes*, schemes of "reading" DNA itself *for* itself, that is, ways to parse nucleotides, back and forth, at different scales and groupings and with different, and sometimes overlapping, starting and ending points—which is to say, ways to parse the organization, *R*, inherent in the DNA nucleotide sequence at different scales. Single genes (say, the gene for hair color), researchers have found, are not composed of continuous strings of adjacent codons, but are distributed about the entire length of the DNA in several codon segments "grammatically" linked. "Genes," in this sense, are rather abstract entities, assembled assemblies of proteins (and other chemicals). Moreover, whole segments of DNA remain dormant until development reaches a certain stage. Then, triggered by events not yet fully understood, the operations that are carried out are carried out not on basic proteins but on already-formed cells and tissue, on whole "cities" of proteins, modifying them to form organs of yet greater complexity and usefulness to survival.[xlix]

A marvellous system this, to say the least.

Attempts have actually been made to measure the degree of organization of the DNA molecule at different grouping-lengths. The picture that emerges is one not very dissimilar from the one we proposed for ordinary language and for music, namely, *R* varies with the scale, or length, of nucleotide groupings. At the smallest scale of analysis larger than individual "letters," we have the group of three letters, the codon (which, at a stretch, is analogous to the syllable or phoneme). Already we find considerable organization. There are 4^3, or 64, different possible ways to arrange four letters in rows of three. The potential complexity of codons, C_{pot}, is equal to $\log_2 64$, or 6 bits per codon. *But there are only 20 amino acids*. It would seem that 44 possible amino acids are missing in action, as it were, they way eight of Quentin's twelve expected colors never showed up. This means that several different codons must—and indeed, we find, do—specify the *same* amino acid. For example, the amino acid leucine is encoded equivalently by AAT, AAC, GAA, GAG, GAT, and GAC.[l] And not one, but three other codons—namely, ATG, TAA, and TAG—specify breaks, "spaces," as we have learned. (It's as though all twelve colors did show up for Quentin, but that he was partially color blind.) In any given string of DNA, therefore, one has an actual complexity, C_{act}, which is less than C_{pot}. Moreover, the 20 amino acids that *are* specified do not appear with equal frequency over a long length of DNA. Empirical measures show that at the scale of codons the actual complexity of DNA per codon for all species is equal to 4.2 bits (plus or minus 0.04).[li]

Why the organization? Or, same question: why the redundancy? Analogous to our tolerance of a certain amount of bad spelling in writing, it is in good part to help make inevitable transcription errors during splitting and protein manufacture less harmful to the organism, evolutionarily speaking. For example, imagine an error that replaces a T with a C in the leucine codon AAT. No problem: AAC encodes leucine too! Other errors are not quite so ineffectual, of course, and some can cause mutations deleterious to the organism. But some are not. Indeed, some have the effect of better adapting the organism to its environment, and the laws of heredity are such that the progeny of that happy individual is more likely to inherit the improvement.

Figure 1.9, re-drawn here from the research of Daniel Brooks, David Cumming, and Paul LeBlond, shows how organization, *R,* increases with group length—from 1 to 6 nucleotides—averaged over 23 actual DNA sequences taken from across the bacteria-plant-animal-human spectrum.[lii] Their claim is that it is characteristic of all living systems that the degree of organization at smaller scales is less than the degree of organization at larger scales. Living systems, that is, are both limited in possibility and fairly random in structure at the scale of smaller components but are highly complex and very organized at larger scales, scales such as that of the organism-as-a-whole or its DNA program. This agrees closely, of course, with what we have said about language and music.[liii]

Figure 1.9 Increasing organization, *R*, with nucleotide group-length (from Brooks, Cumming, and LeBlond, 1988)

Brooks's team went further. They attempted to simulate biological evolution directly (and, course, very schematically) with a *cellular automaton* system.[liv] Would it produce the same patterns of organization that they found in natural DNA? I must leave the details of Brooks, LeBlond, and Cumming's CA experiment to the reader's interest.[lv] But in essence this is what they did: using a computer they allowed a population of 100 identical strings of 9 trinary digits (the "automata"), each representing (the DNA of) a "species," to replicate and mutate with varying probabilities of success over a long period of time. After setting a few parameters and initializing all the strings to "000010000," the system was left to run, i.e., to choose its own evolutionary course.

What happened? As in nature, occasionally, certain mutations, struck upon by chance, were very successful and a new species of identical-to-each-other automata was "born." Other mutations were not conducive to reproduction, however, and the species that inherited them were quickly extinguished. From this blind and mechanical process a large variety of what can only be called "genes" (sub-groups) of various lengths developed and stabilized, and in general the size of the gene pool—i.e. the number of different genes—increased.

Although there have been significant advances in research with cellular automata since,[lvi] Brooks et al.'s early model lent itself to an accurate accounting in strictly information-theoretic terms, terms indeed very similar to ours. What they found is summarized in Figures 1.10a and 1.10b.

Figure 1.10 Increasing complexity and organization of model genes with "evolutionary" time (*n* is equal to gene group length) (from Brooks, Cumming and LeBlond, 1988)

The most important point for us to notice in Figure 1.10 is how both organization, *R*, and (actual) complexity, $C_{act,}$ increase over both evolutionary time *and* gene group length.[lvii] They write:

> The progressive allocation of information and order [our "organization"] to the longer substrings...implies the emergence of higher levels of self-organization. If similar constraining rules alter changes in DNA base sequences on chromosomes, we would expect to find informational structure in DNA sequences similar to that produced by the...model.[lviii]

The process of speciation—the development of new species—adds organization, *R*. For an individual to emerge as the founding member of a *new species* (from being, previously, a member of a "variety" of a certain older species) means for it no longer to be able to mate and reproduce with the main body of its kind. This constraint represents a fissure, a division, a branching point in the evolutionary tree, because it reduces the number of possible "marriages" between a given population of individuals. Chance variation produces a generation of individuals—or just a handful of "mutants"—who can reproduce only with each other. Something similar can happen with the sudden geographic isolation of a small group of individuals from others of its kind. Over a long period of time, the gene complement of the isolated group becomes so different from the parent group that, should they be reunited, they can no longer "intermarry."

Across evolutionary time, the number of distinct species and families of species has increased enormously, this despite numerous cataclysmic setbacks: periods of species extinction that wiped out vast numbers of creatures. (Interestingly, these extinctions were especially hard upon populations of larger animals, the extinction of the dinosaurs in the late Cretacious period being only the most well known example). But, as Figure 1.11 shows, Nature seems never to give up. The long-term thrust is unmistakably towards greater numerosity and diversity of forms of life on earth. One might as well be looking at stock market prices since 1900 (or rather, the number of companies listed). *Life, it seems, "wants" more life.*

Figure 1.11 The growth in number of species over evolutionary time[lix]

In summary, from scientific research to date this much seems certain: first, that the increase in the number of more complex-and-organized species across evolutionary time is real, and second, that the simultaneous increase of their complexity, C_{act}, and their organization, *R*, has been made possible by the increasing length of the DNA molecules that guided their biological reproduction, because that length is what has increased total C_{pot}...all as illustrated in Figure 1.7.

With evolution, potential complexity *and* actual complexity increase, but the latter increases more slowly, so as to keep a "healthy" amount of organization, which is the difference between potential and actual complexity. This is the core of Brooks and Wiley's theory as well as ours, and we shall have many occasions to return to it.

The sources of this increasing complexity and organization are four: first, the onset of chaos and self-organization within the genetic mechanism itself through auto-catalytic molecule production (don't worry about this one); second, accidental mutations, deletions, and additions to DNA which prove to be adaptive to the organism; third, the fortuitous cooperative agglomeration of simpler cells and organisms into larger, more energy-efficient and robust units; and fourth, an environment that for all organisms comes to consist of other, ever-more-numerous and ever-"smarter" organisms. Most of these factors we have already noted. But from whatever source the tendency towards greater actual complexity comes, and from whatever source the tendency towards greater organization comes, the following principle holds: the extent to which the complexification of DNA (and of the behavioral traits that grow out of DNA) actually enhances the reproductive success of the whole organism is the extent to which those more-complex DNA sequences are passed on from generation to generation.

Because this passing-on process was (and is) accretive and cumulative, any particular organism's DNA-sequence today represents a complete record of the genetic history of its species—indeed of all species—from the beginning of biological time.[lx] Given a fixed degree of robustness to transcription and replication error, DNA can only complexify through lengthening. Hardly ever does DNA shortening happen without disastrous effects. This observation will be an important component later of our general theory of value: in a market system, the equivalent of DNA shortening goes on all the time.

IV. Some remarks on the origins of culture

Much of what I have had to say in this first chapter about the tendency for biological evolution to generate more complex forms of life in the long run is common knowledge, even ancient knowledge. For while bottom-up, Darwinian evolution taking millions of years to play out is a modern concept, arranging the creatures and elements of nature-as-found on a scale of *excellence* of some sort—with God (or gods) at the top, then man, then animals below man, then plants, then inert earth at the bottom—goes back to Presocratic philosophy. Indeed, one finds such hierarchies in almost all creation myths, most often with some narrative of serial begetting or temporally extended creation *ex-nihilo*. One might see them as proto-evolution myths. Identifying this "excellence" with complexity and complexity with evolvedness as we can do today, is perhaps only the latest stage in our trying to understand our place in Nature.

Rather recent, however, is the idea that the evolvedness or complexity of the individual members of a species is very much tied to the complexity of the social arrangements between them, first those arrangements that govern reproductive success directly, and then those that govern them indirectly. Which is to say: when animals live in groups, much depends on the timely *communication* of information between them—warnings of danger, locations of food sources, indications of power or sexual availability, transmission of useful hunting or hiding habits, etc. The resulting coordination of behavior increases each individual's chances for survival to reproductive age as well as their offspring's chances for survival to repoductive age (which is why it "pays" for individuals to live beyond their own reproductive years). For the most part, the *capacity* to send and receive such information is transmitted genetically, with some small amount of it learned anew by each creature in its own lifetime. Only with man does the inborn capacity-to-communicate extend to language, and, through language, to the transmission of techniques and technologies that increase the capacity-to-communicate itself.

Many begrudge culture to animals.[lxi] Whether or not this is well-advised, certainly the story of human evolution told only in the biological, genetic realm would be incomplete. Culture was essential and remains so. Man is nothing if not a communicating animal, a gossiping, scheming, plotting, dreaming animal who surrounds himself with the artifactual concrescences of previous generations. Stories are told and retold, skills and tools are handed down, rituals are copied and learned and repeated; questions are transmitted, books are written, schools are established, villages and towns stand firm for centuries. All of these give structure and continuity to human activities. With all this communication and extra-somatic information storage, another whole layer of evolutionary processes emerges from the biological, heredity-based ones. For mankind, the past—the historical past, at least—is engraved not so much in DNA as in the world, in material culture, in patterns of language, architecture, art, and living memory. Here "reproduction" is carried out through conscious learning and imitation. Here, in what is often called *cultural evolution*, it is ideas and practices that mutate, survive, and proliferate (or do not) rather than genes, or genes alone.[lxii] Even *bloodlines*—and what could be more genetic than these?—are maintained with extraordinary contestations of a purely cultural sort, with certain people distinguishing themselves from others solely by habitat, diet, dress, behavior, and language-use and rejecting contact with others who do not or cannot make the same choices.

For humans, biological survival to reproductive age is a necessary but not sufficient condition of life. People also need to be interesting and attractive enough to find mates, which takes a certain amount of artful self-creation as well as a measure of genetic luck. For humans, the only species whose members know that they will die, "more life" means not just a longer life for themselves but also a better life—a livelier, more lifeful, life-promoting life: health, security, attractiveness, intelligence, memorability, for themselves and for their progeny. Indeed, with cultural evolution, it is as though the whole set of processes we call "evolution" had themselves evolved to a more complex and organized plane of operation...with more, apparently, at stake than just the total biomass or numerical proliferation of the species.

In the chapters that follow we begin the transition from thinking about complexity in mathematical and biological terms to thinking about complexity in more psychological, social, and economic terms. In Chapter Two, we will pick up where we have just left off with the idea of cultural evolution, and try to show how biological evolution's major trends—speciation, efficiency, diversity, and so forth—continue to play themselves out in the human, social realm (albeit with new media and by more roundabout means). Once again, information-theoretical concepts will regulate our thinking. I will not try to survey the field of contemporary evolution writing, for once again we have a particular purpose in mind: we will be interested in fathoming the difference between "good" complexity and "bad" complexity. I will do this by proposing that an optimizing rule exists that connects back to the sorts of evolutionary fundamentals we have been studying.

Enter *value*.

NOTES to Chapter One: *Information, Complexity and Life*

---

i C. E. Shannon, "The Mathematical Theory of Communication," *Bell System Technical Journal* 27 (1948): 379, and C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, 1949). As original as it was, they saw themselves as extending the pioneering work of R. V. L. Hartley, in his "Transmission of Information," *Bell System Technical Journal* 7 (1928) 535–563.

----

ii See J. Von Neumann, in C. E. Shannon and J. McCarthy, Eds., *Automata Studies* (Princeton: Princeton University Press, 1956), 43; Norbert Wiener, *Cybernetics* (Cambridge: MIT Press, 1961); Leon Brillouin, *Science and Information Theory* (New York: Academic Press, 1956); W. Ross Ashby *Design for a Brain* (New York, Wiley, 1960); Satosi Watanabe, *Knowing and Guessing* (New York: John Wiley, 1969). Watanabe's book has been my standard reference for information-theoretical analysis. More introductory works include: Colin Cherry, *On Human Communication* (New York, John Wiley, 1957); J. R. Pierce, *Symbols Signals and Noise, The Nature and Process of Communication* (New York: Harper and Row, 1961); Jeremy Campbell, *Grammatical Man*: *Information, Entropy, Language, and Life* (New York: Simon and Schuster, 1982).

----

iii In his *Critique of Judgment*, Kant asserts that man is unique in his being able to distinguish between "the real" and the "possible." To God, every possibility is an actuality *ipso facto*. For God so much as to *think* about something is to bring it into actual being, while for animals to act is only to *re*act, i.e., to act without imagination, reflection, or ambition. Man is located between God and animal. For a discussion of Kant's views *vis-à-vis* possibility, see Ernst Cassirer, *An Essay on Man* (New Haven: Yale University Press, 1951[1944]), 56ff.

----

iv If the bit is the irreducible atom of "information-stuff," then it seems to exist at a mathematical rather than physical level. Indeed, if information permeates all scales of reality, as it seems to do, it might be because information-qua-information is the deepest of Nature's levels of self-description. And from this supposition it becomes not too far-fetched to imagine that something like "bit-fields" or "information fields" are foundational to the whole phenomenal world, foundational in that oldest sense of the verb "inform," which means "to give form to, or to animate." Information *is* order-out-of-chaos. It is *logos*; it is the Word.

For a discussion of bits, or information, as the most fundamental and universal stratum of reality, see John Wheeler's the "it from bit principle" in his "Information, Physics, Quantum: The Search for Links," in *Complexity, Entropy, and the Physics of Information* (Proceedings of the 1988 Workshop on Complexity, Santa Fe Institute, New Mexico, 1989), and the account given of it in John Horgan, "Quantum Philosophy," *Scientific American,* July 1992, 104. See also Gregory Chaitin, *Algorithmic Information Theory* (New York: Cambridge University Press, 1990 [1987]), and the views of Edward Fredkin as reported in Robert Wright, *Three Scientists and their Gods* (New York: Times Books, 1988), 1–110. Philosopher Alfred North Whitehead also came close to this view in the 1940's in his *Process and Reality* (New York: Free Press, 1978 [1929]). At the time of writing, the most accessible survey of contributions to the idea that information (not matter or energy or forces) is the subtrate of reality—that the universe is in some sense a computer—was Tom Siegfried, *The Bit and the Pendulum* (New York: Wiley, 2001)

Widely understood to be responsible for the building up of the elements in the structure we call the Table of Elements, Pauli's Exclusion Principle is a thoroughly informational affair: it prohibits two fermions (such as electrons) of the same spin state from inhabiting the same orbital shell around an atomic nucleus at the same time. This is Nature's enforcement of the indestructibility of a bit of Herself, as it were. For a further discussion of this idea and some wider references, see my "Cyberspace: Some Proposals" in Michael Benedikt, ed., *Cyberspace: First Steps* (Cambridge: MIT Press, 1991), 134–141.

Science may give some day give us practical "quantum computing." Here a bit will be represented by the the persistence/non-persistence of nature's tiniest and most delicate of states—the quantum state of a single atom

----

v Here is Shannon and Weaver's mathematical rendition of the formula for uncertainty, *U*. Let p(h) denote the probability of the coin landing *h*eads up and p(t) denote the *p*robability of the coin landing *t*ails up. Then, they proposed, let

$$\begin{aligned} U_{\text{before}} &= -\,[p(h)\log_2 p(h) + p(t)\log_2 p(t)] \\ &= -\,[0.5\log_2 0.5 + 0.5\log_2 0.5] && \text{........because } p(h) = 0.5 \text{ and } p(t) = 0.5 \\ &= -\,[0.5(-1) + 0.5(-1)] \\ &= 1 \end{aligned}$$

$$\begin{aligned} U_{\text{after}} &= -\,[p(h)\log_2 p(h) + p(t)\log_2 p(t)] \\ &= -\,[1\log_2 1 + 0\log_2 0] && \text{..........because } p(h) = 1.0 \text{ and } p(t) = 0 \\ &= -\,[1(0) + (0)] \end{aligned}$$

$= 0$

This gives us

$\boldsymbol{I}(\text{coin})$ per flip $= U_{\text{before}} - U_{\text{after}} = 1 - 0 = 1$ bit

How did they arrive at this curious procedure of multiplying the probability of an event by the logarithm, base 2, of that probability, adding a minus sign, doing this for all the possibilities, and then adding them up to get *U*? They got it from the work of R. V. L. Hartley (see Note 1), and where *he* got it from no one knows.

----

vi Again, for readers interested in the mathematical demonstration of this behavior: and working from the previous note: if p(h) is known to be 0.8 and p(t) is known to be 0.2, then

$$\begin{aligned} U_{\text{before}} &= -[0.8\log_2 0.8 + 0.2\log_2 0.2] \\ &= -[0.8(-0.322) + 0.2(-2.328)] \\ &= -[(-0.258) + (-0.456)] \\ &= 0.714 \text{ bits,} \end{aligned}$$

which is less than 1 bit, as we might expect. U_{after} remains equal to 0, of course, and so, if we narrowly define *information* as a *reduction of uncertainty*, we can say that the amount of information we have gained by the known-to-be-biased coin's actual behavior is not 1 bit but 0.714 bits. In other words, because we knew something about the coin's likely behavior beforehand, our initial uncertainty about what it would do on the next flip was reduced, and so the amount of information we gained by witnessing the actual outcome of the flip was reduced from 1 bit to 0.714 bits.

It may strike the reader as unfortunate that *the bit*, if it is to be thought of as a unit of information, can so easily have fractional value such as "1.457 bits." How much tidier the theory would be—how much more "cosmic"—if the bit were truly like the quantum, that is, if one could not have a bit of a bit! But if, in any real situation where we estimate information-quantities, bits, like inches or grams, will admit of fractional values, then we should realize that this actually says little or nothing about the situation as it could be described in finest detail. After all, by what microscopic mechanism might the coin be "biased"? Complexity lies hidden in this situation, hidden if not at another scale and/or coming from outside the coin-tossing system, then certainly within the human brain that makes such estimates of p(*i*) and *U*. It is not at all implausible that the finest-grain informational accounting of the whole eye-brain-hand-coin system would indeed give us an integer measure of bits, just as it would, for example, within the mechanism of a computer where bits are explicitly the lowest level unit of informational accounting and no half-bits are possible.

----

vii Written out (and henceforth omitting writing base 2 for the logarithm function), we have:

$$U_{\text{before}} = -[p(\text{"1"})\log p(\text{"1"}) + p(\text{"2"})\log p(\text{"2"}) + p(\text{"3"})\log p(\text{"3"}) + \ldots p(\text{"6"})\log p(\text{"6"})],$$

or, in more concise form:

$$U_{\text{before}} = -\sum_{i=1}^{6} p(i)\log p(i)$$

where, by convention, the upper-case Greek symbol *sigma*, Σ, with its super- and subscripts, commands us to "sum up the numerical values of following expression from the case when *i* = some-integer to the case when *i* = some-greater-integer, by steps of one integer."

----

viii This ignores all the other sorts of questions we might ask of these two objects, such as "what is your chemical composition?" or "what is inscribed on your two sides," in which case the coin might have more to tell us than the die.

----

ix Computed thus:

$$\begin{aligned} U_{\text{colors}} &= -[18/38\log 18/38 + 18/38\log 18/38 + 2/38\log 2/38] \\ &= -[(-0.511) + (-0.511) + (-0.225)] \\ &= 1.247 \text{ bits} \\ U_{\text{numbers}} &= -38[1/38\log 1/38] \end{aligned}$$

= 5.248 bits

----

x Some numbers to illustrate: For the coin, U_{max} is equal to $\log_2 2$, or 1 bit. For the die, U_{max} is equal $\log_2 6$, or 2.59 bits. And for the roulette wheel, $U_{max, numbers}$ is equal to $\log_2 38$, or 5.25 bits. Note that $U_{max, colors}$ = 1.25 bits, as computed in the note above, rather then $\log_2 3$ = 1.59 bits because the three colors—red, black, and green—do not occur with equal probability.

----

xi Readers conversant with the science of thermodynamics will already have recognized *U* as *entropy*, a term invented by Ludwig Boltzmann in the 1850s to describe the degree of unavailability of heat in a system for doing mechanical work within it. The higher the entropy, the less internal work it can do.

The relationship between the "entropy" of thermodynamics and the "entropy" of information theory (for Shannon named his *U*, entropy too, so similar were his equations to Boltzmann's) has been a long and vexing one. Physicists on one side, biologists on the other, computation theorists between, and science writers all over, have variously tried either to collapse or to separate the two roots of the term "entropy" while accusing the others of not *really* understanding it. Many feel convinced that somehow the (concept of) physical entropy—which arises from some of the deepest (descriptions of) irreversible processes in nature—*must* somehow be the same as (the concept of) informational entropy that applies to messages, computation, chemical and biological communication processes, and the like (as well as to cultural and economic systems). I am one of the latter.

In light of the ongoing controversy, however, I use the term entropy only when referring specifically to the Second Law of Thermodynamics (which states that in closed systems entropy increases over time), preferring to speak about "uncertainty," "complexity," etc. as seems best to suit the situation. In doing this I follow Jeffrey Wicken's advice:

> There is a more appropriate alternative to "entropy" in information theory. This is *complexity*. What the Shannon formula measures, simply, is complexity of structural relations.... It is in complex systems that *information* can be stored. Replacing "entropy" by "complexity" in dealing with structured relationships eliminates the connotative field of the second law (of thermodynamics) from arenas of discourse where it does not belong.
>
> From "Thermodynamics, Evolution, and Emergence" in Bruce Weber, David Depew and James Smith, eds., *Entropy, Information and Evolution* (Cambridge: MIT Press, 1988), p. 145.

I discuss the Second Law in more detail in Appendix One and in Chapter Two, where I offer a way to look at entropy that, I think, skirts many of the pitfalls of making complexity and entropy identical. The Second Law of Thermodynamics, I will suggest, describes a *certain set of trajectories* through a phase space, a set along which closed systems become simultaneously more complex *and* less organized in a certain combination, to wit, trajectories along which $complexity^2 + organization^2$ = a constant, and $d(complexity)/dt$ is greater than 0. There are many other trajectories that have similar effects and that are not the Second Law in operation. These do not, therefore, properly invoke the concept of "entropy" which remains the signature, as it were, of the Second Law and the of constraints I have just suggested.

----

xii Many others have made this observation. See especially Watanabe, *Knowing and Guessing*, p. 112.

----

xiii In his 1814 *Essai philosophique sur les probabilités* (*A Philosophical Essay on Probability*), the great mathematician Pierre-Simon Laplace famously said that if we could know the position and momentum of every atom in the universe at a given instant in time, then we would be able to calculate and predict the future of the entire world in perfect detail. Laplace's assertion was very much a thought experiment, for certainly, *obtaining* all this knowledge, as he knew full well, would pose no small problem. Today we might add that, even assuming we could have such knowledge, it is not clear that we could compute the future of the universe from this knowledge any faster than the universe actually *acted the future out*. As much as we might know, we can never "ambush" the world by getting to the future before *it* can.

There are other, more metaphysical questions that derive from Laplace's view—such as: if "probability" is *only* a measure of *our* ignorance, does it follow that the universe itself is deterministic, fixed in its course? I.e., is freedom an illusion? Big Question. To complicate matters further, in the 1920s and 30s quantum physics seemed to rule out the dream of "complete knowing" by *any*one in principle—up to, and perhaps including, God.

But in the classical world, Laplace had a point: at some arbitrarily detailed level of knowledge—if we *could* have such knowledge—many if not all "random processes" might turn out not to be random at all. They might merely be extremely complex: perhaps too complex to take the time to understand, perhaps beyond our capabilities to understand *fully*, but nonetheless "extremely complex"—not random. It is instructive in this regard to remember that what random-number-generators in computers actually generate are not random number sequences at all, only number sequences that are "random enough," i.e. sequences that have no particular order *as far as we can tell, or care to discover*.

With this observation, we come close to the modification of Shannon and Weaver's formulation of complexity/uncertainty offered by Gregory Chaitin in *Algorithmic Information Theory*, namely "algorithmic complexity" or "incompressibility." A phenomenon is complex, says Chaitin, to the degree that the bit-length of an algorithm that would *generate* or *simulate* the phenomenon exactly is not smaller than the bit-length of a direct description or record of it. If the algorithm is shorter than what it simulates, then the phenomenon is said to be compressible. The bit-length of the *shortest* such algorithm, which is one that cannot be compressed any further by any other algorithm, is (a measure of) the true complexity of the phenomenon (although, as Chaitin observes, one can never *prove* mathematically that one has found the shortest algorithm possible). So: a truly random number is algorithmically incompressible because there is no shorthand description of it and no system, method, or algorithm for creating it that would have fewer bits than that number. A random number is its own shortest description or generator. By contrast, a field of a million diamond shapes, or the sequence 123123123123123...can both be created and re-created by the simple instructions "translate diamond on vertexes 10^6 times" or "repeat '123' *N* times." They have "pattern," they are "understandable," we say, precisely when and because we know a shorter, generative description or recipe for them.

Last but not least is this pregnant point: coins, dice, roulette wheels and other random-number-generating devices are the product of extremely careful, non-random, manufacture, and are very symmetrical. We would have to throw away almost all of the dice made by a factory that produced randomly-biased dice beacause *only a tiny fraction of them would be perfectly unbiased.* The very concept of probability, and its two axiomatic constraints—(1) that the system have a finite number of mutually-exclusivepossible states, and (2) that the sum of the numerical weightings that represesent these possible states' relative frequency-of-occurence in the very long run add up to 1 (or any constant)—amounts to a massive pre-shaping of any underlying complexity or randomness into something quite orderly.

This is easy to see.

Take a random string of *N* positive integers, $i = 1, 2, 3, 4,....N$. Calculate the sum

$$M = \sum_{i=1}^{N} integer_i$$

Divide each integer by *M*. This gives

$$\sum_{i=1}^{N} \frac{integer_i}{M} = 1 .$$

This formula converts the initial *N*-length random string into a single probability distribution across *N* possibilites. The probability of any device, like an *N*-sided die, using this probability distribution to produce truly random numbers is very small indeed. It is the same as the probability that the initial random string of *N* integers consisted of one integer *N* times over, i.e. 10^{1-N}

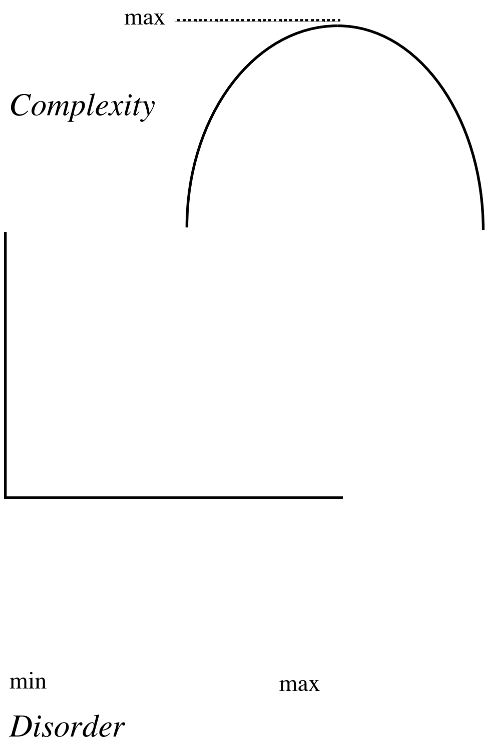
There *are* ways to produce probability distributions that consist of close-to-identical components (that is, where $p(i) \approx 1/N$ for all *i*) "randomly," that is by summing, multiplying, and genenrally chopping up random number sequences and recombining them in complex ways... But that is exactly the problem and the point: complexity knows no origin. Whatever procedure one uses leaves its mark: it is, after all, a *procedure*, which has order by nature. To get real randomness one must open oneself as directly as possible to the effectively-infinite actual complexity of the physical universe at the quantum level. This is what such finely balanced and symmetrical devices as roulette wheels try to do: to open themselves up to the minutest of influences to the complexity teeming at the smallest scales of reality all around them. And even then, structure, control, limitation, rears its (ugly?) head, if only in the mathematical rules of probability theory itself.

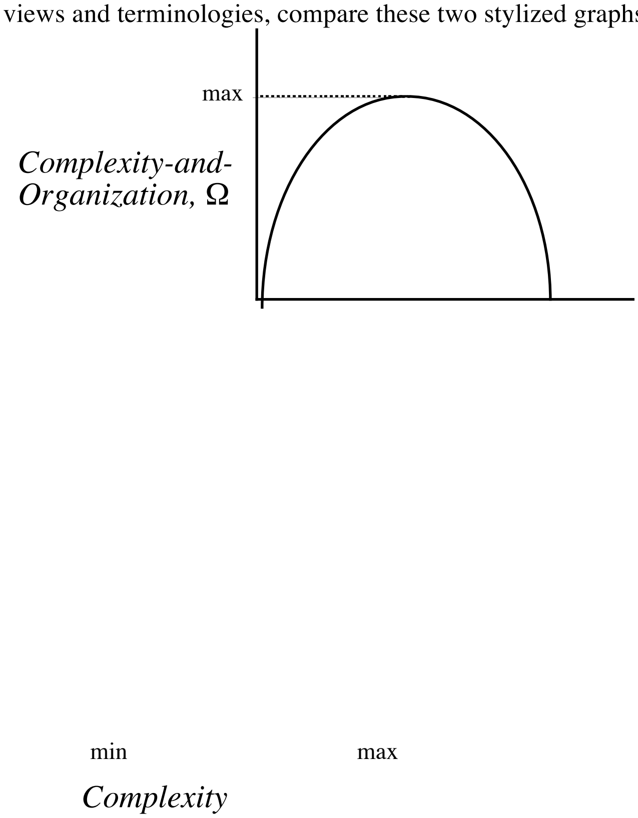
For an accessible account of current thinking about the impossibility of generating random numbers algorithmically, see Malcolm Browne "Coin Tossing Computers Found to Show Subtle Bias," *The New York Times*, January 12, 1993, B5. In this article, John von Neumann is quoted as writing, in 1951, that "anyone who believed that a computer could generate a truly random sequence of numbers is living in sin." See also Gregory Chaitin, *The Unknowable* (New York: Springer Verlag, 1999). See also Note 12 of Chapter Two, and Note 10 of Chapter Three and Appendix Two of this volume. My thanks to Prof. Charles Friedman of the University of Texas at Austin for helping me at crucial moments to think through these questions.

----

xiv The word "complexity" is sometimes reserved for the region that lies *between* complete predictability—order, regularity—on the one hand and complete unpredictability—disorder, chaos—on the other. (For example, Stuart Kaufman, *The Origins of Order* (New York: Oxford University Press, 1993).) Since both extremes are equally boring and useless, this leaves *complexity* (which intinsically somehow good) in the middle. *I do not follow this usage*.

Rather, as I elaborate In Chapter Two but must here preview, I use the compound term *complexity-and-organization* symbolized by Ω (omega) to occupy the middle region. Ω is at a maximum in the middle of the continuum between rigidity (ultimate "order") and chaos (ultimate "disorder") both of which lie at the expremes of complexity. To see the difference between these views and terminologies, compare these two stylized graphs:





<u>The usual scheme</u>

<u>My scheme</u>

I do not expect the reader to find the distinction I am making very clear at this point—or very telling. Chapter Two will shed more light on the advantages of my scheme.

The renowned particle physicist and complexity-theorist, Murray Gell-Mann, calls what I will call complexity-and-organization and Kaufman calls complexity, "effective complexity." He uses the tried-and-true monkeys-at-the-typewriter narrative to make the point:

> Compare a play by Shakespeare with the typical product, of equal length, of the proverbial ape at the typewriter, who types every letter with equal probability. ... (I)t is absurd to say that the ape has produced something more complex than the work of Shakespeare.

No real ape at a typewriter would or could produce a random output, of course, with every letter as likely to show as any other, and not just because he has likings and habits. A computer couldn't do it either, for reasons we studied in Note 13 above. Slight patterns would emerge, and so would many words, in different languages. I would say that the ape has indeed produced something more complex than a work by Shakespeare, indeed too complex by far, just not organized enough to be interesting. One just has to get used to the idea that high levels of complexity are not hard to produce: one just "channels" it from world around using uncontrollable media. Studio artists and architects will know what I mean.

Gell-Mann continues:

> Randomness is not what we mean by complexity. Instead, let us define what I call effective complexity... A random (incompressible) bit string has no regularities (except its length) and very little effective complexity. Likewise something extremely regular, such as a bit string consisting entirely of ones, will also have very little effective complexity, because its regularities can be described very briefly. To achieve high effective complexity, an entity must have intermediate AIC (algorithmic information content) and obey a set of rules requiring a long description. But that is just what we mean when we say that the grammar of a certain language is complex, or that a certain conglomerate corporation is a complex organization, or that the plot of a novel is very complex—we mean that the description of the regularities takes a long time.

Murray Gell-Mann, "The Simple and the Complex," in David S. Alberts and Thomas J. Cyerwinski, eds., *Complexity, Global Politics, and National Security* (Washington, D.C.: National Defense University, 1997), found online at *www.ndu.edu/ndu/inss/books/complexity/ch01.html*.

The difference is not only one of nomenclature. For with Laplace and with Einstein (when he famously averred that "The good Lord does not play dice"), I do not believe that there *is* randomness in nature, just blinding, unimaginable, and perhaps unassimilable complexity at the finest scales, a complexity which *looks to us* like randomness, disorder, chaos, etc. The judgment that "sequence *x* is random" just means that *x* is too complex for us to understand. For Laplace, Einstein, and Shannon, increasing "complexity" is identified directly with increasing unpredictability-by-us and with increasing apparent disorder (inasmuch as *all* disorder is only apparent). Gell-Man wants to get around this by using qualifier "effective," thus conceding in a way to the tradition in physics of allowing randomness to coincide with maximum statistical complexity. Randomness is ineffecitve complexity. His "effective complexity" is my "complexity-and-organization," Kaufman's "complexity," and Chaitin's "organization" (see Note 32).

The reader who wants to throw up his or her arms at all this has my sympathy.

----

xv In Chapter Two I will modify the mathematical definition of R from $R = C_{pot} - C_{act}$ to $R^2 = C_{pot}^2 - C_{act}^2$. None of the qualitiative arguments made in this chapter are affected by this modification.

----

xvi Expressed numerically:

$\boldsymbol{I}$(Paul) = $U_{before} - U_{after}$
= (2 – 2)
= 0 bits;
$\boldsymbol{I}$(Quentin) = (3.59 – 2)
= 1.59 bits.

----

xvii The arithmetic behind this runs as follows: $C_{pot} = \log_2 N = \log_2 M^L = L\log_2 M$ where N is the number of possible states, M is the number of cell-states, and L is the total number of cells in the field. In our example, M is equal to 2 and L is equal to 4. Remembering that our logarithms are always to the base 2, this gives us $C_{pot} = \log_2 2^4 = 4\log_2 2 = 4\ (1) = 4$ bits.

----

xviii The author carried out a pilot study with 20 children aged 5 to 6 years old. The probabilties illustrated in Figure 1.6 reflect those inferred from the data gathered, but should be taken with a pinch of salt. Children of an age to find tic-tac-toe interesting are also of an age not to follow instructions from fun strangers in their classroom very closely. Researchers interested in improving this study should consider observing, and controlling, one pair of players at a time. Why use children? Because adults, of course, are apt to resign the moment the first player marks a corner cell.

----

xix There are 14 white and 14 black pieces and they can each, legally, be anywhere on the board, and 4 bishops that can only be on "their" color; or the square may be unoccupied, which is one more "state." There are other ways to count this. For example, making all white pawns indistinguishable from each other, and all black pawns indistinguishable from each other, so that there are really only 9 states (i.e. one of the noble pieces, or a pawn) for each square, or perhaps 8 states if we take into account that white squares can only have "the white bishops" on them and black squares only "the black bishops", and so forth. The gist of my argument however does not change with this more sophisticated accounting.

----

xx I assume a middle-game-type position with 20 distinct pieces, including the empty state, remaining on the board and (64 – 10) places to put them, regardless of chess rules or chess sense. This gives us $N = 54!/(54 - 20)!$ permutations without repetition of pieces.

----

xxi If almost all possible positions *were* legal in the sense of being possible to move *to* from any other position, the average chess player would have to process around 300 bits of information per move. He would have to process *n* times more than this for every *n* moves he looks ahead—this aside from the internal processing used in judgment, i.e. *comparing* the merits of one move with another. This is not a lot, actually: information pours onto the average television screen at 18 million bits per second, and, thanks to four million years of evolution and the ten-billion-neuron brain that that process has bequeathed us, we see, in this organized torrent of twinkling light, the shapes of cars and people and aspirin. Such are the marvels of our visual system.

At the time of writing, IBM's champion chess-playing program "Deep Blue" is running on an RS/6000 SP computer with 512 specialized microchips working in parallel that can generate and compare 50 billion board positions in 3 minutes. Three minutes is the average time it takes for a human to make a single move, and in that time to evaluate, at best, 20 alternative board positions with any completeness. Clearly, humans don't use brute force to solve chess problems. At least not consciously.

It is easy to be impressed by Deep Blue. The deeper truth is that a computer as large and fast as Deep Blue, and programmed to the limits of its processing capacity, still could not translate an arbitrary sentence from English into French and keep its meaning, nor steer a car through an intersection without crashing. Weighing nearly one-and-a-half-tons, Deep Blue has the intelligence of a small insect, and perhaps even less. That Deep Blue can challenge and beat the world's best chess player, Garry Kasparov, as it did on May 11, 1997—and that *Tinyware*'s Pocket Chess on my Palm Pilot beats me, even on level 2 out of 8—only attests to what terrible chess players human beings are. A two-dollar electronic calculator can beat the pants of any mental-arithmetic whiz. We're just terrible at arithmetic, too.

No, the right thing to marvel at in chess-playing humans is how well they are able to discipline their emotions, focus their perceptions, and direct their myriad errant thoughts to grappling with the clear and arid "problems" of chess. We were "designed" to hunt and weave and dream and laugh at jokes, activities we would know to be more complex by far than chess or mathematics if we could ever see our own brains lit up, end to end, in their enjoyment.

[xxii] This is *not* the case if the last few pieces in the game are intrinsically limited movers, like kings and pawns, in which case the game becomes rather simple.

[xxiii] Note that this figure includes a new complexity, "C_{max}" or maximum complexity. This is $\log_2 N$, where N the mathematically-computed number of all possible legal next board positions. No human player is able to think of every one of these positions, let alone consider their advantages and disadvantages. Only computers can do this, which accounts for their prowess. As illustrated by the example of Paul and Quentin earlier, "potential complexity" is intended to refer to the greatest number of states that an informed person could reasonably imagine being possible for the system he or she is confronting, with all rules-of-the-game and other limits in place. As I note in the text, there is a lot more information to be had from a real spinning coin than whether it will land heads or tails. That we say there are only two states possible for the coin (once landed) is a significant *a priori* limitation on what we see the "system" *as*.

[xxiv] It is also the trajectory of many battles in war. Unless one is a surviving victor, and a member of a warrior culture, one finds it hard to call such battles “beautiful.” This is because, although the battle’s pattern might be like a life or a game of chess in the abstract, war is no art form, and the pain and death it actually deals by far outweighs any life-enhancement dervied from contemplating its pattern.

[xxv] Most notable recently has been Robert Wright's *NonZero* (New York: Pantheon Books, 1999) which I will review in the Chapter Three. Others include astrophysicists David Layzer in *Cosmogenesis* (New York: Oxford University Press, 1990) and Lee Smolin's *The Life of the Cosmos* (New York, Oxford University Press, 1997), engineer Susantha Goonatilake's *The Evolution of Information* (London: Pinter Publishers, 1991), computer theorist/biologist Tom Stonier's *Information and the Internal Structure of the Universe* (London: Springer Verlag, 1990), environmental economist Robert Ayres' *Information, Entropy and Progress* (Woodbury, New York: American Institute of Physics Press, 1994), physicist Murray Gell-Mann's *The Quark and the Jaguar: Adventures in The Simple and the Complex* (New York: W. H. Freeman, 1994), chemist/physicist (and arch-rival to Gell-Mann on this topic) Ilya Prigogine's *Exploring Complexity: an Introduction* (New York: W. H. Freeman, 1989), *Order Out of Chaos: Man's New Dialogue with Nature*, with Isabelle Stengers (New York, Bantam Books, 1984), and *The End of Certainty* (New York: Free Press, 1997) also in collaboration with Stengers; and the science writer Jeremy Campbell's *Grammatical Man: Information, Entropy, Language, and Life*, (New York: Simon and Schuster, 1982) and Brian Goodwin's *How The Leopard Changed Its Spots: The Evolution of Complexity* (New York: Charles Scribner's Sons, 1994). See also: Nicholas Georgescu-Roegen, *The Entropy Law and the Economic Process* (Cambridge, Mass.: Harvard University Press, 1971), Jeremy Rifkin, *Entropy: a New World View* (New York: Viking Press, 1980) and Rudolf Arnheim, *Entropy and Art; an essay on disorder and order* (Berkeley and Los Angeles, University of California Press, 1971).

For a constantly updated listing of publications in complexity science, the reader might visit the websites *www.cpm.mmu.ac.uk/~bruce/combib/* and *www.vanderbilt.edu/AnS/psychology/cogsci/chaos/cspls.html*

[xxvi] Indeed, complexity as a quantifiable, basic, property of nature first entered the world of physics with the founding of *thermodynamics* in the 18th Century and with such concepts—developed in the 19th Century and still current—as *entropy,* which is an entirely physical property of all aggregates of matter. "Information," as we have seen, is a member of the same family of terms.

The terminological connection between "entropy" and "information" was the result of happenstance. John von Neumann rather casually suggested to Claude Shannon, who worked down the hall at the Institute for Advanced Studies at Princeton, that he call his new uncertainty function "entropy" because it was very similar to Boltzmann's entropy function. Since no one understood what entropy *was*, von Neumann joked, it would be hard to argue against! Ever since then, shelves of books have been and continue to be written under the rubrics of "chaos" and "self-organizing systems" about the relationship of *entropy*, understood thermodynamically, to *information* understood statistically, computationally, and cognitively.

The reader should be aware that in sober scientific circles, skepticism abounds with regard to how much light "information theory" and the "sciences of complexity" can shed on human phenomena such as social organization, meaning, and valuation. For example, here is noted science writer, John Horgan, in a review article directed mainly at the hype surrounding the work of scientists studying complexity at the Santa Fe Institute:

> **Information theory.** Created by Claude E. Shannon in 1948, the theory provided a way to quantify the information content in a message. The hypothesis still serves as the theoretical foundation for information coding, compression, encryption, and other aspects of information processing. Efforts to apply information theory to other fields, ranging from physics and biology to psychology and even the arts have generally failed—in large part because the theory cannot address the issue of meaning. ("From Complexity to Perplexity," *Scientific American*, June 1995, pp. 104--109)

Needless to say, I do not agree with Horgan that "information theory cannot address the issue of meaning." I hope this chapter is beginning to show that.

[xxvii] As I began this chapter remarking, information (and therefore complexity) might be a constituent of the universe in a deep physical sense, what Descartes or Spinoza would have called a substance. It would seem, for example, that, *contra* Einstein, light *can* travel faster than "the speed of light" (300,000 kilometers per second); but that it is *information* that cannot. If true, this would give a whole new slant as to what the General Theory of Relativity is about. See James Glanz, "Faster Than Light, Maybe, But Not Back to the Future," *New York Times*, May 30, 2000, D1. This is a report on D. Mugnai, A. Ranfagni, and R. Ruggeri, "Observation of superluminal behaviors in wave propagation," *Physical Review Letters* 84, 4830-4833 (2000). See also: http://helix.nature.com/nsu/000601/000601-5.html#1

[xxviii] Today "systems engineers," "systems integrators," and the like are apt to be computer experts of some sort dealing with "information systems." Management schools are another place where "systems thinking" still holds sway *under that name*. For the fact of the matter is that systems theory has been so completely absorbed into our language, culture, and economy that it's now tantamount to common sense, and its once-radical name has been discarded.

[xxix] Such propositions are not entirely susceptible to proof or disproof by empirical evidence. A measure of faith is involved, and of temperament. Indeed, many think that if evolution is defined by the phrase "the survival of the fittest," and "the fittest" are defined as "those that survive," then the whole theory of evolution itself is based on little more than a tautology—something like "the success of those that succeed," or, a bit better, "the improved reproductive success of the offspring of those who have had reproductive success."

On this view, evolution does not reliably produce anything new or better or different or more (or less) complex. The best one can say of evolution as a principle is that it shows how proliferation is the reward of proliferation; and whatever qualitative trends evolution seems to have exhibited thus far are secondary, transient, and no more meaningful than the faces we see in clouds.

The unsatisfactory nature of the "survival of the fittest" formulation is discussed at some length by Robert Jastrow, ed., in *The Essential Darwin* (Boston: Little Brown, 1984) xvi. According to Jastrow, the phrase "survival of the fittest" is actually Herbert Spencer's and was borrowed by Darwin and inserted into *Origin of Species*, something Darwin lived to regret.

[xxx] I am here following Robert Ayres's summary in *Information, Entropy and Progress* (Woodbury, New York: American Institute of Physics Press, 1994) pp. 115-116. He gives six. I have combined his numbers 4 and 5. Ayres also adds another description of evolution's direction: "The biosphere tends towards zero natural waste (perfect recycling), at least over the long run." (p. 116) With solar radiation—its absorption and re-radiation—with tidal forces caused by the moon and with geothermal energy also rising to the surface, the biosphere is energically an open system. But, meteoroids and spacecraft aside, the earth as a whole is a closed system materially. The waste of one creature is eventually food or habitat for another. That there is free oxygen in the atmosphere at all for us and all the animals to breathe, for example, is the result of micro-organismic activity billions of years ago. I do not include this observation in the main body of the text since I don't think it qualifies as a direction for evolution.

See alo Francis Heylighen, “The Growth of Structural and Functional Complexity during Evolution” in Francis Heylighen, Johan Bollen, and Alexander Riegler, eds., *The Evolution of Complexity* (Boston, Mass.: Kluwer Academic Publishers, 1999) and available online at *http://pespmc1.vub.ac.be/Papers/ComplexityGrowth.html*

[xxxi] A new *physical* niche for an organism may not be more complex than the old niche and therefore more demanding of, and rewarding to, more complex behaviors. Indeed, a large number of successful population migrations, one must suppose, were from difficult to benign environments, environments which would, if anything,

allow the relocated population to grow large in number and to spawn simpler variants of itself that were just as successful. In *On the Origin of the Species*, Darwin called change driven by struggle with the physical environment *abiotic*, and did not see in it any necessary spur to increasing organismic complexity. Evolution in this mode could go either way, complexity-wise. It was *biotic* competition—competing with other living creatures and populations—that would reward growth in brain size, neural and anatomical complexity, etc. Darwin's distinction is still not enough for Gould, however. Gould will have no truck with the idea that complexity has either a causal origin *or* reward, let alone any teleological thrust. See Stephen Jay Gould, *Full House* (New York: Random House, 1997), 142.

The thinker perhaps most antithetical to Gould's ideas was Pierre Teilhard de Chardin, the Jesuit mystic and paleontogist who died in 1955. Author most famously of *The Phenomenon of Man* (completed in 1940, published in 1955), Teilhard argued for a cosmological evolution that included the development of human consciousness, the rise of Christianity, the emergence of a *noosphere* or globe-encircling supermind through global communications, and finally a convergence in the distant future upon the "Omega Point," which is the reconciliation and unification, in extreme complexity-consciousness, of Man, God, and a now-totally-living-universe (an idea from the ancient Greek Christian theologian Origen). This cosmological unity constitutes also the second coming of Christ. Somewhat less deliriously, Henri Bergson had expressed similar views in his 1907 *Creative Evolution* for which he won the 1927 Nobel Prize in Literature; as did Oxford philosopher Samuel Alexander in his 1935 *Space, Time, and Deity*.

----

xxxii There can be little doubt that humans occupy the most complex end of nature's productive spectrum. We *are* the smartest of all animals when it comes to thinking, planning, and making things. But in terms of sheer reproductive success, ruggedness, longevity as a species, total biomass, and functional importance to the rest of nature, bacteria and insects are superior to humans by far!—a point made, I think strangely over-forcefully, by Gould, in *Full House*.

The reader may be wondering why sheer body size does not overwhelm any accounting of complexity. After all, an elephant is perhaps 200 times larger than a chimpanzee, and is therefore made up of quadrillions more molecules that *could*, each, be doing "it's own thing." What a vast C_{pot} difference this represents! Why aren't elephants or whales, or, for that matter, giant redwood trees, at the top of the complexity tree?

The answer has to do with how similar to each other and repetitious are the *assemblies* of molecules that make up a creature's cells and juices and organs. 1000 pounds of muscle-tissue is no more complex or organized *per pound* than 1 pound of muscle-tissue. The most complex and organized creature is one not only that is a certain *size*—for size *does* help, potentially—but one whose *parts*, as articulated within that overall size, are (1) *different* from each other in structure and function, across at a wide range of scales, as well as (2) *coordinated* to communicate and work togther to a certain extent, i.e., are organized but not *too* much.

This is why total DNA-length is a fairly good measure of organismic complexity. Think of DNA as a computer program that runs a machine that manufactures a certain animal tissue, hormone, or whole organ. If we let a program of a certain size, say X bits, keep running, it will start repeating itself. Like a stuck record, it will churn out the same protein patterns over and over—pound after pound of whatever tissue-type it was "written" to produce. A longer program, say Y bits > X bits, will not repeat itself so soon, especially if it also "knows" how and when *to turn itself off and turn on another program*, one that produces another tissue type, written into another part of the DNA string. And so forth.

Thinking along these lines, the theorist Gregory Chaitin has defined what he calls *algorithmic complexity*. The basic formulae and formalisms of algorithmic complexity theory are identical to classical (i.e., Shannon-Weaver) information theory, but the interpretations of them differ somewhat. Rather than talk about probabilities and possibilities in the minds of observers (even including computers as observers), Chaitin is interested in the idea of the *smallest computer program* that could *produce* the phenomenon in question from scratch. Complex things take more instructions to manufacture, and the bit-length of the shortest possible "instruction book" (read: program or DNA string) that will do the trick of producing one complete instance of the thing—not how long it takes or how much energy it uses up—could be used to *define* the complexity of the thing manufactured. For an extended exposition of this idea, see his "To a Mathematical Definition of Life," (1970) and "Toward a Mathematical Definition of Life"(1979), collected in in his *Information, Randomness, Completeness: papers on algorithmic information theory* (Teaneck, New Jersey: World Scientific Publishing Co., 1987) 79 –104. In the second paper, he gives to "organization" the qualities we will give to a similar quanitity, "complexity-and-organization" in Chapter Two.

----

xxxiii Ayres, *Information, Entropy and Progress*, 115.

----

xxxiv See Robert E. Ricklefs, *The Economy of Nature*, 3rd ed. (New York: W. H. Freeman and Co., 1990) 112 – 116, and Paul Colinvaux, "The Efficiency of Life," from his *Why Big Fierce Animals are Rare* (New Jersey: Princeton University Press, 1978) 46.

----

xxxv See Notes 11, 14, and 26 above. The reader might want to review Appendix Two, where the relationship of "entropy" to "complexity" is more formally shown to be, for nearly all intents and purposes, identity. I continue to use "entropy" in these paragraphs because Brooks and Wiley do (as does Ayres and many others) and I am here

---

discussing their work.

[xxxvi] It follows that for Brooks and Wiley there is no such thing as "negentropy" or "negative entropy," a term popular in the 1950s and '60s. There is only the difference, or the lack of it, between maximum/potential and actual/observed entropy-complexity.

And whence the cosmos's increasing C_{max}? It is ultimately, I believe, the direct result and expression of the expansion and aging of the universe since the presumptive Big Bang.

[xxxvii] Rupert Sheldrake, in his book *A New Science of Life: The Hypothesis of Formative Causation* (New York: St. Martin's Press, 1987), claims that people can look at two pages of text, one being in a randomly generated pseudo-language with language-like spacing and punctuation, and the other in a foreign language a totally unknown to the subject, and know unerringly that the first is nonsense and the second is a real language. Ditto with spoken psuedo- and real langauge. One wonders whether subjects could also distinguish between semantic nonsense in the foreign language and a meaningful story or article in that same language.

Infants as young as 8 months old, listening to four tri-syllabic nonsense words strung smoothly together and spoken in random order for two minutes, are able to distinguish with close to 100% accuracy what the individual "words" are. Such is the statistical prowess of the brain. (Report on NPR's *Morning Edition*, 7.40am, December 13, 1996)

[xxxviii] The reader may recognize the monkey-and-typewriter parable here. To wit: it is possible, though very far from likely, that a monkey given enough time could type all the works of Shakespeare. How much time? William R. Bennet Jr. has calculated "that if a trillion monkeys were to type ten keys a second at random, it would take more than a trillion times as long as the universe has been in existence merely to produce the sentence 'To be or not to be: that is the question.'" (William R. Bennet Jr., *Scientific and Engineering Problem Solving with the Computer* [New Jersey: Prentice Hall, 1976], cited in Campbell,*Grammatical Man*, 116.) This reminds me of the story of the man who goes to a lecture on cosmology. At one point in the lecture he jumps up and cries out, "Excuse me for interrupting, but did you just say that the sun will burn out in 10 million years?" "No," says the lecturer, calmly, "I said '10 *billion* years'." "Oh, thank God!" says the man and sits down.

[xxxix] In the interests of clarity of exposition, I have not used the idea of "conditional probability." This is written in the form $p(i \mid j) = x$, and reads "the probability of event *i* happening *given* that event *j* has or will have happened, is *x*." Most events or field-states or symbols-in-a-string that unfold in time and that are the manifestations of natural or logical (if complex) processes present themselves to us as having probabilities that are "conditional" upon what we know has *already* happened. In reading a sentence, for example, our predictions of what word(s) will follow are constantly updated. It is as though reading (or intelligently following any time-extended phenomenon) were like driving at night with one's headlights on and usefully illuminating a finite width and distance of the road ahead.

A more basic example: in a long string of symbols each chosen at random from among 26 Roman letters, the two letters "X" and "E" have a certain probability of being found adjacent to one another:

$$\begin{aligned} p(\text{"X" and "E" adjacent}) &= p(\text{"X"})p(\text{"E"}) \\ &= 0.038 \times 0.038 \qquad (1/26 = 0.038) \\ &= 0.0014. \end{aligned}$$

But in English the letter "X" is less frequent than "E". Given the actual frequencies of their individual occurrence in the English language,

$$\begin{aligned} p(\text{"X" and "E" adjacent})_{\text{English}} &= p(\text{"X"})p(\text{"E"}) \\ &= 0.0013 \times 0.103 \\ &= 0.00013. \end{aligned}$$

(where the probabilities 0.0013 and 0.103 are derived empirically, by letter counts. This means that in English strings of letters the chances of finding "X"s and "E"s together are about 10 times less than finding them together in random strings of letters. But there is more structure in English yet. For the letter-*pair* "*EX*" occurs in English with far greater frequency that "XE", the ratio being somewhere around 13:1. A good way of expressing this is in the formalism of conditional probabilities, thus:

$$p(\text{"X"} \mid \text{"E"}) \approx 13\, p(\text{"E"} \mid \text{"X"}).$$

This predominance of one ordering over the other is precisely wherein the English language is providing new organization and positive information at the scale of (two-letter) phonemes or syllables. Given the differing usage frequencies of the letters of the alphabet, the space, and the period or comma, the actual complexity, C_{act}, of written English at the scale of individual letters is around 4.27 bits per letter.

Of course, with zero-organization strings such as sequences of coin flips, all conditional probabilites are constant and equal to the raw, unconditional probabilities. That is, with a tossed unbiased coin, p(h|t) = p(t|h) = p(h) = p(t) = 0.5

[xl] A mathematical constraint also ensures that $R_{\text{letters}} \leq R_{\text{syllables}} \leq R_{\text{words}} \leq R_{\text{phrases}} \leq R_{\text{sentences}} \ldots$, and that simultaneously that $C_{\text{etters}} \leq C_{\text{syllables}} \leq C_{\text{words}} \leq C_{\text{phrases}} \leq C_{\text{sentences}} \ldots$ This is because of the spatiotemporal embeddedness of one scale, or group-length, in the next-larger scale, the next-longer group, and vice versa: the larger *consisting of* the smaller.

[xli] Claude Shannon, "Prediction and Entropy of Printed English," *Bell System Technical Journal* 30, no. 1 (1951): 50–64. See also Norman Abramson, *Information Theory and Coding* (New York: McGraw Hill ,1963) 33 –38, and William R. Bennet Jr., who is cited and excerpted in Jeremy Campbell, *Grammatical Man*, p 116 –118. Perhaps no one has explained the hierarchical-with-respect-to-scale nature of evolving systems more elegantly than Herbert Simon in his classic paper "The Architecture of Complexity," in the *Proceedings of the American Philosophical Society* 106, no. 6 (December, 1962): 467–482.

[xlii] Information theory is itself a reading scheme, we realize; and so is our theory of value based on measuring "complexity" and "organization." With both ,*we* read nature *in a certain way*. Any method—any reading scheme, any computer program or operating system, any language, theory, interpretive system, or sense—*because* it is methodical, includes and omits, collects and discards, adds, subtracts, and averages...and thus imposes a certain amount of its own organization to the subject.

But "imposes" is not the right word. For where do reading schemes come from? They too evolve, along with their subject matter, and thus have no less legitimacy as a source of order in the world than that which they read. Indeed, because the reader and the read are part of the same world, we ought not to be surprised that DNA is a single string of code part of which "reads" the other. When it comes to the Book of Life, certainly at its smallest scales, the reader *is* the book and the book the reader. Some of this ambiguity survives at the cultural scale, i.e. with regular books, readers, critics, and *their* readers..., as poststructuralist philosophers of the 1980s and 90s liked to play up.

[xliii] Dan Sperber and Dierdre Wilson, *Relevance:Communication and Cognition* (Oxford: Blackwell, 1986).

[xliv] The square root of this product will be called Ω in the next chapter.

I would also venture this more radical hypothesis: that, like energy, *in any closed system (or universe as whole) the total amount of information in it is conserved*.

With maximum information content identified as entropy, this hypothesis flies in the face of the Second Law of Thermodynamics, even as I have elucidated it in this text, because it says that the entropy of a closed system increases over time to maximum (for its temperature). What I am proposing is not that "information really is different from entropy," or some such conceptual adjustment, but that entropy/information/organization *can move from one scale to another*, and thus seem to appear or disappear to the observer who is concentrated on viewing one scale only. In the canonical example of the ideal gas, upon which the Second Law is based, the internal energies and complexities-of-structure of the gas molecules are ignored, but this is where the randomness—the entropy—is coming from. These microworlds exist "outside" of the system as defined—not spatially, but in size. In this sense, information is like dust that can be swept under the rug but cannot be eliminated, or like a guest who can be ushered from room to room—scale to scale—but will not leave.

[xlv] Interestingly, while there is information in the tone or pitch of the speaking voice, there is no equivalent to musical chords in language semantically. Two or more different words spoken simultaneously do not combine to give rise a harmonious or dissonant "meaning chord." Lovers of operatic duets might want to disagree.

[xlvi] How exactly we judge "relevance" for music is an interesting question, but we do seem more quick to notice when music lacks tunefulness, integrity, consistency, or interest, than when speech or writing lacks the same. It seems to be part of our neural programming. (See the research of Sandra Trehub of the University of Toronto, reported online at *www.msnbc.com/news/433442* and at *http://cogweb.english.ucsb.edu/Abstracts/Tonality_96.html*). This is probably because language, going beyond the syntactic, has a semantic density and referential specificity that music lacks. With language we are focally aware of the organization of that-which-is-referred-*to* rather than of organization of that-which-refers.

[xlvii] In the nucleus of each living cell, copies of the organism's unique DNA-string lie bundled into *chromosomes* (together with the protein histone). Similar in structure to a fragment of a single string of DNA—but more like a fragment of (one side of) it—is another molecule which copies off segments of larger DNA molecule, namely, *messenger-RNA*. Messenger-RNA, as the name suggests, effects a continuous flow of "messages" from nuclear DNA to other parts of the cell, messages that are chiefly about the building of protein molecules that are required for other life processes.

[xlviii] A nucleotide is an A-T or G-C *pairing*, joined by a "tine" that is itself made of sugar and phosphate molecules.

[xlix] A tiny fraction of these are cancers: single-type cellular replication run amok.

[l] Example drawn from Robert E. Ricklefs, *The Economy of Nature*, p. 297.

[li] Research of the author, 1997. See also page 3 of Chapter Three.

[lii] Daniel R. Brooks, D. David Cumming, and Paul H. LeBlond, "Dollo's Law and the Second Law of Thermodynamics," in Bruce Weber et. al., *Entropy, Information and Evolution*, p. 206.

It should be noted that these authors use a way of calculating C_{max} (they have no C_{pot}) that makes it considerably higher than ours, as it adds in all possible N-tuples, including non-adjacent elements. This only becomes clear from studying the principle author's more extended exposition of this research in Daniel R. Brooks and E. O. Wiley, *Evolution as Entropy* (Univ. of Chicago Press, 1986; second edition 1988). Calculating C_{max} in this way tends to make R appear much higher for longer groups. Nevertheless, the point (i.e. that there is variation in R with scale) is well demonstrated.

[liii] See also Note 39 above.

[liv] First described by John von Neumann in the 1950s, "cellular automata" are mathematical entities: cells, in a field of cells, each of which change their state (or do not) in accordance with the state of the cells immediately around them. Receiving no instruction from a central controller, each cell rather "looks around" at its neighbors and decides what to be. Programmed appropriately in a computer, characteristic groups of such cells can be made to replicate themselves and multiply, forage and die, resulting in what many call "artificial life:" digital "creatures" of their own accord dancing on the computer screen, migrating, competing, succeeding and failing, as the case may be to be fruitful. Theorists of complexity employ artificial life (AL) cellular automata to test and demonstrate propositions, at an abstract level, about the evolution of life forms and their constrained behaviors over thousands of generations. We discuss AL systems again in Chapter Two.

[lv] See Brooks, Cumming, and LeBlond, "Dollo's Law...", and Brooks and Wiley, *Evolution as Entropy*, 182 ff.

[lvi] See Steven Levy, *Artificial Life: The Quest for a New Creation* (New York: Pantheon, 1992), and *The Whole Earth Review*, 76 (Fall 1992): 4–47.

[lvii] They do not account for the general decrease in the slope of C_{act} for N = 6, 7, 8 and 9 after the initial spurt except to say that it is a function of population size: a plateau is soon reached at which the actual variety of larger "genes" is very close to the system's potential if not the theoretical maximum, C_{max}, with which the authors calculate organization (which they call Information). It goes without saying that C_{max} (they do not distinguish between C_{max} and C_{pot}) increases quickly with group size.

[lviii] Brooks, Cumming and LeBlond, "Dollo's Law...", 199 and 201. A look at more recent work with cellular automata by Stuart Kaufman and Christopher Langton must await Chapter Three.

[lix] Figure adapted from Edward O. Wilson "Is Humanity Suicidal?" *The New York Times Magazine*, May 30, 1993, 26; illustration by Richard Downs.

[lx] And so we ought not to be surprised to learn that humans not only have the longest DNA sequences of all animals, but share most of their genes with other mammals, many with fish (the gene for hemoglobin, for example) and good few with insects, plants (such as the gene for photoreception), and bacteria. Our entire complement of genes, it is said, differs by less than one percent from those of chimpanzees. Only one percent? Clearly, from our anthropocentric point of view, this is a very *critical* (and, it seems, reading-scheme) one percent: the same one percent that has given us language, music, and, what is surely our supreme skill, getting into a car while closing an umbrella.

> It is clear in many respects that the distance between [man and chimpanzee] is very great indeed. Scarcely a bone in the body is the same...and the chimp's brain is smaller. [Because the protein-coding sequences in the DNA are so similar], (d)uring evolution, the changes which lead to such significant differences must have been not in those parts of the DNA text which specify the protein directly, but in the passages containing the algorithms which control the timing and use of existing structural genes [our "reading schemes"]… King and Wilson declared in 1975,

"Most important for the study of human evolution would be the demonstration of differences between apes and humans in the timing of gene expression during development of adaptively crucial organ systems such as the brain."

The following year, the biologist Emile Zuckerhandl reported that the brain of a human being is so similar to that of a chimpanzee in its basic chemistry that it may not contain a single protein with a genuinely novel function. "The uniqueness of the human brain must be due to structural and functional variants of pre-existing proteins...and to variations in the quantity, timing, place and coordination of the action and interaction of these proteins."

Mary-Claire King and Allan Wilson, as reported in Campbell, *Grammatical Man*, 137, 138.

Reading schemes again

[lxi] But see J. T. Bonner, *The Evolution of Culture in Animals* (Princeton, New Jersey: Princeton University Press, 1980). Bonner makes an excellent case that animals have culture too, and that such cultures are tightly interwoven with their genetic evolution.

[lxii] Richard Dawkins has famously called these ideas and practices *memes*, in allusion to genes. See his *The Selfish Gene* (New York: Oxford University Press, 1976). "Meme" is a term I use sparingly.
