

Biocalc 1.0

By Mark White, MD
Copyright Rafiki, Inc. 2008

Introduction

Biology is the scientific study of life extending from the molecular level to the extraterrestrial level. It is a large and complex field of study, to be sure, and so it has many languages, yet it has no common language. The absence of a common language in biology has caused many problems; not only in the communication of biologic concepts but in the proper formation of essential concepts as well. Therefore, the task at hand is to invent a language of life. The primary tool for doing so, set theory, is the same as the primary tool for inventing most mathematical languages. Note that the basic task in the invention of this new language of life is to first create sets and then to give them names; names that themselves contain meaning. But sets of what, and precisely how shall they be defined? How shall they actually be named? The answers are not simple, of course, and in fact any systematic answer must embody the language formation process itself. However, it should be clear that the creation of any new language in this area will constitute the creation of an entirely new model of biology, a creation that is desperately needed.

The first name we will invent is the name of the language to be invented here, a language that is biologic in nature, and yet one that borrows heavily from the foundations of mathematics. So we shall call it Biocalc 1.0. Like all languages, this one will surely evolve, so we have also given it a quaint version number: 1.0. As a synthesis of biology and a meta-mathematical language vaguely akin to traditional calculus, Biocalc will be both dialectal and algorithmic. The purpose of Biocalc is to define sets and the logical relationships between them, give them each names, and to thereby derive instantaneous “meaning” from their proper usage. Ultimately, this language will provide the necessary structure to advance human understanding of biology, the study of life.

Life has an undeniable logic structure, and so the language we use to describe it should share that structure in the most efficient way possible. At the very least, Biocalc will finally provide us with a mechanism to communicate ideas clearly, propose theorems, and test those theorems on the basis of logic alone, an essential ability that has been sorely lacking from the entire history of biological sciences.

The foundation of the sets to be created in Biocalc must start with the always nebulous concept of a material object. After all, the original definition of “atom” was intended to be applied only to that which cannot be divided. Of course this belies the physical and epistemic absurdity of sub-atomic particles, but the foundation of Biocalc is intended to actually reside at the atomic level of human understanding of the universe. Building the language upward from that foundation, we will pursue our goal of maintaining a reasonable degree of internal consistency so that names can retain a large measure of logical meaning. Objects enter our language at first as the atoms portrayed by

the perfect order within the periodic table of physical elements. New objects then quickly emerge as sets of other objects, and they expand into the logical combinatorical sets in biology, which is of course a logical subset of all things in the universe.

Atoms combine into what are traditionally considered to be molecules, but molecules themselves require arbitrary definitions in many ways. There are many different kinds of molecules, and a molecule is not really a molecule until we say it is. Worse still, there are many different ways to define a molecule, and the system of definitions will greatly impact on how we see, think and work with molecules. However, these complex combinations of atoms must then expand into more complex combinations until ultimately we can include the classical sets of biological taxonomy that formed the very language - the system of thought - employed by Charles Darwin. Life is a system of molecules, but the system Darwin used to understand and describe it had no concept of a molecule. It is instructive to note, however, that his system provided him with the essential data, organization and insight for his seminal work in the then relatively new field of biology. Darwin's mode of thinking has obviously organized our thinking in modern biology, so our language should logically be founded on that particular mode of thinking. In other words, Darwin used exactly the right system, and so should we, only bigger and better. The language we invent will span the gap between classical chemistry, physics, mathematics and Darwin's grand view of biology, an ambitious goal, to be sure.

Paradoxically, our aim is to transcend the material world so that we can ultimately introduce a metaphysical domain of pure "information," a domain in which the actual processes of life will ultimately be located. Yes, it is materialism at the core, but a kinder, gentler form of materialism, one that accepts the elusive and ephemeral nature of reality, especially living reality.

To achieve this lofty goal, a program of pure complexity will be pursued. In other words, there will be no sets - and therefore no names of sets - that are based on absolutes and utter simplicity. Each set construction is complex. Each set is bound to be defined and redefined continuously based on shifting relationships. In addition to the defined objects in each set, systematic definitions will be conditioned on three primary descriptors: existence, time and relationship. In other words, a set can only be described by the objects, the existence of those objects, the context of time, and the many necessary relationships between specific parameters of any set with those of all other sets. In this way no absolutes can be created beyond the axioms of the language itself. In this way all things are ultimately defined relative to all other things. It is pure comparison, or a purely metaphorical language, as all languages must be. At first this may appear to make things much more difficult than they need to be, but in the end, this is the only path to the systematic simplification of biology, a goal long sought but far from obtainable with any current language. After all, before the situation can be discussed and debated, a language capable of doing so must exist. Life is itself a language, and yet we are now in pursuit of a language that can be used to describe life. Therefore, we find ourselves in desperate need of a language of languages. There currently is no coherent discussion about this language or the vital issues it might resolve; consequently, this step of goal-specific language creation is the necessary first step forward. A lengthy discussion must precede the creation of the actual language. Vigorous debate will surely follow. That is the goal.

Language

The creation of any language must take account of four things: syntax, semantics, pragmatics and semiotics. Unfortunately, this is much easier said than done. Exactly where, say, syntax ends and semantics begin is never entirely clear in any language. Blurred lines will be a prominent feature of Biocalc, but much less so than in every existing language in this field. The lines in Biocalc will be less clear than a pure symbolic language, for instance, but clearer than any language yet employed by biologists at any level. Actual lines are more imagined than real in any language, but the currently favored languages of biology rarely acknowledge the need for any lines at all; the obvious exception being the case of classic biologic taxonomy, a system of nomenclature that closely parallels Biocalc in structure and epistemic value. In fact, the classic system of taxonomic nomenclature will append quite nicely to the high end of Biocalc, or perhaps the high middle, but the global system itself must be built from the bottom up; *not* from the top down as it has historically been built.

The semiotics of a language pertains to the overall “purpose” of the language. The semiotics of Biocalc relate to its ability to describe the natural universe in such a way that complex living systems become clearer in our thinking and communication. Biocalc is designed to replace the patchwork system of languages that currently dominates modern science and is, in fact, convergent within biology. Biocalc is designed especially to deal with the complexity and confusion that now reigns supreme at the molecular level of life. Think of it as a computer language of biology, except the computers that will run it are walking, talking, writing and fighting humans in the flesh.

The pragmatics of a language must address the actual language usage as a real world form of communication. It is the actual history and future of the language itself. Biocalc will surely be used both in written, formal, algorithmic pursuits as well as verbal, dialectical and colloquial pursuits. In other words, it will be used in concise symbolic form as well as expanded verbal form. No language could ever be optimal for both simultaneously, so Biocalc will require a system of compromises, balancing one with the other. Additionally, perhaps the most significant concern regarding the pragmatics of Biocalc is the issue of how it will deal with all of the many confusing and contradictory legacy languages in biology. It is in no way practical to abandon all legacy languages completely, yet on a purely pragmatic level of understanding, the legacy languages perhaps ought to be abandoned. They suck – big time. They are causing more trouble than they are worth, but again, the inventors of Biocalc will seek a reasonable policy of compromises with legacy languages that maximize efficiency and ease of use, and minimize the confusion that will inevitably follow the false familiarity of concepts arising from our beloved legacy languages.

Semantics involves definitions and meanings. Semantic concerns are always metaphorical, or more precisely, semantics always involve comparison. Syntax, on the other hand, addresses the formal rules of a systematic language – the formal rules of comparison. Biocalc is a language founded on systematic comparative nomenclature, so the line between semantics and syntax hardly exists at all, let alone could it exist in a way considered easy to define. This will become clear as we get further into the language creation itself, so rather than more explanation here, we shall jump in and begin building the language from the ground up. The essential task of building Biocalc is first to create

and then to name sets. We will then create methods for using sets to form valid sentences. A set can be anything we chose. Most sets already exist in mind and name, and of course those sets will need to be defined again in Biocalc if the system is to be effective. Each set will consist of objects and basic descriptors of the objects. The three types of basic descriptors include: existence, time and relationship.

The Three Basic Descriptors

Existence descriptors:

- Ideal
- Actual
- Rational
- Nonrational

Time descriptors:

- Qualitative
 - Past
 - Present
 - Future
- Quantitative
 - Instant
 - Continuous
 - Period
 - Phylogeny
 - Ontogeny
 - Evogeny

Relationship descriptors:

- Intra
- Inter
- Extra
- Meta
- Systemic
- Global
- Function
- Code

Existence Descriptors

Biocalc will primarily be used to define, quantify and analyze various forms of information. Life is composed of information, especially molecular information, but molecular information remains a curiously undefined entity. Information in general has many algorithmic and dialectal forms, and Biocalc is designed to finally bring valid forms of information into the field of biology. However, the foundation of information science involves identification of sets of things to count, methods of selecting things from the sets, and knowledge of the probability of making any selection. The probabilities depend on the various forms of existence of things. Therefore, without any knowledge of forms of existence, information can never be properly defined, quantified and tracked. Ergo, we have no workable definition for molecular information today. It is a major problem that must be addressed before progress of any type can be made.

The four basic existence descriptors in Biocalc are: ideal, actual, rational and nonrational. We chose nonrational instead of irrational because we prefer – where possible - to have a distinct initial letter to symbolically stand for each descriptor: I, A, R, N. This is especially useful when the language is reduced to efficient symbolic formulas. These “existence descriptors” provide an obvious analogy to number theory, so we will describe them partially within that context.

Plato described ideal forms, so the perfect solids are named Platonic solids, and the concept of ideal forms as the foundation of mathematics is a popular school of thought known as Platonism. A cube is an ideal form, for instance. In geometry, points, lines and planes are ideal forms of things to be used as tools to prove the existence of other things by way of relationship to ideal things. The concept of three things involves the ideal form of three, a thing believed to be unchanging when applied to any and all objects. In the world of molecular biology there are many examples of ideal forms, like the double helix of DNA, or any illustration of the molecular structure of ATP is a good example of an ideal form.

Actual existence involves something that actually exists, like a three inch block of pure ivory cut into a cubic shape. Likewise, sodium and chloride atoms form actual cubes in a salt crystal. The ideal form is a perfect cube, but the ivory and the sodium-chloride atoms are actual forms that inevitably are not perfect. The concept of three is perfect, but actual applications never are. For instance, I can claim to have three golf balls in my hand, but the truth of this statement depends entirely on my definition of a golf ball, not on my understanding of the concept of three. Three is easy, but golf ball becomes surprisingly hard. Perhaps our chosen definition of a golf ball turns entirely on the weight of the object itself, and perhaps one of the three objects in my hand weighs “too much” to qualify as a golf ball. It would then be a false statement that I have three golf balls in my hand. We can similarly study actual DNA and say that it is in the form of a double helix, knowing full well that the ideal form is a double helix, but a significant portion of our actual molecule is not in that form.

Rational existence describes all of the possible ways something might actually exist. Naturally, there is always a huge measure of subjectivity involved in deciding the rationality of the number of possible ways in which something might exist. We can have three golf balls, three hats, three children, three cubes of ivory, but how many ways can

three exist? There must at least be a one-to-one correspondence between the rational existence of something and the natural numbers. To be rational in this context is to be countable. To understand the information content of a molecular structure, one must first know the rational existence of that particular structure. Without that, there is no real information, only pretend information.

Nonrational existence corresponds well with the concept of irrational numbers. If one defines a rational number as a relationship between actual, “natural” numbers, then there are many numbers that we know about and talk about freely, but that cannot exist as actual numbers. Pi is perhaps the best example. Phi is my favorite. The square root of two, for instance, is nonrational in this context because we cannot find two actual numbers to be used in a concise representation of it. A double helix of DNA with all of the bonding partners inverted 180 degrees is similarly nonrational. There are many different cubes of ivory that cannot exist because we can perhaps only describe them in theory, yet not decode any rational path to their actual existence. Perhaps the technology does not exist to produce an actual example. We cannot, for example, propose a cube of ivory the size of our solar system. Similarly, the concept of a perfect cube of ivory is also nonrational because ivory is itself an imperfect material and cannot therefore be used to make a perfect cube. It is a purely constructivist point of view: if you cannot make it then it does not exist. Until you prove that it can be made then we cannot merely assume that it can be made. What about the set of three infinities? It too is nonrational because we may or may not be able to find a single thing that can be considered a single infinity, but we cannot logically then organize three of these things into a single coherent set. What about the set of all possible DNA? This too is nonrational because DNA has no natural bounds in terms of possible components and possible combinations of those components. Who is to say that DNA must consist of four nucleotides? More importantly, there is not enough space in the known universe to contain such a set, so we cannot have any actual set of all possible DNA. Only when we precisely define DNA can we have a rational set of it. Nonrational sets are especially important when we begin to think about relationships between molecular sets, more so perhaps than the sets themselves. These concepts must form the basis of molecular information and the codes in which it is trafficked. Anything less is nothing short of a hoax.

Time Descriptors

Life is evolution. Evolution is the organization of information through time. Molecular information provides the logical foundation of life, so life at bottom is the organization of molecular information through time. We cannot, therefore, define or describe life without addressing the issue of time at every level of life. Consequently, Biocalc provides a robust set of time descriptors that break time information into two classes: quantity and quality. The qualitative descriptors are intuitive – past, present and future – but the quantitative are less so.

We can talk of an instant in time, a period of time, or all of time as a continuum. Time animates the universe, so an instant in time de-animates the universe for us and our analysis of it. A period of time must contain at least two instants: a beginning and an end. Biocalc, by convention, will call a beginning the alpha instant and the end will be

the omega instant. A period of time can exist as a past, present or future period relative to the object of interest. These are called respectively phylogeny, ontogeny and evogeny (p, o, e). For example, your ontogeny stretches from your conception to your death. Your phylogeny extends from the beginning of life on earth (or perhaps to the beginning of earth) up to your point of conception. Your evogeny (a newly coined term) stretches from your conception to the death of all your descendants. Obviously your alpha ontogeny and alpha evogeny are the same, but the alpha ontogeny is also the omega phylogeny.

A continuous time element has no known beginning or end. However, we can still speak metaphorically of the alpha and omega points of a continuous time function. For instance, the universe operates on a continuous timeline, as far as we know, yet we have no compunction when speaking of the alpha and omega points of the universe.

Relationship Descriptors

As we begin to define sets we begin to realize that there are different kinds of relationships that will eventually interest us. The most basic kinds of relationships are those that exist within sets – intra - and those that exist between sets – inter – and those that exclude sets - extra. Then there is the curious relationships where a set participates in a relationship, without doubt, yet the set itself cannot define or be defined by the relationship itself - meta. The epitome of a “complex” relationship is a meta relationship. New sets and their properties can emerge as meta relationships of other sets. The properties of metaset simply transcend the properties of the sets that create them. Nothing more can be said. However, this also brings to the fore the logical notion that sets are virtually always merely sets of other sets. So we now also need a way to define the extent or scope of any relationship or set of relationships. A relationship can exist between many but not all things within or between sets. This is a systemic relationship. When a relationship exists between all things in question, then it is a global relationship. Whether a relationship is systemic or global depends entirely on the definition of the sets in question.

A function is any relationship between two sets. A code is an algorithm that defines a precise relationship between two sets. To clarify: functions can be mapped. Codes can precisely “explain” the mapping of functions. For instance, set A contains 2, 4, 6, and set B contains 7, 15, 23. A map of the function between these two sets will look like this: 2-7, 4-15, 6-23. The code for this function is: “multiply A by 4 then subtract 1.”

Our First Taste of Basic Sets

We have now laid the basic ground rules for putting Biocalc into action. The vast majority of work is done in defining sets, and it is no small task, to be sure. However, once sets are defined, statements or theorems can be made, and these statements can be analyzed as to their value of true or false. They can be logically or empirically proven true or false. It is not an emotional matter of how you “feel” about a statement but of

whether it is right or wrong, smart or stupid, productive or counterproductive. However, because the task of naming is so critical and so overwhelming, we will focus at first mostly on nouns and adjectives, but we must also create common conventions for using adverbs, verbs, prefixes and suffixes, etc., just like we must in any language.

Unfortunately, to do so will require that we import many of the words already familiar to us from our romantic legacy languages. The obvious place to start is by importing the taxonomy of biology. It is one of the few languages that actually works for its intended purposes, and notably works quite well.

Species – rational set of individual organisms of arbitrary definition.

Genera – actual set of species of arbitrary definition.

Family – actual set of genera of arbitrary definition.

Order – actual set of families of arbitrary definition.

Class – actual set of orders of arbitrary definition.

Phylum – actual set of classes of arbitrary definition.

Kingdom – actual set of phyla of arbitrary definition.

Unfortunately, this system is woefully incomplete, and it has more than a small opportunity to confuse us, so we will modify and extend this particular system. Please note, however, that this was the simple system itself that provided Darwin with his fundamental insight regarding natural selection as a mechanism for evolution. Put yourself in his shoes. Darwin was a naturalist. It was his job to review actual specimens of organisms living and dead, and to then place them *precisely* within this system of nomenclature. Sounds simple, but is hard – impossible, really. Being a smart man, he realized that his job was logically and literally impossible. In other words, “species” is the “bottom” set yet defies logical definition within this system of classification. So it is a never-ending game of “catch up.” However, Darwin realized that life itself was organized by a nested network of relationships, and yet the foundational relationships clearly could not be defined. Darwin knew that it was nature that made the actual system and so it was obviously nature that needed more definition. He did not find the ever elusive definition of species; he simply defined the process that leads to the origin of species within a natural system that is poorly represented in a formal system of nomenclature. What part of that is hard to understand? Yet there is a big difference, and until that difference becomes clear, the student of biology will forever remain in a fog.

Unfortunately, it appears at first blush that we will need some of these powerful names elsewhere, to be sure, so we must clear a path toward distinguishing the terms used in this context from the same terms that must be used elsewhere. We will do this by always capitalizing the name and prefixing it with the lowercase “c” plus “-“ to stand for “classic,” or perhaps “continuous.” We can then write and talk of c-Species and c-Family. This will be pronounced “see species” and “see family.” When we do not use the “see” convention within this language, then it will be clear that our usage of these words is not meant in the classic context. After all, many of these highly descriptive words will be more useful in other contexts, so the burden of appending “see” to them here will only mitigate the confusion they might cause elsewhere.

Next we will extend the system at the high end. We will define c-eLife (pronounced “see-e-life”) as the set of all c-Kingdoms in the universe originating from

earth, and define c-uLife (pronounced “see you life”) as the set of all life in the universe. Of course the highest set is then Universe, so it contains c-uLife and all else. Finally, we will expand the system at the low end. We will base the system on the atom, so the sets will emerge from the universe and then start with c-Atomic, move up the hierarchy through c-Molecular, c-Subsystemic, c-Cellular, c-Systemic and c-Individual before merging with the classic system. For the sake of efficiency, by convention we will collapse c-Genera to c-Kingdom into c-eLife, and further by convention and convenience, we will also loosen the name to c-Life. In other words, c-eLife plus c-uLife are synonymous with c-Life. In other words, we are co-opting the classic system and intentionally making it symmetric – like the color wheel – and decidedly bottom-heavy. Now, the basic hierarchy of expanded classic sets looks like this:

Universe – rational set of all things.
c-Atomic – rational set of atoms.
c-Molecular – rational set of c-Atomic.
c-Subsystemic – rational set of c-Molecular of arbitrary definition.
c-Cellular – rational set of c-Subsystemic of arbitrary definition.
c-Systemic – rational set of c-Cellular of arbitrary definition.
c-Individual – rational set of c-Systemic of arbitrary definition.
c-Species – rational set of c-Individual of arbitrary definition.
c-Life – rational set of c-Species of arbitrary definition.
Universe – rational set of all things.

All of these sets are defined as rational, continuous and global. The alpha and omega points are all the same. This does not, however, preclude their use in any context we may choose down the road. That is the whole purpose of Biocalc, to create definitions that can be used for other definitions. This particular set of definitions is mostly arbitrary. In other words, there are an infinite number of possible definitions to choose from, but a set of definitions must at some time be chosen. There will always be stylistic differences, because there is no accounting for taste. Some are lumpers and some are splitters. However, the semiotics of Biocalc are designed to focus attention on c-Molecular, a very low-level set indeed, but in order to do so, one must have a context for this level in relation to all else. It’s a tough nut to crack, but this is the best way to get it cracked. We promise.

The definitions here have been made as broad as possible and the Universe has been added to the upper and lower bounds of the global hierarchy to provide a much needed sense of global symmetry. These things have been done in anticipation of some of the larger conceptual difficulties that we can anticipate within any system of this kind. For instance, we know that the universe is the ground level for any epistemic, scientific, theologic, mathematical, or philosophical system of word or thought even partially capable of explaining life, but at the same time, Universe is the highest level as well. A situation most curious, at the very least. But, in-between the conceptual Universe at bottom and at top, relationships will exist that are prone to inversion based on conflicting needs leading to conflicting definitions. Note that man is part of life, yet life is the foundation of man. Likewise, if we chose to define the Universe as a logical system of pure information, then we can perceive a hierarchy of information systems distinguished

by their ability to somehow develop a model of the universe. This hierarchy might appear as follows:

Universe – rational set of all things.

Life – actual set of all biomolecules.

Man – actual set of all men.

Mathematics – actual set of all mathematics.

Computers – actual set of all digital technology.

Universe – rational set of all things.

Note the differences - that this particular hierarchy is now based on an actual time of appearance rather than rational forms of any set. The exact form of the hierarchy is now made simply by estimating the relative alpha point for each actual set. It is not particularly difficult to do so. Yet, the overall pattern formed of such a thing is striking and could hardly be denied. Moreover, we can now see that the pattern of phylogeny and ontogeny for each of these sets is exactly the same. They are nested, or fractal, so they are self-similar in many significant ways. However, if it all boils down to pure information, and in this case - by definition - it does, then the pattern clearly repeats. Yet the period of repetition is also clearly becoming compressed. In other words, the general pattern is accelerating through time. This is a big clue. This is the basic pattern of life. These are the fundamental patterns we seek to examine at the heart of biology. These are the patterns that have heretofore perhaps been obscure, but become obvious in Biocalc.

In plain English: The pattern of the Universe leads to Life in the same way that the pattern of Life leads to Man in the same way that the pattern of Man leads to Mathematics in the same way that the pattern of Mathematics leads to digital technology in the same way that digital technology allows us to better discern the many complex patterns of the Universe. Each “leads to” is exponentially faster than the last. In other, other words: we use computers to model the universe. The universe uses us to model itself with computers. Duh. Take a moment to realize that there is an undeniable consistency to these patterns. It is complexity and emergence writ large. The inter-set and intra-set relationships are remarkably consistent; however, we might also say that each is a meta-set of the previous. This is the ultimate meaning of “meta” in this context. What is life? What is the “meaning” of life? One might logically conclude from this context that information in the Universe is rapidly evolving in a direction that allows the Universe to better know itself – metaphysics, pure and simple. What could be simpler?

If this is not Platonic, what is? If this is not Constructivism, what is? If this is not formalism, what is? If this is not pure mathematics, what is? In fact, it is even possible to now draw direct and meaningful parallels between the development of mathematics in Man and the development of molecular life in the Universe. In other words, the Universe is obviously doing some kind of math with molecules. There are molecular symbols, functions and codes, just like formal human mathematics, and so now we seek to understand these awesome systems through our puny language called Biocalc. We shall nonetheless struggle forward in an attempt to translate and transcribe this naturally occurring molecular language into a human language that can possibly communicate the raw beauty and awesome complexity of the Universe.

Note, however, that the order is not the same in the classical “see” hierarchy as it is in the information hierarchy. This too is an example of how sets must be perpetually and precisely defined to convey meaning. If the definitions are followed closely then we should find no logical problem with this seeming oddity. But, whether or not this constitutes a paradox or a true and problematic form of contradiction in our system of communication, it is too early to determine, so we shall continue to grind our way forward in building our fabulous new language. We shall leave these truly meaty foundational questions to the philosophers among us.

Adding Terms

We will now begin to add a taste of some additional terms that will provide needed utility moving forward. For instance, the term clonal will refer to a set of many things in which all things are taken to be identical by definition. Communal will be used to refer to a set of many things that are taken to be different by definition. For instance, the same set can be defined as clonal or communal depending on our linguistic needs at hand. To give an abstract example, we can say the clonal set of threes, or we could say the communal set of threes. In the first case, clonal, we mean by definition the set of all possible things grouped into threes that can be taken as identical. In the second case, communal, we mean the set of all possible things grouped into threes where they cannot be taken as identical. This demonstrates that the exact same thing can be defined in diametrical opposite ways. It is not the things themselves but their definitions that matter most. This may sound silly now, but it is vitally important when we begin to apply this language structure to molecules and complex living systems.

The term “global” will only apply to a case when something is true for all things under consideration. If there are exceptions, then the statement is not global but systemic. For instance, “sequence determines structure” is a global statement for proteins today. The statement has been repeatedly tested and repeatedly proved false, yet the statement is still made as if it were globally true. Why? The problem is not with the thinking but with the language itself. After all, there are many words and many “true” sentences that depend on the truth of this one statement to maintain their meaning. Rather than abandon the words and the sentences, we choose to abandon the truth. This is no way to run a rodeo. Folks, we have a big problem, and Biocalc is the solution.

We have recently made a list of no less than twenty-six synonyms for the term “translation” and it appears that perhaps we might need them all, and then some. This ain’t easy.

When we speak of molecules we must often speak of their spatial arrangement. Structure is the general term we will use when we speak of spatial arrangements, and we will use the suffix “form” to mean structure. To briefly illustrate, if we encounter a molecule that we intend to name “Erwin” then we can talk of the Erwinform. Furthermore, several conventions of sub-structure will also be observed. For example, if Erwin happens to be a molecule that can be partially described by merely listing a sequential list of substructures – like DNA - then we will refer to this particular kind of substructure format as sequence. The suffix “string” will be used to specifically mean sequence substructure, and so now we can talk of the Erwinstring. Note that there is an

ideal Erwinform, perhaps an actual Erwinform, a rational Erwinform and a nonrational Erwinform. There also are actual Erwinform phylogenies, ontogonies, evogenies, an alpha Erwinform and an omega Erwinform, an actual Erwinform ontogeny versus a rational or an ideal Erwinform ontogeny, etc... The concept of an Erwinstring also implies that there are sets of components that go into the construction of the Erwinform. A common component of a string sub-structure will be defined as a monomer, and the suffix “mer” will stand for an individual component of a larger structure. So now we can speak of Erwinmers in Erwinstrings representing Erwinforms. The need for these conventions will quickly become apparent as we struggle to redefine central and “familiar” things from our romantic legacy languages, things like DNA, tRNA, protein, and the genetic code.

Note that the world of biomolecules is dominated by consistent structures of two basic types: polymers of nucleotides and polymers of amino acids. They create a dynamic tension that creates all else. The first type will be given the prefix “nucl,” and the second type will be given the prefix “proto.” So we can now speak freely of various brands of nuclomers, protomers, nucliforms and protoforms, rational protostrings and actual nuclostings, etc... In other words, we can begin making sense of a complex living universe. Of course, in the romantic legacy languages we already have funny terms, like codon and peptide bond, that have limited meaning. We also have nonsensical terms, like protein, gene, molecular information and genetic code that have no meaning at all. Fortunately, these things can now finally all be logically defined within Biocalc. They can now carry actual meanings that can be logically operated upon and subjected to vigorous debate and empirical investigation. The self-contradictory tautologies can finally be laid to rest. Thank God. We can finally ask and answer important questions, difficult questions like, what is the genetic code? Where is the genetic code? These are questions that have never been asked, remarkably enough, but the reason for this is simple: logical answers are made impossible by illogical languages used to give answers. Biocalc will strip away the false veneer of scientific authority and reveal these anemic current models for what they are: vacuous place holders in the history of science. The homunculus lives, but not for long. We must first solve the simple riddles in need of linguistic structure before meaty answers of universal truth can have any epistemic value.

Truth up, Bro.

This paper is merely a sketch, not of what is but of what might be with the creation and development of Biocalc. It is the ghost of what is yet to come, an opening salvo in the call for biologic linguistic revolution. However, one must note that the current situation with our ill-conceived romantic languages of biology has a clear precedent. It harkens back to a time when Roman numerals were the standard for all of mathematics. They were fine in as far as they could go, for simple ciphers and such, but numeric systems of superior efficiency and clarity were obviously available, and so Roman numerals finally went the way of the abacus. The language of life is the same way now as Roman numerals were then – hopeless - yet until we begin to explore the infinite number of possible systems leading to superior alternatives, we will continue to speak jabberwocky to each other.

Let the games begin. Please come join us.