Everybody Knows the Genetic Code
An ornery myth of science

By Mark White, MD

Introduction

Everybody knows that two plus two equals four. The earth is round. Everybody knows that the round earth is about four and one half billion years old, and it revolves elliptically around the sun. These are just a few of the things that everybody knows to be true in one sense or another, at least everybody who knows anything knows these things. It would seem then that certain things in nature are not only inherently knowable, but everybody already knows them. It does, however, beg the simple question: What does it mean to know something?

The genetic code is another one of those things that everybody knows. It is now an established fact of science that the genetic code is simply the conversion of molecular sequences – a simple conversion from nucleic acids to amino acids. Everybody knows that the genetic code is linear, one dimensional, and so it is wholly embodied in a single table of data known as a codon table. The idea, of course, is derived from other ideas, and it is bolstered by dictums, such as "sequence determines structure" and "molecular information is sequence." The remarkable simplicity of the whole idea behind the genetic code allows us to view DNA as the elusive central "secret of life." These vital molecular secrets possessed by DNA are translated by the genetic code bit wise into protein, shedding excess information in the journey away from DNA like burning rocket fuel during liftoff.

Irresistible speculations are derived from our knowledge of the genetic code, speculations regarding its origin and evolution which lead to still larger speculations of life in general. There are philosophical musings about the genetic code's true "meaning" and how this meaning places life awkwardly within the grand puzzle of the entire cosmos – the bond between physics and metaphysics. Ironies abound. For instance, despite its simplicity, universality and central importance within the impenetrable complexity of life, the genetic code is to be considered linear and mostly arbitrary, the result perhaps of

a single chance event, a miraculous accident so powerful that it is now essential to all living cells. We don't know how it happened, but we do know that it did happen and probably only once; otherwise, we could not possibly see it the way we do now. This vital molecular code is seen as the default position for nature when performing the dance of life, the functional imperative, the frozen accidental legacy turned hyper-competitive advantage in the grand battle for organic materials. To be sure, the largest of ironies is that the genetic code has become central to our thoughts but remote from our curiosity. The reason why? It is dogma, and it is demonstrably false dogma. Contrary to what everybody thinks they know; nobody knows the genetic code in any meaningful sense, and perhaps nobody ever will.

The purpose of this essay is to demonstrate that the genetic code today is nothing more than an ornery myth of science, following in the grand tradition of such embarrassments common to all branches of science. Various meanings of the word 'myth' are applicable to the ubiquitous narrative of the genetic code. It is fictitious and it also perhaps serves as an allegory or parable. Of course all meanings of the word 'ornery' apply, since the genetic code myth is common, irritable and insufficient.

Granted, it is not too terribly difficult to contend that the genetic code remains unknown if one holds, as I do, that the past is inherently unknowable and the future is inherently unpredictable. The central folly of science – when it becomes folly, as in a case such as the genetic code myth – is merely revealed in the hubris of knowledge. A conceit wrapped up in the notion that a tiny part of certainty can obscure much greater parts of the still unknown. It is as though knowing something constitutes knowing everything. It is as though details that refuse to abide must be perversely taken as evidence for faith in our larger enlightenment. The crux becomes simply a problem of knowing what it means to know something. In this case, what does it actually mean to know the genetic code? How does one know when they know it? Who decides that something of this overwhelming complexity and importance is actually known? I contend that the genetic code is perhaps one thing that is inherently unknowable. However, by pretending to actually know it, we completely fail to understand it in any meaningful sense. Furthermore, our unwavering confidence in our current knowledge serves as a poor instrument for guiding our future knowledge.

Knowing Mathematics

$2 + 2 = 4$

Look at the symbols and read them aloud: two plus two equals four. What does it mean to know this? There are four things to consider: 1) you, the reader, 2) me, the writer, 3) the symbols themselves, and 4) something external to the symbols that we will call reality. We must consider the relationship that the symbols have between me and you, or the shared meaning of these symbols between at least two communicating humans. We will call this public knowledge. This is what "everybody knows." There is also the relationship between you and the symbols. You have internalized their meaning. You can now intuit a clear meaning from these symbols without any significant effort.

We will call this private knowledge. Then there is the relationship between the symbols and reality. This is the foundation of mathematics, and there are many schools of thought about this. We will briefly consider three schools of thought about the foundations of mathematics: Platonism, Formalism and Constructivism.

Platonism holds that there is an objective truth or structure to reality, and a formula such as 2 + 2 = 4 merely uncovers part of that truth. Conversely, Formalism is agnostic as to the nature of reality, and it holds that the symbols only have meaning relative to each other. Mathematics is seen in this way only as an internally consistent set of rules regarding symbols and their relationships to each other. It is a self-consistent game with no application to objective truth. Constructivism is perhaps a form of intellectual empiricism. It is somewhat of an experimentalist's approach to mathematics, a pragmatic stance that only things that can be clearly demonstrated can be considered to be true.

Regardless of their true foundation, for there to be any significance to the symbols, 2 + 2 = 4, we must be able to do something with them, expand on the concepts they represent or apply the logic of the formula in some way. For instance, we can generalize the formula by replacing some of the constants with variables. We can say that a + b = c. We can then insert any manner of things into the shell of the original formula and draw meanings by comparison. We already know that if we replace 'a' with '2' and say a = b, then 'c' can be replaced with '4' and we have produced a true statement. We might also venture out into a real world, say an apple orchard and pick some real apples. When we pick two apples with our left hand and two apples with our right hand we know that we are holding four apples. We know this without even counting all of the apples because of the logic of the formula. But what about other things that are counted - do they all work in fundamentally the same way? Is there a two-ness and a four-ness to reality - not to mention the properties of addition and equality - that always allows us to do these types of thing? What about things like cups of water? If we add two cups of water to two cups of water, naturally we expect to end up with four cups of water. But when we add two cups of water to two cups of alcohol we get less than four cups of liquid - granted, this is mixing apples with oranges. Yet, ice, it might be argued, is a form of water, and when we add two cups of water to two cups of ice we eventually end up with either more than two cups of ice or less than two cups of water. Adding two plus two in mathematics is linear because there is only one dimension of information involved. Adding water and ice can be non-linear because there is more than one dimension of information involved. The molecular structure of water changes with temperature, and there are so many other fascinating properties of water that must be taken into account before we can begin adding cups of water as if they were linear. The take home message is that simplicity is often a crutch or a veil for our ignorance of actual complexity when our sterile world of knowledge collides with messy reality. We must expect to see paradoxical formulas like, 2 + 2 < 4, and we must seek to understand the difference between our sterile expectations and the messiness of reality.

Numbers are derived from a line. The logic of most mathematical formulas depends on the inherent logic of a line. Real lines, like real triangles, squares and circles, are much harder to find in nature than they are to find in mathematics. The geometry of mathematics is dominated by points, lines and planes whereas nature's geometry is arranged throughout time and space in the forms of solids, fields and lines that are not

straight but fractal. The sterile geometry of a line serves as a good basis for simple arithmetic, but within the sterile line there is no space; there is no time, and all interactions occur within the line itself. Yet the geometry of nature is complex in both space and time. The logic of nature is rarely linear, so the sterile linear knowledge we hope to impose upon messy reality is perhaps not up to the task.

Knowing the Genetic Code

A + C + G = Threonine

In the same way that we know that two and two is four we know that A and C and G is threonine. The number line of arithmetic is merely replaced by the linear relationship between nucleic acids and amino acids. The formula is simple: a + b + c = d, where a, b, and c are any one of four nucleic acids, and d is a specific assignment of one amino acid from a set of twenty possible amino acids. Since there is a set of only four possible nucleic acids, there is a set of only sixty-four possible combinations of three of them. Each specific instance is called a codon. When we place all sixty-four instances into a single codon table we have fully described the conversion of nucleic acid sequences into amino acid sequences, which is what we call the genetic code. In this way we can now write the sum of all codon equations as:

Codon Table = Genetic Code

Everybody knows this. But in what sense do we know this? Obviously, it is public knowledge. Please note, however, my private knowledge is quite different – quite different indeed. I do not see the genetic code when I see a codon table. I have not internalized the dogma, so I cannot see any kind of depiction of reality when I see an ordinary codon table. All I can see privately is molecules, and molecules do not translate well from my brain onto paper.

What about the obviously correct formula ACG = Threonine. How do we know this? The answer is that when we run experiments where we start with nucleic acid sequences they always correlate with threonine in the corresponding amino acid sequences. It is a purely constructivist point of view, to be sure, but it is one that is willfully ignorant of the truth. Surely there is no Platonic reason that ACG = Threonine, so we do not know the formula in a Platonic sense. Formally, ACG = tRNA attached to threonine, so it is the antithesis of formalism as well. Yet the real constructivist might even argue that man has proven that by changing a particular protein we can also change the equation so that ACG = Alanine. So we do not even know that ACG = Threonine is a true statement in any sense of the word. It is not consistent with Platonism, Formalism or Constructivism, so in what sense do we know it? It's just one of those things that everybody knows; it is public knowledge, that's all.

A better question is this: In what sense do we find truth in any equation of codons and amino acids? Or perhaps the best question of all: In what sense do we find truth in the equation between the codon table and the genetic code? The answer traditionally

proceeds as follows: Every sequence of codons leads to a sequence of amino acids, and every sequence of amino acids folds into a specific protein.

We can see this simple logic much more clearly if we replace every codon with an integer. As long as we discipline ourselves to observe placeholders and a consistent use of leading zeros, then we can write formulas that determine integers for proteins based on the equivalence between the codon table and the genetic code. Here is one simple example:

27 + 04 + 61 + 17 + 09 = 2704611709

We can easily write a formula where the integers for five codons combine with placeholders to produce an integer for a particular protein. In this way codons actually equal proteins. But notice that since molecular information is sequence and sequence determines structure, we will end up producing many unique integers for proteins that are actually the same protein, so formulas like this one 2704611709 = 3004601810 will now have meaning and will be perfectly logical. This oddity, of course, is due to the redundancy of the genetic code. More than one codon can stand for the same amino acid. This is easily corrected by replacing all of the codon integers with integers for each amino acid. The formula now merely concatenates individual amino acids into strings of amino acids. How could that possibly fail? After all, a protein is merely a sequence of amino acids. The answer is simple: Codons do not equal amino acids, and proteins are more than mere sequences of amino acids. In other words, not all "synonymous" codons for threonine are equal. When one threonine codon is replaced for another, something unexpected might happen in reality. For instance, the protein might fold in a new way. Furthermore, the production of that protein might change in some abstract, statistical or generally unforeseen manner that proves to be significant on a larger, non-linear scale of molecular calculus. The first logical conclusion to be drawn is that the genetic code is not redundant in the way we believe it to be. Silent mutations are not silent. Silent mutations can and do change protein synthesis in ways we simply failed to imagine. So now our simple equations become broken in more destructive ways where $a \neq a$, and the products of formulas lead to contradictions, like $2704611709 \neq 2704611709$. We now have a genetic code logic where all three of the following statements can be true simultaneously:

$2 + 2 = 4$
$2 + 2 > 4$
$2 + 2 < 4$

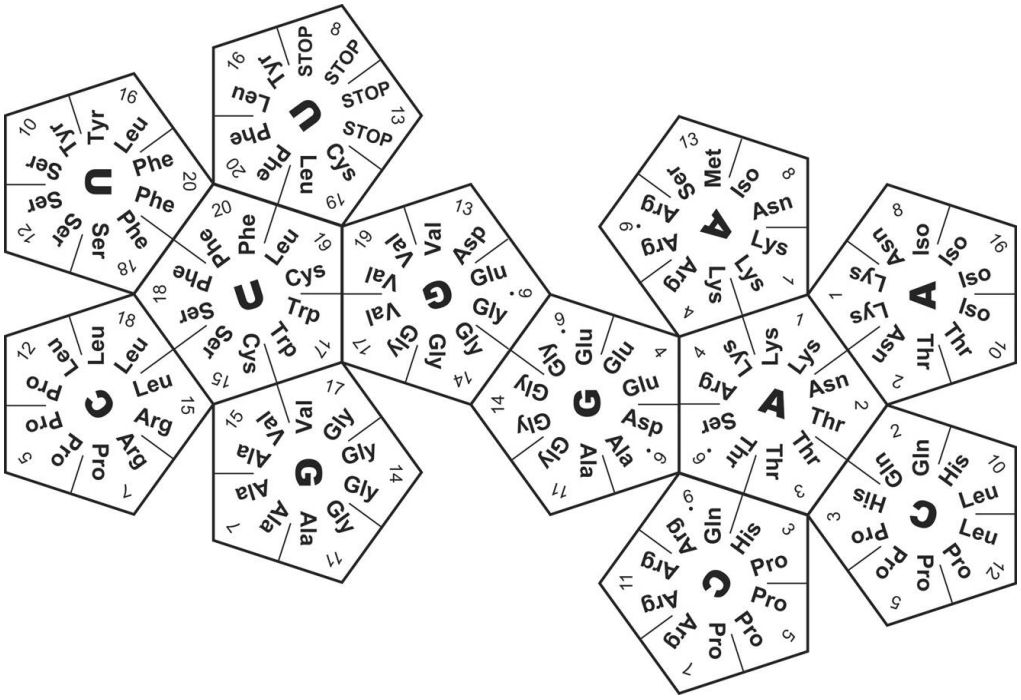
How do we know the logic of this? What is the point in now believing that we actually know the genetic code?

Separating Myth From Truth

My question, again, how do we know the genetic code? How do we know the logic on which it operates? How do we know its origin and evolution? How do we know

its proper place within life and the cosmos at large? The answer to all of these questions is that we don't. We are deluding ourselves to think otherwise. In no sense of the word do we know the genetic code today. We know that it exists, it works, and it is central to our understanding of life, but beyond that we know very little about it. Mostly, what we have today is not truth or knowledge but dogma and false knowledge. The dogma insists that we must know a non-linear phenomenon strictly as if it were linear. There are right ways and wrong ways to know something, but nothing can be any more wrong than this.

We do, however, know what the genetic code is not. It is a fairly reliable diagnosis of exclusion, and that is at least something. The genetic code is not simple; it is not linear, it is not one-dimensional, it is not universal, it is not arbitrary, and it is not frozen out of the heated competition for organic materials today. Most importantly, the genetic code is not a codon table. Codons do not equal amino acids. One-dimensional strings of amino acids do not equal proteins. Proteins can and do fold in many ways. Sequence does not determine structure. Molecular information is not merely sequence. Molecular information is not simple; it is complex in both time and space. Molecular information is non-linear and involves structure through time. DNA is not the central secret of life - it is the earth not the sun. Protein is the sun, but the earth and the sun must always move relative to each other through the organic universe. The real secret of life is the relationship between DNA and protein. The relationship between DNA and protein is Platonic:



The genetic code need not have arisen only once on earth and then became frozen. The genetic code perhaps converged, and now continues to compete aggressively within that logical channel of convergence. In other words, there are forced moves within the universe of competitive molecular codes. The varieties of this code today remain great,

extending well beyond differences in codon assignments with amino acids. For instance, transfer RNA constitute a vocal part of the genetic code, and tRNA populations can vary quite a bit. The competition is perhaps more apparent at levels of tRNA and protein than at levels of codons and amino acids.

The amount of information translated from DNA to protein is yet unknown, primarily because a working definition for molecular information is yet unknown. However, the fractal nature of molecular information is apparent. The relationship between time and information is also apparent: molecular information increases with time. This probably holds true for protein synthesis as well, but only if the genetic code is properly understood, and it obviously is not.

It is an ornery myth of science that everybody knows the genetic code. The myth is identical in form but worse in epistemic devastation when compared to the homunculus myth of embryology that lived and died not too long ago. The pathetic truth is that nobody knows the genetic code, and the prospects for knowing it in the near future are bleak, made all the bleaker by the fact that seemingly nobody can admit that they don't already know it. Everybody knows that everybody already knows the genetic code. But there is a more profound problem involved in the challenge of knowing the genetic code: Is this really something that we could know? In what sense could we know it? Can the genetic code be written down onto paper, or is it too spatial, too temporal, too complex? Regardless, the elegant truth will only be known when the ornery myth finally dies.