The Genetic Code
A Good Example of Bad Science and Why it Needs to Get Much Better.

Mark White, MD
Copyright Rafiki, Inc. 2006

"All religions, arts and sciences are branches of the same tree. All these aspirations are directed toward ennobling man's life, lifting it from the sphere of mere physical existence and leading the individual towards freedom."

Albert Einstein

Introduction to Knowledge

Our world is founded on universal laws that lead to complex phenomena. We study this using the scientific method. This creates scientific data that must somehow be "modeled" to be understood. We must first place the data and our understanding of it into a model, or into a structure that makes those things understandable in some way. In the broader scope, we must have a formal system, or a type of language to effectively communicate all of these things. Our systems of communication always depend on various icons. Ultimately, the entire process leads to human knowledge.

The act of acquiring knowledge has been around for thousands of years, but recent advances in culture, science and information technology have made the process far more complex and much less intuitive than perhaps we realize. As a result, the evolution of human knowledge has led to a need for a formal academic discipline called informatics. Informatics is a little hard to conceptualize because it is quite broad, but we might define informatics as the organization, presentation, visualization, interpretation and analysis of information and information systems. In my mind it is nothing more than the scientific method on the steroids of modern information technology.

The pressing need for informatics is nowhere more acute than in the area of life sciences, where information is being accumulated at exponential rates. There are thousands of examples of bio-information and thus the need for bioinformatics. The central paradigm of this field - grandfathered into the field by default - is a concept known as the genetic code. The genetic code is a simple

human metaphor for a complex natural phenomenon. It is a living language of molecules that all cells on earth depend upon to manufacture proteins. This process somehow has "molecular information" being stored in DNA and somehow has that information being translated into the many important molecules that we now know as proteins. Our knowledge of this system is quite recent, but our understanding of it is woefully inadequate. Before it can improve, we simply need a better understanding of knowledge itself.

The idea of a code of life was first popularized in the 1940s by the brilliant quantum physicist, Erwin Schrodinger. By the 1950s, the genetic code had been formally modeled and partially elucidated, and by the late 1960s, the genetic code had been formally and completely described. Today, the genetic code can be quickly visualized, organized, analyzed and conceptualized by a tiny data table known as a codon table. This is today the universal icon for the genetic code. Our all-too-simple concept of a genetic code, along with its now all-too-familiar visual icon, together perhaps provides the first true paradigm of bioinformatics, and so it is used to teach the fundamental concepts of biological information. In this area, the codon table completely informs our thinking today.

Unfortunately, the genetic code is merely a good example of science and bioinformatics gone bad; therefore, we need a better understanding of where and how it went wrong, and a better understanding of how and why it must change.
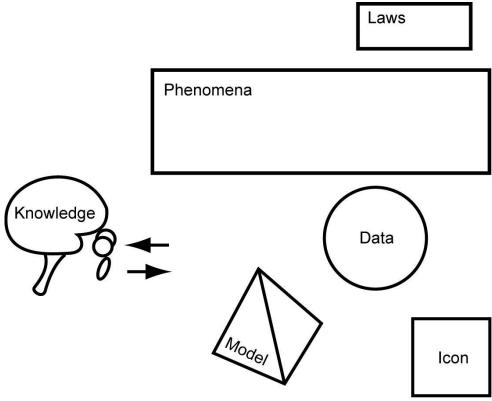

Figure 1.

Figure 1. introduces a system of visual icons to help us conceptualize the structure of human knowledge. The universe is built upon basic principles that we perceive as physical and logical laws. For instance, there must be six faces on a cube, 2 + 2 = 4, two objects with mass will be attracted by gravity; these are examples of the kinds of basic laws that appear to construct our universe. These laws operate over vast times and vast space to create physical phenomena of our reality. Two hydrogens will bond with one oxygen and oceans form, sunlight hits chlorophyll and photosynthesis results, protons collide in a chain reaction and an atomic bomb explodes; these are all examples of physical phenomena. We use science to formally and systematically study natural phenomena, and this generates observational data - gobs of it. However, this data will always reflect but a tiny subset of the actual phenomena. That is the nature of nature.

Science provides a narrow window through which we are allowed to peer at the natural world. The specific area of natural phenomena, the type, character, quality and amount of data will all be dictated by human knowledge and human prejudice. Human knowledge is at bottom dictated by data, models, language and a strong set of beliefs.

We use models to guide our observations of the universe. The earth rotates elliptically around the sun, sodium and chloride occupy alternating points on a cube to form a salt crystal, a pulley is an atypical form of a lever as a mechanism to apply force; these are all examples of the simple conceptual models that we use to guide our observations. These models require systems and languages for our communication as well as our understanding of them. Knowledge always requires languages. Icons are therefore created. The word "red" is a linguistic icon for an electromagnetic wavelength of approximately 750nm, and a patch of specific pigments on paper can be a more literal example, or a powerful visual icon of that same phenomenon. "F=MA" is yet a more robust scientific icon that symbolizes complex general relationships quickly and powerfully. Words are indeed powerful icons, and graphical symbols are still more powerful. After all, a picture is worth a thousand words.

Basic understanding of the interrelationships and the juxtaposition of these essential components of human knowledge lay at the heart of science and its new tool, informatics. At bottom, informatics is the practice and art of using information to construct and communicate human knowledge. It is a contemporary prosthetic device to augment science primarily through the use of modern information technology, but it is at all times grounded in the first principles of human thought and our understanding of natural phenomena.
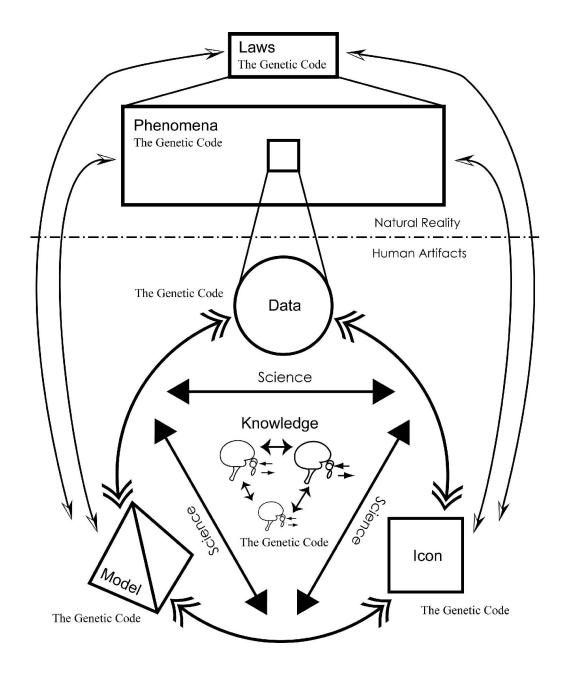
Figure 2.

Figure 2. illustrates the interrelationships of the components of human knowledge and the central role that science plays in the overall scheme. It illustrates the various and logically distinct aspects of our metaphor of the genetic code. When we speak of things, it is always important that we first know the exact type of things of which we speak. The genetic code can mean different things to different people in different contexts. It is all-too-easy to become confused about whether our thoughts and words are referring to something real

or something we merely imagine to be real. Knowing and understanding are two entirely separate things, and they are tied together by a strong belief that we understand that which we think we know.

Models create icons and icons serve as the currency in the economy of building and communicating models. Models and icons are derived from data but also dictate the acquisition of data. All of these things are input to and output from human systems of thought and human interaction in the ever-escalating creation of human knowledge. Models, data, icons and knowledge are human artifacts of our interface with the natural world. They are the sum total of our insight and understanding of natural laws and phenomena. However, these human artifacts do not necessarily reflect a literal or accurate picture of those things in nature. They are merely ghosts of reality. The more tightly all of these components are integrated and the more internally consistent they are with each other, the more accurate will be our perceptions and knowledge. As these things accumulate and become more refined, the process of generating new knowledge becomes greatly accelerated.

The genetic code is a paradigm of bioinformatics, perhaps a paradigm of science in general. It is the very image of science itself that man should somehow be able to decode the secret code of life. The name is a linguistic icon that can be applied to all components of our knowledge system regarding genetic information and its translation. The codon table is a powerful visual icon that solidifies and amplifies the broader system of knowledge. These familiar and preferred names and visual icons are today quite simple, and so the model and data are also simple. This presumably reflects the reality of simple natural phenomena and laws. However, this has broken down in profound ways and now merely creates a demonstrably false context for our perception and understanding of this system. It has misguided our thoughts about the laws and operation of this important system of molecular translation. The entire knowledge engine breaks down here at its core, at the visual icon. The codon table *is* the genetic code for most people, but it is an inadequate icon with an inadequate structure for understanding this incredibly complex and robust natural system of information. Consequently, the model and data that we currently possess is inadequate toward an explanation of this phenomenon. Further investigations of these important natural laws and phenomena have been diminished by it. The codon table is a good example of informatics gone bad. The genetic code is a good example of science gone bad.

If we are to advance we must question everything. We must break the system apart and reassemble it in entirely new ways. We must question our data, our models, our language, and most importantly, we must question our beliefs. After all, it is our strong but wrong belief in the validity behind our metaphor of the genetic code that has led us into this uncomfortable situation.