Are We Alone ?

Neil Breakwell

Is there life on other worlds, and maybe other sentient beings such as ourselves ? Or is life very rare, perhaps even unique to the planet Earth ? This is a question that must surely have been pondered by most of us. Indeed, it is a question that directs much astronomical research and space exploration. There is something in the human spirit that yearns to believe that other civilizations exist and longs to make contact with them. But just how likely is it that such civilizations are really out there ? Before we can attempt to answer such a question, we must first consider two more fundamental questions:

1. What exactly *is* life ?
2. How did life originate on the Earth ?

Our knowledge of life's fundamental building block, the cell, has increased so rapidly over the past few decades, that the first of these questions has become far more difficult to answer than could previously have been imagined. In the time of Charles Darwin, the living cell was seen to be something quite simple, comprising a substance which he referred to as "protoplasm", which he envisaged to be something like a homologous blob of jello, that might conceivably be created from relatively simple chemical processes. Our current technology has revolutionized our understanding of the living cell - we are discovering a staggering level of complexity comprising molecular protein machines and digital information processing systems that we do not yet fully understand[1]. An ex-Microsoft engineer recently told me how Microsoft are now learning from the data processing mechanisms within the cell to develop new software. Our answer to this first question of what life actually is influences our answer to the second question of its origin - a question which inevitably crosses the line traditionally drawn between science and faith. Broadly speaking, there are two possible answers to the question of how life originated on Earth - either it is the result of undirected "stochastic" (random) processes, or else, as the author of this article firmly believes, it was brought into existence through an intentional act of creation. A growing number of scientists from multiple diciplines now present objective and compelling arguments that the first life and its subsequent presumed evolution could not result from random processes[2]. If this is indeed the case, then the question of the likelihood of life elsewhere would simply be a question of the will of the Creator, which would make for a very short article. So, in the interests of a longer article, we will examine objectively what would be required for life as we know it to arise from random processes and attempt to quantify the likelihood of such an event actually occurring. Many argue that life *must* have arisen elsewhere because the Universe appears to be *so* vast and *so* very old. Science though is about quantifying things objectively and any scientific discipline which avoids putting numerical values to such bold claims is of questionable validity. So, can we assess more quantitatively the probability or "likelihood" of life having arisen elsewhere (assuming that this could indeed happen from random processes) in a similar way that we can determine probabilities for getting a head when flipping a coin or of rolling two sixes on a pair of dice or of winning the National Lottery ? To do this, we must first review a little mathematics, specifically a technique known as "probability theory".

Single Event Probabilities

The probability of an event happening is defined by the ratio of the number of desired outcomes for the event compared to the number of possible outcomes. For example, consider the flipping of a coin, which can result in heads or tails. The probability of getting a "head" is one chance in two, because there is one desirable outcome (heads) and two possible outcomes (heads or tails). This is expressed as "1 chance in 2" or "1:2" or as a fraction "1/2". Clearly, the probability of getting a "tail" is also 1/2. Now consider rolling a die, an event which has six possible outcomes. In this exercise, the probability of any specified outcome (eg rolling a 6) is only one chance in 6 (or 1/6). The probability of rolling any other specified number is also 1 chance in 6. We can see that the more possible outcomes there are for an event, the lower the probability of any specified or desirable outcome.

Multiple Event Probabilities

Now let us consider the scenario where there is more than one desirable outcome for an event. For example, what if we want to know the probability of rolling either a '6' OR a '2' on a die. Since the probability of a '6' is 1/6 and the probability of a '2' is also 1/6, we can say that the probability of rolling either a '6' OR a '2' is 1/6 plus 1/6 ie 2/6 or 1/3, ie 1 chance in 3 of rolling either a '6' OR a '2'. So, if we want to find the probability of one event OR another occurring, we can simply ADD the individual probabilities. It is easy to see that the probability of either a head OR a tail when flipping a coin is 1/2 + 1/2, ie 1, a dead certainty of either a head OR a tail (neglecting the miniscule probability of the coin landing on its edge). Likewise, the probability of rolling a 1 or 2 or 3 or 4 or 5 or 6 when rolling a die is 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 - again a dead certainty that one of the six possibilities will result.

Now what about the scenario where we want to find the probability of multiple events occurring, for example if we roll a pair of dice and we want to get a '6' on both of them. In this case, when we want to know the probability of one event AND another occurring, we MULTIPLY together the individual probabilities, so to roll a '6' on two dice simultaneously (or one after the other for that matter), the probability is given by 1/6 x 1/6 ie 1/36 or 1 chance in 36 of rolling two sixes on a pair of dice.

In summary then:

- the probability or one event OR another is determined by ADDING the individual probabilities
- the probability of one event AND another is determined by MULTIPLYING the individual probabilities.

Cumulative Probabilities

Now, the likelihood of something actually happening, ie its "cumulative probability" depends not only upon its probability of occurrence as a one-off event, but also upon the number of times that we try. For example, the probability of rolling two sixes on a pair of dice *on a given attempt* is one chance in thirty-six. The more times we try though, the more likely we are to succeed. If we have two attempts, our cumulative probability (overall likelihood) of succeeding is 1/36 + 1/36 or 2/36, ie one chance in eighteen (1/18). If we try eighteen times, then our *overall* chances of success improve to one chance in two (or fifty-fifty). Any more attempts and we are more likely to succeed than to fail (although the probability of success *on any given attempt* remains at one chance in thirty-six). If we try dozens or hundreds of times then we are virtually certain to succeed. By this reasoning, it does indeed follow that no matter what the probability of life arising on its own, the prospects are improved if there are more suitable planets on which to "try" and if the Universe is older so that such planets have had time to make more "tries". However, the prospects are made worse if the probabilities of the individual events that would be required are so small as to offset the number of "tries" (in turn determined by the number of suitable planets and their age). For example, if the probability of life arising on a suitable world were one chance in a trillion (we shall see later that it is in fact much lower) and we searched more than a half-trillion worlds, then we would be more likely than not to find life. Our chances of success would outweigh our chances of failure. However, searching the same half-trillion worlds, if the probability of life arising on any one turns out to be only one chance in a trillion trillion (instead of the previously-assumed one chance in a trillion), then we would be a trillion times less likely to find life than to find it. So, we see that there are two opposing factors - the number of Earth-like worlds and their age increase our chances of finding life, while a low probability of life actually arising *on a given world* decrease our chances. We must quantitatively assess both of these opposing factors in any attempt to estimate the probability of finding extraterrestrial life.

The Drake Equation
An attempt to quantify the likely number of civilizations that might be able to communicate with us in our Galaxy (N), taking into account these opposing factors, was formulated by Frank Drake of Cornell University. This Drake Equation[3], as it is known, is expressed as follows:

$$N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$$

where ..

- $\mathbf{R^*}$ is the rate of star formation in the galaxy
- $\mathbf{f_p}$ is the fraction of stars which are accompanied by planets
- $\mathbf{n_e}$ is the average number of Earth-like planets in planetary systems
- $\mathbf{f_l}$ is the fraction of Earth-like worlds upon which life arises
- $\mathbf{f_i}$ is the fraction of those worlds upon which intelligent life arises
- $\mathbf{f_c}$ is the fraction of those intelligent species which learn to communicate across space
- **L** is the average lifetime of such a communicative species

Now, we know from recent discoveries that the first three factors in this equation are significantly high, making our prospects of finding other species initially seem quite good.

In the remainder of this article, we will focus upon the factor $\mathbf{f_i}$, which might easily offset a large value for the product of the first three factors if this factor itself turns out to be extremely small.

Information
At this point, it is necessary to say something about an entity known as "information". We live in a world today which is saturated with information, but what exactly is it ? One thing we can say is that it is tightly associated with probability. The more improbable an event or sequence of events is, the more information is required to specify it (less information is required to specify the result of flipping a coin than is required to specify the result of rolling a die or of winning the lottery). The more information something contains, the less likely it is to be generated at random. Well, this second statement isn't entirely true. To understand why, consider the following two strings of characters:

1. l2otj h39m klop hjew ;qa ltjk 0p-[kjgmja suirllpt x\$4=
2. In the beginning, God created the heavens and the Earth.

Both strings contain an equal number of characters (56 including spaces and punctuation) and thus the same amount of information is required to specify either. Each is equally improbable to arise from a random selection of characters, although I generated the first sequence from a random selection of characters. Each contains the same amount of information and each would require the same amount of storage in a digital memory. So why would we find the second string more remarkable than the first if it were generated at random, say by randomly throwing scrabble tiles onto the floor ? The difference between the two strings is that the second conforms to an independently given convention, namely the English Language, such that it conveys function or meaning. We say that the second string exhibits *functional* or *specified information*. The first sequence is equally unlikely to arise by chance, but it does not specify anything and thus would not be remarkable if it were generated by chance. So, we should correct the claim made above to say that "the more *specified information* something contains, the less likely it is to be generated at random". There are many more ways of arranging characters into stings which represent gibberish than there are ways of arranging them into meaningful sentences in English (or any other language for that matter). In the language of mathematical probabilities, the number of "possible outcomes" vastly outweighs the number of "desirable outcomes". Throwing a bucket of scrabble tiles out onto the ground is *overwhelmingly* more likely to result in a meaningless random arrangement such as the first string of characters above than it is to result in a meaningful sentence. This is understood intuitively, without the need for debate or argument. It is our intuitive and consistent experience that aperiodic functional or specified information never arises from random processes, but is invariably the product of conscious activity - indeed, the Search for Extraterrestrial Intelligence (SETI) is based upon this very premise. This intuition though seems clouded when it comes to seeing the unlikelihood of functional biological molecules arising spontaneously, so let us try to clarify the similarity.

Consider the building blocks of living cells, which are long organic molecules called proteins. Such molecules would have been essential for the first life on Earth before anything resembling a self-reproducing cell could arise which might subsequently be capable of Darwinian Evolution. These proteins are strings of smaller units called amino acids. We can think of the proteins as analogous to sentences, comprising long aperiodic (non-repetitive) strings of amino acids, which we can think of as analogous to letters. Now amino acids can be generated by known undirected chemistry, as demonstrated by Miller and Urey[4] in the 1950s, but it certainly does not follow that undirected processes could string together these amino acids into functional proteins. To infer such is similar to inferring that a machine capable of randomly generating scrabble tiles could also be expected to assemble them into meaningful sentences. The problem of creating the first functional proteins by random assemblies of amino acids is exactly the same problem as creating meaningful sentences by randomly stringing together alphabetic characters. In the same way that we understand intuitively that there are vastly more ways of stringing together alphabetic characters into gibberish than there are ways of stringing them together into meaningful sentences, it is equally true that there are vastly more ways of stringing together amino acids into polypeptide chains which serve no functional purpose than there are ways of stringing them together into functional proteins. Again, *specified information* is required to define the useful sequences. In living cells, this specified information is encoded onto long molecules called DNA. DNA molecules comprise a sugar-phosphate backbone, which takes the form of a double helix and a chain of nucleotide bases between the two helices. The sugar-phosphate backbone itself does not specify anything, but serves an essential function - we can think of it as the medium upon which the genetic information is "written". It is the nucleotide bases, or rather it is their sequential arrangement, that store the information as to how to string together amino acids into functional proteins. The bases are arranged into groups of three called "codons", where there is a defined "language" to transcribe the codons into corresponding amino acids in a protein under construction. This genetic language is now turning out to be different between different organisms, contrary to earlier understanding[9, 10]. The codons function exactly like alphabetic characters in a written language or digital characters in a computer code in that they represent prescriptive information and convey meaning in the form of precise assembly instructions. But how did this information arise ? Without it, a cell cannot build the proteins that it needs and cannot reproduce itself. Chemistry allows nucleotide bases to be arranged within the double helix in any order and allows amino acids to be strung together in any order into polypeptide chains - the information stored within each cannot be explained as a result of the chemistry or physical geometry of the constituent molecules in the same way that the meaning of words on a printed page cannot be explained as a result of the chemistry of ink bonding to paper.

We should also note that the information stored within the DNA and the molecular machinery required to read that information and translate it into amino acids and proteins are interdependent and thus must both have arisen together. To understand this point, I refer to a story told by a friend of mine about a cine film projector, which was supplied with a short film containing ... guess what ... instructions on how to load a film ! It is of course necessary to know the information contained in the film in order to be able to play it and subsequently reveal information that must already be known ! A somewhat analogous situation exists with the DNA, which contains the coded information on how to build the molecular machines in the cell, including those machines which are required to read the information stored within the DNA itself and to replicate the DNA strands - so, it appears that the machinery required to read the genetic information stored in DNA requires that information has already been used for its own construction !

Estimating Times for Success

Consider a combination lock, such as might be used to lock a bicycle and a persistent thief who is determined to steal that bicycle. He knows that on a four-dial lock, on which each dial has ten numbers, there are only 10 000 possible combinations (numbered 0000 thru 9999). Now if he can try one combination every 5 seconds, then he can try 5 000 combinations (half of the total number of possibilities) in 5000 x 5 seconds, ie about 7 hours. Thus to try all combinations and guarantee success, he would need about 14 hours, but we can say that after 7 hours he is more likely to have succeeded than to have failed. However, if he were to encounter a lock with ten dials instead of four, again each with ten positions, then there are now 10^{10} or 10 billion possible combinations and again trying one combination every 5 seconds would require over 1500 years to try all combinations ! In this case, then even if our thief were to devote his entire life to the task, it is clear that he would be overwhelmingly more likely to fail than to succeed.

In a similar way, if we know the probability of proteins forming by chance, we can estimate the time which would be required to be more likely than not for a random search through all possible sequences to succeed. Now there are 20 different types of amino acids used in proteins found in living cells - thus in our letters and sentences analogy, we have an alphabet of 20 letters. Most proteins are typically two or three hundred amino acids in length and their function is critically determined by the sequential arrangement of these amino acids. Let's consider building a shorter than average protein of just 150 amino acids in length by chance alone. We have one chance in 20 of selecting the correct amino acid for the start of our chain, one chance in 20 x 20 of getting the first two correct, 1 chance in 20 x 20 x 20 for the first three and so on. In order to get all 150 right on a single try, our probability of success is one chance in 20 to the 150th power, written as one chance in 20^{150}. This equates to one chance in 10^{195} or 1 chance in a thousand trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion trillion. However, this defines the probability of sponaneously forming *a specific* 150 unit long protein by chance. The odds of forming *any* 150-unit long protein which might conceivably fold into something which has any function have been estimated at slightly better than this - "only" 1 chance in 10^{164}, as suggested in the Illustra Media DVD "Origin"[5, 6].

But, as we have mentioned, our cumulative probability of success depends upon the number of tries, which in turn would be determined by the number of Earth-like worlds throughout the cosmos and their ages (ie, how long they have been trying). Now let us consider an imaginary Earth-like world, as represented by a thought experiment in the DVD "Origin" from Illustra Media[5, 6]. The oceans of this imaginary world are stocked to capacity with amino acids - all its carbon, nitrogen, oxygen, hydrogen and sulfur atoms are available to form 10^{41} complete sets of the 20 amino acids used to form proteins in living cells. Now we will suppose that these molecules are protected from damage due to ultraviolet radiation and chemical contamination. In this thought experiment, we will suppose that a chain of 150 randomly sequenced amino acids assembles in just one second. On our imaginary world, 6 million trillion trillion trillion attempts to form the target protein take place every minute. That means in 4.6 billion years (the oldest estimated age for the Earth), the number of attempts would exceed 10^{58}, assuming that every attempt produced a different combination, which there is no reason to suppose would be the case. We are still 10^{164} / 10^{58} (ie 10^{106} times or a "million googol" times) more likely to fail at producing a protein than to succeed, even given a planet with ideal conditions and an age similar to that of the Earth.

So just how long would it take to be more likely than not for our target protein to arise by a random undirected search ? The answer in years is so large a number as to be meaningless, so let's try to put it into perspective by means of another thought experiment also presented in "Origin"[7]...

Let us consider an amoeba, setting out to cross the observable universe, a distance of some 90 billion light years, traveling at a speed of one foot per year. Now, how far would our amoeba get before the odds became more likely than not that a functional protein would have formed on our hypothetical Earth-like world by chance ? Would it travel 10% of the distance across the Universe (at one foot per year) ? Yes. Would it make it half way across ? Again, yes. How about all the way across ? Definitely. How about all the way back again ? No problem. Now suppose that we let the amoeba travel across the Universe and back again, but this time carrying with it a single atom. It turns out that in order for our Earth-like planet to be more likely than not to produce a single functional protein by chance alone, our amoeba could travel across the Universe and back (at one foot per year), carrying with it each time a single atom, enough times to transport every atom on Earth, every atom in the Solar System, every atom in the entire Milky Way Galaxy and every atom in the known Universe more than 56 million times[5] ! That's how long it would take for chance alone to stand a reasonable chance of producing a single functional protein, even on a planet where the amino acid building blocks are abundant. It is OVERWHELMINGLY more likely that such a structure will NOT arise by chance than that it will. And that's not even a living cell - that's one isolated protein; a lifeless chain of amino acids, which by itself can serve no function anyway. The simplest living cell requires many hundreds of functionally interdependent proteins, which must all be formed in the same microscopic space, enclosed by a protective membrane (itself made of other specialized proteins). And the probability of that happening by chance has been estimated not at 1 chance in 10^{164} but about 1 chance in $10^{340000000}$ [11]. As any gambler knows, there comes a point when something becomes so overwhelmingly unlikely as to be considered impossible.

Conclusions

In answer to the question posed by the title of this article then, on the basis of the reasoning presented herein, I would submit the rather non-committal answer "probably". The reader should appreciate the tremendous amount of organization and functionally specified information that is required to build even a single protein from its constituent amino acid building blocks, never mind a living cell which might subsequently be capable of self-reproduction and evolution by natural selection. A planet with the right ingredients and conditions would be a *necessary* condition for life to arise, but clearly not a *sufficient* condition. Perhaps though, proteins and the first living cell could arise in a slow step-by-step process, perhaps not unlike natural selection. Of course, natural selection itself cannot occur until a self-reproducing cell already exists, but might a similar process of small incremental steps account for the origin of the first proteins and the first self-reproducing cell ? Richard Dawkins attempts to assign some plausibility to this suggestion by running a computer simulation in which he sets the goal of generating the English sentence “Methinks it’s like a weasel” by random selection of letters and then having the simulation select those sequences of letters, which by accident are included in the target sentence[8]. It can be shown quite easily that such a simulation does indeed converge toward the target sentence. However, the experiment proves nothing for two important reasons. Firstly, the simulation has a target sentence to which it is comparing the random variations – in other words, it is working toward a specific meaningful (future) condition. A simulation of processes proposed simulate natural selection should not select on the basis of attributes of a future state, which offer no function in the present. Secondly, proteins cannot be built in a slow step-by-step process like this anyway, except in the highly organized environment of a fully functioning cell. Partially completed chains of amino acids would serve no useful purpose and nature would have no reason to preserve them. Even a fully functional cell does not live indefinitely, but lasts for a limited time only and must reproduce itself within that time. Likewise, partially completed protein molecules will not "live" indefinitely and wait around for the next useful piece of the chain to be added. There's another problem too, which we have not mentioned up until now - Charles Darwin envisaged life originating in a "warm little pond", but it is now apparent that a water environment is actually not condusive to the formation of peptide bonds in proteins[5]. Proteins are actually made today in a highly structured production-line type environment *within the living cell*, controlled by DNA and the associated molecular machinery - the constituent parts do not easilly stick together in the environment of a "warm little pond". Conditions which we typically recognize as desireable for supporting life may in fact be hostile to its initial formation.

Perhaps the next time we read a news article implying that life may exist on some newly-discovered Earth-like exoplanet in the so-called "Goldilocks Zone" of its parent star, or on a moon of one of the planets in our solar system because it has the right environmental conditions and maybe even the right ingredients, we should reserve judgment and remember that whether its origin be undirected processes or divine creation, life from non-living chemicals requires much more than the right conditions and the right ingredients. It requires information. The search for life elsewhere and understanding the origin of terrestrial life will ultimately come down to a question of the origin of that information. Perhaps there are processes in nature of which we are currently unaware which can assemble amino acids into aperiodic functionally specified sequences, but attributing this result to random chance alone is perfectly absurd. Random processes are inadequate as a plausible explanation of the origin of even a single protein, never mind the structures found in life. Meaningful aperiodic information is never observed to arise from random processes. Granted, such statements do have implications toward theistic belief, but it is important to recognize that they are not *based upon* theistic belief - rather, they are based upon objective observation, our consistent experience of cause and effect and reasoning from mathematical probability. Even if the theistic implications are disregarded, this reasoning, as presented here also has profound implications for the question of the likelihood of finding life elsewhere if it were indeed derived from random processes. As in our combination lock analogy, we can never say that it is impossible to hit upon the right combination by chance, but the odds are *overwhelmingly* stacked against.

References

[1] "How to Build a Cell"
https://www.youtube.com/watch?v=VhZ0oIY88W8
(excerpt from Illustra Media DVD "Origin")

[2] "Unlocking the Mystery of Life"
https://www.youtube.com/watch?v=tzj8iXiVDT8

[3] The Drake Equation - Wikipedia
https://en.wikipedia.org/wiki/Drake_equation

[4] The Miller-Urey Experiment - Wikipedia
https://en.wikipedia.org/wiki/Miller%E2%80%93Urey_experiment

[5] The Mathematics of "Origin"
http://www.originthefilm.com/mathematics.php
http://www.neilbreakwell.net/pdf/MathOfOrigin.pdf

[6] "Origin"
http://www.originthefilm.com

[7] The Amoeba's Journey
https://www.youtube.com/watch?v=oxxs_kFfS-s

[8] The "Weasel Program"
https://en.wikipedia.org/wiki/Weasel_program
https://www.evolutionnews.org/2016/09/dawkinss_weasel/

[9] "What Happened to the Tree of Life" - Paul Nelson
https://www.youtube.com/watch?v=Mf01GXUl-mg

[10] "Darwin Was Wrong - Cutting Down The Tree Of Life" - (New Scientist - Jan. 21, 2009)
https://www.newscientist.com/issue/2692/
https://www.newscientist.com/article/mg20126923-000-editorial-uprooting-darwins-tree/

[11] Illustration of Probability of Protein or Cell Arising by Chance
https://www.youtube.com/watch?v=nuMvRExazAw