The Designer Inference

Berkley E. Gryder

Introduction

Design: to create and shape for a planned purpose

My dad loved arrowheads. He used to take me and my two younger brothers to the mountains in South and North Carolina for entire days in the summertime to walk through tall grass in sloping meadows on a treasure hunt for Native American relics. We would spread out, dig through dirt, and leave no stone unexamined for marks of artistry and function: chisels in symmetric patterns, a sharp tip, sharp edges, and two indentations near the base of the rock for attaching it to a wooden shaft. Most of the objects we thought quality enough to warrant an "over here!" were mostly a mix of approximation and imagination, but every now and then (less than once per day) an unmistakably true arrowhead was discovered.

The ability to distinguish between rock and arrowhead is an intuitive inference. The "Design Inference." By inspection, we reach a conclusion about an object's origin as arising either from purely undirected physical forces (thermodynamics, physics, chemistry, etc.) or from a directed and purposeful use of physical forces. Inferring design helps to unlock the story of the designer. We look and wonder, "Why was it made? What is this for?" Objects created with intelligent intent tell stories: an arrowhead speaks to hunting, a cannonball speaks to war, jewelry speaks to culture, paintings capture the experiences of life and the movements of history. Together with written history, intelligently designed objects leave a clear mark where the unformed earth has been met by an intentional hand and given form. Order imprinted on chaos: creation. Putting together the human story from the innumerable acts of human ingenuity and creation gives the world color and meaning as we try to find our place in the grand epic of humanity. Archeologists recreate a story of past civilizations by studying artifacts they infer were shaped by human hands and ignoring dirt piled up from wind erosion. Crime scene investigators solve murders by figuring out what is a "clue" pointing to human action (a fingerprint) and what is "just there." The Search for Extraterrestrial Intelligence (SETI) has scanned the sky for alien intelligence from other planets by looking for patterns in electromagnetic radiation that couldn't be explained as just noise from the stars. When you look at your phone, you rightly infer a lot of engineering and intelligence went into making it; it did not arise by chance.

But what about life itself? When we look at the machine of the human body, or the computer code written on the DNA molecule inside each of our trillions of cells, do we see design and purpose? Or, as Richard Dawkins suggests, do we say "biology is the study of complicated things that give the appearance of having been designed for a purpose," that somehow arose by undirected physical forces? These questions are critical because if we arose by chance we will also disappear without ultimate meaning or eternal purpose – just an accidental blip on a lucky rock spinning in cold empty

space. If on the other hand we are the intentional product of an intelligent force, a masterful designer, a God, a super-engineer, then our lives and our biology have the potential for meaning that reaches out beyond our day-to-day slog against forces of decay.

I hope to illustrate that the tools of science point more and more to a Designer, and the shovel of biological research is digging up clues that do not demand either nihilism (the belief that life has no meaning) or pure naturalism (the belief that spiritual or supernatural forces and beings do not exist). I also aim, throughout this discourse of examining the evidence of life, to lift my own soul up to a place of wonder, recapturing the marvel and awe that wears out as I get older.

1. The Designer Inference

In the introduction, we looked briefly at the "Design Inference," a phrase coined in the 90's by mathematician William Dembski to describe the intuitive human ability to detect the difference between designed and random events, objects, and signals. Going one step further: because the designer infuses their own character qualities, we can learn much about the designer by studying attributes of the created object. This is the **designer inference**.

We will look at engineering to learn principles of design and designer inference. The designer inference is not limited to engineers of course, because incredible design goes into poetry, paintings, landscape architecture, sculpture, culinary explorations, choreography, fiction and non-fiction writing, composing music, etc. All of these we can investigate and conclude not only that they require intelligence, but we naturally form a picture of what the creator is like. Because at the molecular level the mechanics and computation inside living organisms more closely parallel human mechanical/chemical/computer engineering than (say) a painting, I will focus there.

The Engineer

To help with our inference of the designer, let's think about what it means to "engineer" something. It means to sit down with a goal in mind (abstract, perhaps vague – like "I want to build a machine to allow me to fly"), draw up blueprints, build a prototype, test the prototype to see how it works, learn the rules and turn them into formulas that describe the real world (the science and physics of aerodynamics), then go back to the blueprint and make corrections

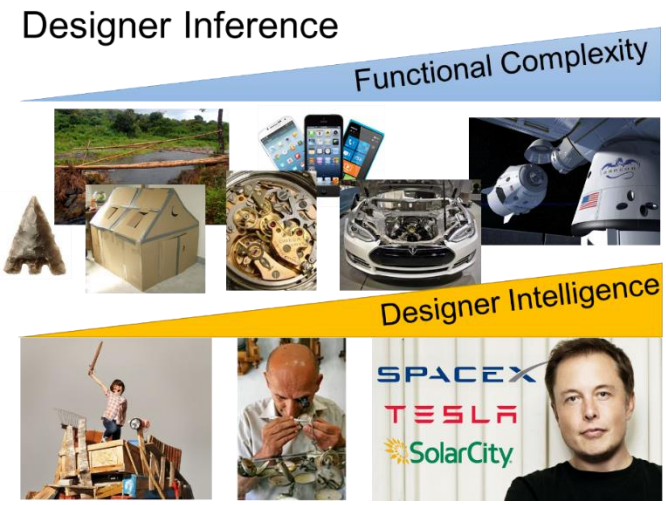

Figure 1. The Designer Inference, where the functional complexity is used to infer the relative intelligence of the designer.

before building a new prototype. The more knowledgeable and experienced the engineer, the more times they have gone through this engineering cycle, and by their gathered knowledge of the true representations of the way the real world works can make objects of ever-increasing complexity working towards the envisioned goal. By studying products, we can learn about the relative level of intelligence, perhaps something about the surrounding cultural influences the engineer cares about, the financial resources available to the engineer, or even their color preferences. In cases where an engineering product has multiple aims to simultaneously satisfy (not only "it must fly" but also "it must be safe" or "it must be fast" or "it must consume no fuel" or "it must look sleek and elegant"), then by comparing different designs you can infer the different priorities that the engineer cares about, and what people care about gives you a window into who they are and what they are like.

Grade Adjustment

Because knowledge can be codified by language, life-times of lessons learned can "jumpstart" a young engineer and save them centuries of trial and error. This impacts designer inference by adjusting for the historical moment. At any single given point in history, because passed-down knowledge is largely shared, in order to rightly infer the skill and intelligence of an engineer you also have to normalize by (or, subtract out) the contribution of the current shared knowledge pool. Thus, if Einstein came along today and rediscovered all the same things, he would not be revered as a genius and would hear "this is already known." Conversely, this is why we can marvel at scientist-engineers such as Michael Faraday who made fundamental and profound discoveries linking electricity and magnetism to engineer an object like the dynamo (that converts mechanical action into electric power by a rotating magnet), precisely because at the time he was truly shaping knowledge out of the unknown. Engineering of the Pyramids at Giza or the Colosseum in Rome are wondered at for similar reasons. True genius can be recognized in the work of anyone who appears to be "ahead of their times." The same goes for age considerations – a 4-year-old who builds a function-rich cardboard fort may receive as much admiration as a 40-year-old who builds a function-rich tree house for that 4-year-old.

Minimal Intelligence

The designer inference is also only capable of discovering the "minimum intelligence" of the designer, not necessarily their full capacity or brilliance. Elon Musk could engineer and build a cardboard fort, a tree-house or a space station. Even if he made all three, the fort shows one minimum intelligence requirement, the fort yet a higher minimum, but the space station shows the greatest amount of minimum intelligence. We may be smarter than one of our creations can reveal, but not dumber.

Functional Density

So, what metrics can we use to ascertain a high-quality engineer who deserves a raise and more responsibility, versus a low-quality engineer who should be given more time, training and careful oversight? Or, an engineering company worth investing in, versus one we should avoid? History of engineered projects can give us insight, as engineering quality appears to inevitably increase over time. The metrics one might look for include ***efficiency*** (ratio of input operational cost to output function, i.e. miles per gallon), ***elegance*** (clever streamlining rather than ad hoc or redundant parts), ***operational durability*** (how many years does it take for a microwave to stop working?), ***functionality*** (how many different things can it do, and in how many different conditions?), and ***size*** (equivalent function with smaller size require more detailed, precise and difficult engineering – think of a jukebox versus an iPod nano). When either functionality increases or size decreases, the ratio (function/size) increases, growing in

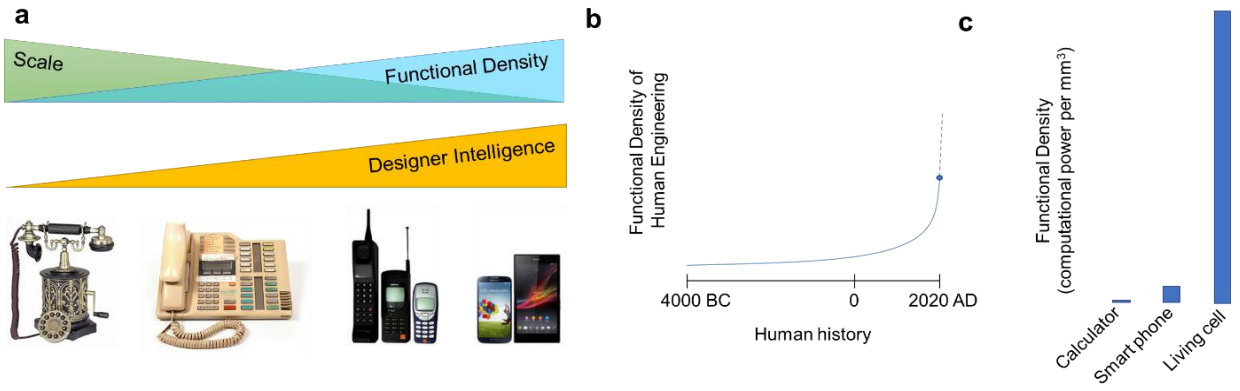

Figure 2. Functional density. **A**. Increased functional density as a decrease in the size of the object capable of carrying out the same function. **B**. Functional density of human engineering, skyrocketing in the modern information age. **C**. Functional density as a consideration of computational power and size.

functional density. When ***both*** function increases and size decreases, an even greater improvement in functional density has occurred – this is the success sought after by many modern engineering companies.

Let's look at one great example of **increasing functional density**, the telephone. Before electricity-based wires, telephones began as two tin cans connected by wire conducting mechanical vibrations from speaker to listener (known as the "lover's telephone"). In the late 1800s, the first primitive attempts at an electric based audio communication device were simple, expensive, and lacking basic features (such as a "ringing" feature to inform you of someone on the other end). Improvements and patent wars led to rapid improvements throughout the turn of the century, and in the 1930's phone makers made a major leap by adding a number rotor clicking in a circular motion. Phones stayed largely unchanged until rotors were replaced with touch-tone dialing technology in the 1960's, but performed largely the same functions. The first cellular

telephone (cell phone) was demonstrated in 1973, but early versions called “bag phones” were quite cumbersome. Handheld mobile phones in the 1980s were large, military style devices with poor signal coverage and no additional features. Yet they were symbols of wealth and financial acumen in the dawn of a new era of technological innovation that began bringing primitive computing technology from the labs at NASA into the home. Each year of the 1990s saw cell phones steadily diminish in both size and price, fostering rapid proliferation. Functional density increased because of size reduction, but not necessarily because of feature addition. But, once size and weight reduction reached its optimum in the 1990s, the only available next leaps in functional density in the 2000s were addition of more and more functionality. The first ubiquitous new function given to mobile phones was texting, consuming the function and existence of beepers. Then phones added clocks, alarms and timers, largely reducing the wrist watch to a functionally redundant fashion accessory. Functional density increased with backlit dial pads, screens, color screens, eventually touch screens. Cell phones broke the yoke of costly long-distance calls, and networks expanded coverage beyond big cities. Now they are breaking the yoke of international calling costs, via internet based digital voice packet exchanges like Google hangouts and Skype. Perhaps the largest single jump in functional density was the first iPhone – giving you a music library, the internet, document storage, and an endless swarm of applications (apps) whose marketplace spawned a new profession, all in a single device. Functional density improvements since then have inched forward incrementally – but persistently marching toward higher resolution screens, more memory, faster processors, better speakers, more powerful cameras, face recognition, and much more.

Imagine an outsider or alien studying the history of this technological revolution. Applying the designer inference, they could surmise that each increase in functional density not only bears the marks of a remarkable engineer leading a talented team, but also surmise the collective intelligence and design capabilities of humanity as a whole are increasing. If we then apply the functional density test to biology, it suggests a super intelligence. The super computer of the cell has far greater functional density than anything the full intellect and collective knowledge of humanity could possibly produce.

God of the Gaps

As science and human learning progress, superstition and wishful thinking are pushed out of bounds. Everyone benefits. We can interact with the world more fully by understanding more accurately. Moving deeper into human history we see a larger amount of phenomena being explained as “actions of the gods”, or “spirits”, and these gods and spirits had to be appeased. Thus, as Zeus (or many others to whom lightning bolts has been attributed) was replaced with a theorem of electric current, and as we understand charge separation and discharge, we worry more about water and metal fences than whether or not we’ve offended the Bolt Thrower in the Sky. Thus, the “Gap” (where an occurrence can’t be explained by science) has often been filled in by a supernatural force or entity (often God). Any arguments for God that are based on Gaps (we can’t explain X, therefore God must have done it) are criticized as spurious. This leads to the mocking claim that believers in any religion use their deity as an excuse to cease the hard work of trying to figure out “how things really work.” For any God of the Gaps, our towering accumulation of knowledge has squeezed Him into a very small Gap indeed. We’ve plunged the depths of the atom and pulled out quantum mechanics, and reached into the stars and pulled out relativity. Yet, there is one field of study where the Gap gets wider and deeper the more we delve into it: Biology. The more we learn of the inner life of the cell (not to mention the rich communications and interactions between cells in a multicellular organism), the more we unravel new layers of complexity, information, sophistication, and **functional density**. This drives the demand for a designer up to an astonishing level, namely, an entity with a super intellect with engineering capabilities far above anything humanity can create. This also sinks the possibility of random chance and undirected natural forces to explain its existence. Any one of your cells has more functional density than your smartphone. Only God can fill that Gap.

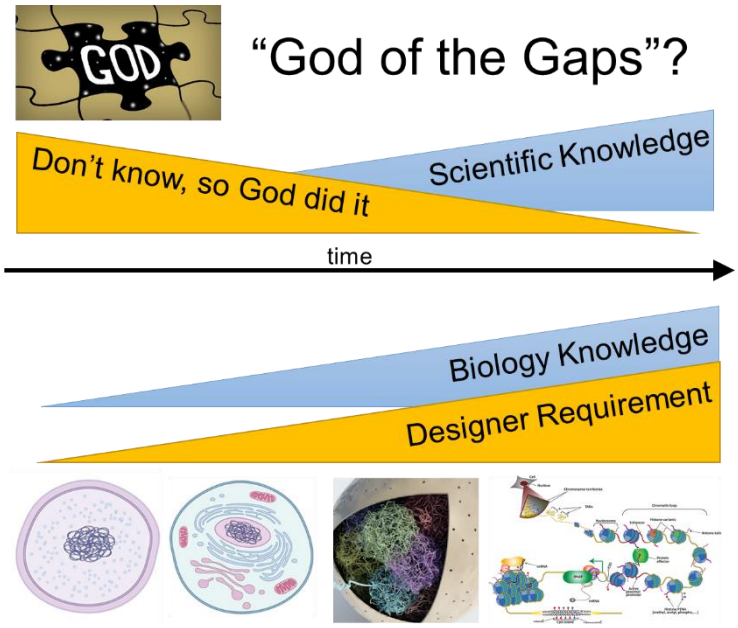

Figure 3. The God of the Gaps. Above, the observation that, over time, an increase in understanding of direct scientific causality supplants the frequency of attributing direct causality to a Divine Agent. Below, the reversal of the trend, as the increased scientific details of biology reveal information systems with greater functional density and an ever-increasing requirement for a designer.

2. Biological Computers

If no natural cause or combination of chemical reactions can produce computer code, we aren't just calling God into the gap when we see computer code written in the cell. It isn't an argument from ignorance, but from knowledge. We ***know*** what causes computer code - a mind! All abstractable and semiotic (sign based) information sets (novels, speeches, software) come from intelligent minds. Our repeated experience confirms that. We do not find a single exception to this in all of scientific history.

So far, human attempts to create biological machines capable of calculation are far more rudimentary than computers humans have made with metal circuitry. Yet, biology itself is already performing computational tasks of exceeding complexity, and with

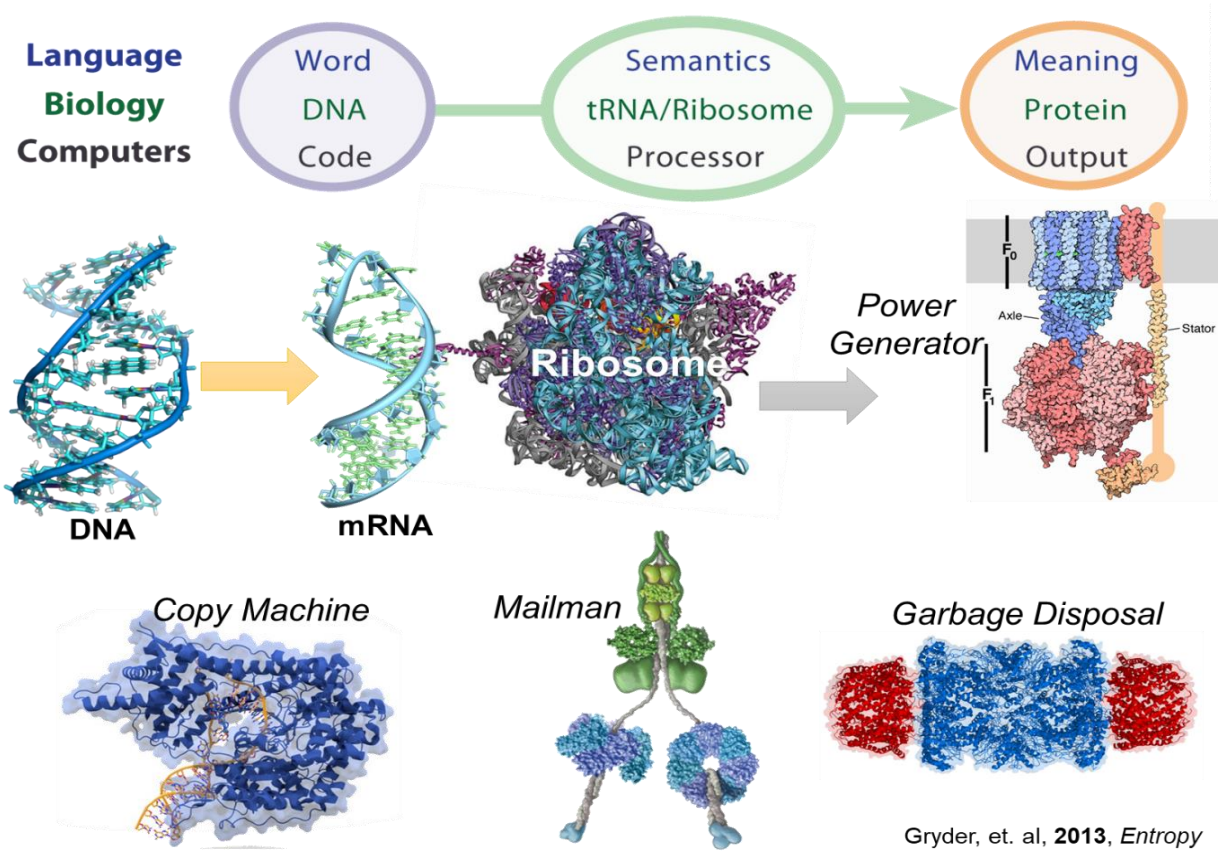

Figure 4. Semiotics, the study of sign systems. Just as in language and computer technology, biology is rooted in a 3-part semiotic relationship between the DNA, RNA and resultant proteins. This coded language of life has hundreds of thousands of meaningful specifications for nanomachines with complex functionality, such as copying machines,

remarkable accuracy. Here I will undertake a survey of the components involved in the computation of living cells, with a focus on the super-computer of the nucleus. It is a machine that, just like a computer, operates to store, manage, interpret and process *coded information*.

Information and Semiotics

Why do stories and computer code require a mind? Because they come from a higher plane of existence than mere molecules and energy. They originate in the abstract world of information and representation: thought. The world of ideas, ideals, purposes, plans. To help us consider biological computing in living organisms, we can learn by comparison to language and computer code. All three systems form an information relay that communicates meaning, and performs function. The formal study of the information exchange in these systems calls is "**semiotics,**" the study of sign systems. These systems all involve three essential parts: they all have a set of "signs/symbols," a "decoder/interpreter" and a set of "meanings." Each symbol is mapped to at least one meaning. A system of many signs with distinct meaning, all coherently passing through the same decoding device, form the basis of complex communication. A semiotic system allows meaning to be stored as signs (a code, a word, a string of DNA bases). Storage of a real object (a rose) or an experience (a smell) or an idea (the arbitrariness of names) into a physical medium (a written word) allows for storytelling (*Romeo and Juliet*). A feature of coded information is that the signs themselves are arbitrary (a rose by any other name would smell as sweet). Another such feature is that although words and code are arbitrary, they are generally not very flexible (in order for language to accomplish the goal of communicating one particular meaning, and not another). Yet another key feature of symbolic information is that it can jump from medium to medium (thoughts can form spoken words, which can then be written with pen and paper, then captured by a photograph, transformed into electrical pulses streaming through a smartphones processor, then transformed into radio waves of light, and ultimately pass through the eyes and into the mind of your friend an ocean away). Code, language and biological code also all heavily depend on **context** to inform meaning.

The three semiotic spheres (of human language, computer code and biology) are distinct and yet parallel, such that comparisons between them is mutually illuminating. All three systems have different physical entities for their signs: **language** is codified into spoken words, written alphabets, hand motions, or drawn symbols; **computer code** is codified into bits, 0's and 1's, which are assembled into higher level code languages that are "human readable," **biological information** is codified into chemical strings made of the base pairs A, T, C and G. Languages are *developed* through social interactions and history. Computer codes are *invented* to achieve technological goals. Biological information is *discovered* by molecular-level investigation with the rigors of scientific observation.

Both language and computer code are markers of intelligence, and although they evolve, the evolution is in the abstract world of ideas through mental connectivity. Cars can transition from simple to more complex, but the evolution is not being passed on from car to car, but from engineer to engineer.

The Genetic Code

There are, in fact, many levels of information exchange in biology. Your brain sends signals, like the signals needed to wave your hand, and endocrine glands send signals via hormones in your blood to regulate all manner of biological activities. Proteins on the surface of a cell ("receptors") can receive information communicated by all shapes and sizes of molecules that bind to them specifically: different receptors exist that can recognize single-atom ions like calcium, multi-atom small molecules like caffeine, or larger strings of molecules (called proteins) like insulin. Inside the cell, countless molecular binding events and tightly controlled chemical reactions are all happening in a fluid orchestral dance, and where these events are traced in reproducible lines of succession, we term them "signaling cascades" or "pathways" that are relaying meaningful chemical information.

Yet, underneath all these layers of biological communication, there is a "central dogma" that gives life its ability to exist from generation to generation: the genetic code. Each of the hundreds of thousands of biochemical machines called proteins are encoded in the twenty-thousand genes (from the Greek word *genos*, meaning birth). For life to be life, it must make copies of itself, to be fruitful and multiply. Yet, protein machines that carry out the functions of life (for instance, enzymes that chop up your lunch into the small chemical nutrients that are the building blocks) are not able to copy themselves and ***cannot be copied*** by others. Instead, all proteins are "encoded" into the alphabet of DNA: this is important because DNA is a molecule that although it cannot copy itself ***it can be copied*** by a dazzling ensemble of proteins that make up the "replication machinery."

The DNA molecule itself is chemically suitable for copying, because it is a "double helix": the DNA molecule is double stranded, like two sides of a zipper, but linked together by the chemical complementarity of each base pair. Each "A" molecule fits like a jigsaw puzzle piece only with "T" on the opposite side of the zipper, each "C" pairs with a "G." DNA comes in long strings (many millions of base pairs) of code, where one side is both a "reverse" and a "complement": a code reading "ATGT" on one side would read "ACAT" on the other side. To be meaningful, these chemical letters are not arranged randomly. The parts of DNA that can make proteins are built with a long series of 3-letter words called "codons," where each codon represents a particular amino acid (of 20 possible). The four base pairs (also called nucleotides) are strung together in careful code-like order, such that a specified series of DNA codons is interpreted into a string of amino acids (the individual molecules) that make up proteins. There is very little chemical difference between base pairs, but amino acids have very different chemical properties (some negative, some positive, some neutral, some bent, some oily, some watery, some small and others large). This allows for a simple chemical alphabet to form a complicated 3D mechanical language.

The DNA isn't interpreted directly into proteins, however. An individual gene in DNA is copied into an RNA molecule, by the process of transcription. A single chromosome

may have ~1000 protein-coding genes on it, of which about half may be “active” and have RNA transcripts. This creates a pool of RNA molecules (the “transcriptome”) that is 8 times more abundant than the DNA molecule itself. The DNA is the hard-drive, sturdy and double stranded, coming in 23 different long strings of information (we have 2 copies of each of the 23 chromosomes, averaging 140 million nucleotides each). The RNA is the random access memory (RAM), is unstable and single stranded, and has 2000 nucleotides on average.

RNA molecules that code for protein carry a semiotic “message,” and are referred to as mRNA, which are decoded (translated) from nucleotide sequence to amino acid sequence by a machine called the ribosome. The name mRNA helps to discriminate it from the many other types of RNA in the cell that are used for other functions: ribosomal RNA (rRNA) is the most abundant, and forms a 3D structure interwoven with ribosomal proteins to form the completed ribosome. Transfer RNA (tRNA) is also quite abundant, as it is the molecule that has an anti-codon (decoder) on one side and a specific single amino acid on the other. One tRNA is used up in the ribosome for every amino acid in every protein ever made. There are many other types of RNA as well.

Self-referential systems require forethought and design

The central dogma--DNA to RNA to Protein--is itself a self-referential system. In other words, each of these three parts has no meaning without the other parts. The DNA molecule is only a “code” filled with useful information because of the existence of the mRNAs, rRNAs, tRNAs, and hundreds of other proteins required for its interpretation. Yet, these RNAs and proteins are all coded in the DNA itself. So, which component could have come first? None. All three simultaneously had to have been put in place together. For this primary reason, the origin of life from non-living chemical matter remains completely devoid of a plausible naturalistic explanation: no naturalistic process can create semiotic systems.

Self-reference, or circular interdependence, is a feature of semiotic systems more abstractly, although they aren’t all self-propagating the way biology is. Computer code can only have meaning if it can be interpreted by other computer code, and the dictionary is only useful if you can already read a bit. Unlike human language, the connections and inter-relations in biology and computers are “built in from the beginning,” and the information ecosystems within which they operate are not able to rewrite their semiotic conventions. In this way, biology is closer to computers than to human language: they are more fixed and rigid. Self-referential systems cannot “evolve,” as all the parts must be placed together at the start in order for any individual part to be functional. Systems of abstract representation with symbols are created: these are the domains of logic, mathematics, computing and language, all of which originate as designs in the mind.

The forces of chemistry and physics do not create codes of any kind on their own. In chemistry, the fundamental principle is that reactions move until they reach the lowest

energy state. In physics, the states of matter and energy tend only to decrease free energy and increase in chaos (entropy). These forces can create types of order (the orb-like shape of a water bead arises from minimized surface tension; the repeated 3D order of atoms arranged in a lattice creates a crystal), but not *systems of abstract symbolism*. The arbitrariness of semiotics is itself a function of not having chemical constraints. This breaks down a little bit when considering the chemistry of life, as we are not smart enough to imagine another kind of coded chemistry, so it seems non-arbitrary. Nevertheless, the protein code is arbitrary, and could have been designed in other ways (4 bases per amino acid, or the 3 base-pair codon for one amino acid could just as easily have been assigned to a different one) to achieve the same proteins.

Information is itself an immaterial thing: you can erase your laptop's hard drive without reducing its physical weight or energy level. Non-living things consist only of energy, matter, time and space; living things have information as one more orthogonal parameter. The protein coding world in a living cell is a remarkable amount of ordered molecular function that goes against the laws of disorder and chaos that run the cosmos. Humans create order out of chaos using their minds to enact artistry and design into the world; in the same way, a Designer with capabilities well beyond any human must have built the technologically superior and artistically stunning world of biological life.

The Myth of Junk DNA

In the 1980s and 1990s it became clear that the portion of the genome that codes for protein (mRNA) was quite small. The actual number was in debate until the Human Genome Project neared completion, leaving us in amazement that only ~2% of our 3.2 billion base pairs code for protein. Even before the final numbers were tabulated, the existence of non-coding DNA became a favorite argument against the view that biology was designed. Richard Dawkins notes that the genome "consists of multiple copies of junk, 'tandem repeats,' and other nonsense which may be useful for forensic detectives but which doesn't seem to be used in the body itself. Once again, creationists might spend some earnest time speculating on why the Creator should bother to litter genomes with untranslated … junk tandem repeat DNA" ("The Information Challenge" *The Skeptic.* 18,4. Autumn 1998). "Junk DNA" was considered as garbage left over from the random and blind mutational process proposed in neo-Darwinian theory. Or again, in a book attempting to showcase evidence for an undirected process in the formation of biology: "it is a remarkable fact that the greater part (…95%) of the genome might as well not be there, for all the difference it makes." (Richard Dawkins, *The Greatest Show on Earth: The Evidence for Evolution*, 2009). At the same time, scientists who viewed life as the result of intentional and careful design assumed the "junk DNA" wasn't junk at all, just that we hadn't figured out what it was doing. William Dembski writes, "…on an evolutionary view we expect a lot of useless DNA. If, on the other hand, organisms are designed, we expect DNA, as much as possible, to exhibit

function… Design encourages scientists to look for function" ("Science and Design," *First Things*, 1998).

As science has progressed in the era of genomic data (especially since 2010, as new technologies developed), the prediction from Design came true, and junk DNA was abandoned as a viable concept, and this resulted in an earthquake in the world of biological science. The functions of junk are now so numerous it is hard to keep track. A watershed moment in the history of science is summarized nicely in an article titled "ENCODE Project Writes Eulogy for Junk DNA" (Elizabeth Pennisi, *Science*, 07 SEP 2012 : 1159-1161). The ENCODE project, and its results, shine a bright light into the "dark matter" of the human genome, and reveal a level of sophistication and function that boggles the mind. In my own research, I used the insights and algorithms from ENCODE to map the epigenetic state of childhood cancer epigenomes, shedding light on how "junk" DNA is a critical driver of circuits that cause these life taking diseases (Gryder et al, *Cancer Discovery* 2017, Gryder et al *Nature Genetics* 2019).

The amount of functional density in the genome is constantly growing, as we discover more and more functions in the junk. To date, most of these functions can be divided broadly into three major categories: (1) layers of "non-protein-coding" code devoted to the control of RNA transcription, (2) non-coding DNA that nevertheless makes RNA molecules that are a functional end unto themselves, (3) DNA sequences with structural and mechanical roles needed for its non-transcription-based functions, such as DNA replication, folding, and maintenance. As my personal scientific pursuits center around the discovery of mechanisms involved in the control of RNA transcription, I will focus our discussion on the three "codes" discovered in the junk DNA regulating gene expression: the **gene body code**, the **enhancer code** and the **origami code**.

Codes Controlling the Code

DNA is code. The base-pair sequences code ultimately for protein nanomachines, and they do this by instructing the sequence of amino acids; this sequence then instructs the folding and structure of the protein, giving each of these nanomachines a purpose and role to play in the metropolis of the cell. The process of transcription itself is elegantly controlled. In the last few years, new technology has propelled the genomic sciences at an ever accelerating rate. We are now beginning to discover and recognize that the DNA is multi-layered with codes on top of codes, all of which work in concert to govern the timing and magnitude of RNA expression. Functional density increases the more functions we find in the same string of DNA letters. The elegance of a computer code ecosystem is a marker of intelligence, experience and skill. The genome bears all the marks of elegance, most especially when considering it has multiple overlapping codes: a remarkable feature that allows the designer to pack even more into a very small space.

The Gene Body Code

The first “code on top of the code” is the Gene Body Code. The Gene Body Code was the first to be discovered, yet our appreciation of its complexity is still increasing. It has to do with the non-coding DNA that immediately surrounds the genes, and is thus much “easier to find” than the Enhancer Code (which works over quite long distances) and the Origami Code (the code for DNA folding at larger scales).

Each gene (a region of the genome that gets turned into mRNA) has some sequence that itself is a kind of “code” that talks to the machines that make RNA (called RNA Polymerases) and tells it to “start transcribing here!” or “stop transcribing, you’re done!”. Each gene has a “promoter” and “transcriptional start site” code at the beginning, which has sequences that are interpreted as a “bind here” message that gets ignored by many proteins but is listened to by the proteins that are supposed to start transcription. There is often a TATA box recognized by the TATA-binding protein and many other general **transcription factors** (TFs) needed to get RNA Polymerase to set up.

Another feature of the Gene Body is the “splicing code”. Long sections of the DNA copied into RNA do not correspond to the amino acid code; these are **introns**. The parts between introns that have meaningful corresponding “protein coding potential” are called **exons**. One protein might be encoded by a single exon, or it could be made up from 9 exons that had introns spaced between them. Those introns have to be “spliced out” before the RNA can be decoded into a protein. How does the cell know which sequences need to be spliced out, while others are kept in the RNA molecule? It turns out there is a “code” for that as well: a special sequence where an exon stops and an intron begins! As the cell is copying the RNA from the DNA template, it forms a lasso out of the intron, then cuts out the lasso and glues together the two adjacent exons. What makes this mechanism so special is that it allows for a single set of exons (a single gene) can be alternatively spliced, allowing for a set of exons to combine in many different ways, resulting in many “splice forms”, resulting in a variety of protein isoforms. We don’t fully understand how different isoforms are selected for, or explain it when a cell choses to pick exons 1, 2, 3, 4, 5, 7, 8 and leave out exon 6. But we watch and record it happening. A single gene can become 2 or 20 different protein isoforms, which is how our only 20,000 genes give rise to 200,000 different transcripts! This modular recombination increases the functional density of life even more, and resembles a clever engineering trick to maximize complexity and diversity.

The Enhancer Code

How can the same DNA information create so many different kinds of cells, tissue types and organs? It turns out that each cell type has a custom setup that uses a specific set of genes (while each cell also actively prevents most genes in the DNA from coming on at all). The cell is like a computer in this way: the same computer in a university library gets used by hundreds of different students, and each of these students can do very different tasks by turning on different sets of software programs. Some may turn on

music and write an English paper, while another is relaxing and watching a movie, while an architecture student uses AutoCAD to design the finishing touches on a four-story hospital wing. In similar fashion, our cells turn on different sets of programs to do different actions: beta cells produce insulin for your blood stream, nerve cells relay electrical signals from your fingertips to your brain, muscle cells in your eyelid constrict and relax to blink. Each of these cells have the same DNA, but each of them has a very distinct set of RNA molecules that code for proteins used to perform these diverse functions. How? Well, surrounding genes needed to make a muscle cell are DNA stretches with sequences recognized as binding sites for muscle-making transcription factors proteins. When these sequences are recognized by the transcription factors, they open up the DNA in these regions. Why isn't all DNA already open? Well, each multi-million letter string of DNA is not naked: it is wrapped around histone proteins, like beads on a string. These beads get two wraps of DNA each, and they bind tightly, preventing access to the information inside DNA as a default. Pioneer transcription factors can push these histone beads out of the way, creating an opening for regular transcription factors to bind their target DNA sequence. To keep these sites open, and thus active, other factors are recruited which can chemically modify the histone beads and make them greasy, so that they can be pushed around more easily. These sites are called **enhancers**, and help activate nearby cell-type specific genes. Enhancers can be nearby or hundreds of thousands of DNA bases away from their target gene; to work, these enhancers form loops over long distances to connect with their target.

The Enhancer Code still contains many poorly understood mysteries, making it exciting to study. But, we do know that enhancers are cell type specific, just as the genes they activate are cell type specific. How are some enhancers on and others off? That is in the Enhancer Code. For an enhancer to become active, it requires two things: (1) a combination lock, or "enhanceosome": a dense packing of many transcription factor binding site sequences all next to each other, and (2) the combination keys: large quantities of the right transcription factor proteins available in the right combination. A small number (~10 or 20) of "master" transcription factors exist in each cell type that run the show, while most of the 1,500 transcription factors are barely on, if at all. The masters also auto-activate themselves (each *gene that encodes a master transcription factor* has enhancer sequences that are recognized by the protein product of that gene). This creates a biochemical memory. This also forms the biochemical basis of "transcriptional computation" and cell state switches. This is how cells remember who they are. This is where signals telling cells to become a new type of cell are registered. That is one wonderful achievement of the Enhancer Code, which unfolds elegantly and precisely during the development of a baby from a single cell.

The Origami Code

In 1999 the human genome was sequenced near completion, and in 2009, the first genome wide maps of 3D genome folding were published. This new data painted an elegant picture of a highly organized 3D structure. Some of the most exciting

brainstorming discussions of my life happened on the heels of that work: I and several other like-minded biologists and computer engineers began to wonder if the secrets to the data-management of the genome were hidden in the 3D structure. The maps were initially very blurry, but higher resolution and clearer structure was achieved 4 years later in 2013, revealing a new kind of code that controlled 3D looping of DNA. I call this the Origami Code.

The Origami Code started to arise after a key discovery: the first 1,000 base pair resolution maps revealed clear loop anchors, where two spots on the DNA were interacting at very high frequency. Imagine you find a shoestring with blue marker coloring at one place, then red marker coloring at another spot, and folded so that those two spots touch with a rubber band holding them together. Now, imaging there is a sequence of letters along the shoestring, and imagine you find hundreds of such shoestrings with one loop each. Now, you notice the letters at the blue spot spell the words “left stop” under each blue spot, and then quickly look under each red dot and find they spell the words “right stop”. Well, with all the DNA loops, we looked at the loop anchors and found the code phrase “CCGCGNGGNGGCAG” come up over and over again! It turns out, there is a protein called CTCF that reads the genome until it finds the phrase “CCGCGNGGNGGCAG”, and sits down there and binds so strongly that it won’t get back up. How could it form loops, though? Well, it turns out that there is another protein that works like a rubber band into which the DNA string is inserted. The rubber band is called cohesin. Cohesin also has a motor built into the rubber band, and extrudes the DNA through the middle of it. Then, when cohesin runs into a “CCGCGNGGNGGCAG” site where the CTCF protein has latched on, it stops extruding the DNA. But, that is just on one side of the DNA; the other side it will keep pulling through until it reaches a second, anti-parallel CTCF stop sign site. One stop sign on the left side, one stop sign on the right side, and one big DNA loop in the middle. This amazing process happens in all of our cells constantly. It prevents DNA from tangling up like spaghetti, and creates little neighborhoods and cul-de-sacs where the right enhancers can talk with the right genes. That is just the surface of what the Origami Code is doing, but it is a marvel to study.

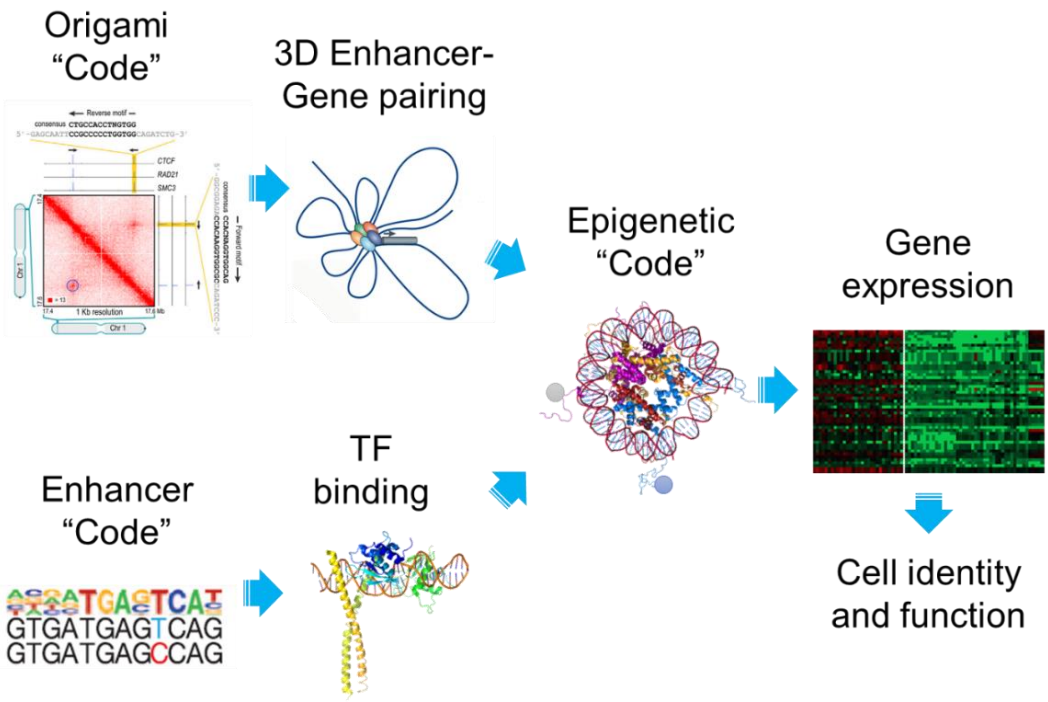

Figure 5. Multiple overlapping codes beyond the central dogma that control gene expression patterns and thus cellular identity and function. TF = Transcription Factor, a sequence-recognition protein that activates a region based on non-coding sequences.

In 3.2 billion base pairs of genetic information, how does the right data get accessed efficiently? Well, the Origami Code paired with the Enhancer Code can help ensure the genes inside are more easily found. This is similar to a data-speed trick that is used in computers: we give files a data-address so that when looking for the file the computer doesn't have to look everywhere. Reducing the number of places you have to search greatly increases speed. Many more of the tricks and clever engineering tools that we invent for solving problems in the world of computing may provide inspiration for us to look for similar strategies in the human genome.

Biological Information and Entropy

Now, how can all this elegant information, and the systems to control it, arise? Can we really believe that random forces of mutation and natural selection can make a super computer? Many people and many scientists still imagine that all this organization and super-high functional density is not the result of design. But, even so the marks of design are recognizable by those who do not think it was designed: “Biology is the study of complicated things that give the appearance of having been designed for a purpose”, remarks Richard Dawkins in *The Blind Watchmaker*.

The debate of how information got inside the genomes of living creatures revolves around the role of mutations in genetic information.

What happens when genetic information gets eroded? We grow old, wear out, and get cancer. One of the ways we figure out what a biomolecule does is to study what happens when it breaks. Usually that means comparing a disease-causing mutation to a healthy, non-mutated control. Now, just like in human language, some mutations are worse than others at reducing the amount of useful information.

Owh ucm negantim an ory tact ofrm ish tencesen?

Hwo mcuh meaiinng cna yuo exatrct form tihs setnence?

In biology, we infer that sites where almost no variation exists in the human population are sites that are most essential for function. How much meaning can you extract from this sentence? Some mutations that completely erode the meaning and function of a critical protein are simply never observed. This is for the same reason that airplanes returning to military bases in World War II all had bullet holes accumulating in the same few places, and zero bullet holes ever showing up in certain places: the zero-bullet hole components weren’t harder to hit. They just resulted in a crash and those airplanes never returned home. Natural selection works this way, as a purifying filter that weeds out the very worst mutants that cripple the cell. Approximately 2,000 genes are essential to every one of our cell types, while at the whole-organism level many more genes are essential for successful embryogenesis (development of lungs, heartbeat, the central nervous system, the digestive system, the circulatory system, and the skeleto-muscular system) and a healthy birth. Genetic diseases that manifest at birth or in early childhood development are the result of bad genetic mutations, but these mutants are not as lethal as the mutations we never see. Most mutations that happen are “slightly deleterious and near neutral”, meaning that they may add up to a negative fitness, but are not so acutely damaging that Natural Selection can prevent them from building up in a population.

If mutations are causing information loss and reduction in species fitness at a rapid pace, how can a mechanism like that create new biological computer code? It is only imagined that positive, beneficial mutations would come to the rescue. But, do we see this in the data? No. It remains a topic of inference and assumption, not observation.

The comic characters in the Marvel X-men, where superhuman powers arise from random mutation, remains squarely in the realm of science fiction. Not observed, not science. A full review of the scientific literature surrounding population genetics reveals that mutations accumulate in a downward spiral toward extinction, not upward spiral toward better new types of species. I refer the more interested reader to find a full treatment of this topic by geneticist Dr. John Sanford in his book *Genetic Entropy*.

Purpose Inference

What is an object made for? That is the practice of teleology, the intuitive activity whereby we discern "why" something was made by studying what it does or what it is functionally capable of. The telos of a knife is to cut, and so on. Thus, in addition to the Designer Inference offering insight in to the sophistication and capability of the designer, it may also be useful to infer the Designer's purposes. The genome bears the marks of design is because it was designed for a purpose. Life itself, in all its forms, sculpts beings with many purposes. If it were designed by a creator for a purpose, what would some of those purposes be? And how many of those purposes might we learn from studying the design?

Purpose forms at every level: the purpose of the genome is to bring life to a cell, the purpose of the cell is to bring life to an organ; the purpose of organs is to work together for a symphony of higher functions such as cognition, emotion, sexual desires, hunger and thirst, breathing and blinking and growing stronger. These higher functions lead to even greater purposes, such as love, trust, morality, art, literature, music, mathematics and science, history and philosophy, community and social structure, diverse occupations, economies, technology, sports, film. But, beyond these, what is the greater collective purpose of humanity, and indeed all of nature and all creatures? If there is no Designer then there cannot be any greater purpose, as eventually heat-death of the universe burns the whole game of life to a non-remembered end. If we can deduce a Designer, and qualities about the designer from what has been made, then why not also attempt to infer notions of what we were made *for*? Humans, in particular, appear to have a deep biological need for personal purpose. If this is indeed the case, then the study of biology can inform philosophy, sociology, and other domains of rational discourse.