

Information, Intelligence, and the Origins of Life

Randy Isaac



Randy Isaac

The term “information” has a connotation of knowledge in the midst of ignorance, an order that arises amid disorder. Information exists everywhere around us, and we spend our lives acquiring, storing, transmitting, and processing it. Yet it is hard for us to define or describe it, in part because the word can be used in so many different ways. In this article, four main categories of usage of the word “information” are explored, paying specific attention to its relationship to intelligence. Thermodynamics includes information on all possible physical microstates; capacity of information refers to the maximum number of physical states possible in a system corresponding to pre-established conventions; syntax refers to the particular physical state of that system at a point in time; semantics are the meaning, function, or significance of that physical state. Living systems, in particular, are complex information systems. A look at how living cells process information provides some clues, but not yet a solution, to the mystery of the origins of life.

The explosive growth of information technology in the last several decades impresses on us the potency of information transfer. Lest we think this phenomenon is unique to our generation, we must recall that the ability to exchange symbolic information among individuals for collective learning is one of the crucial enablers of the development of humankind. Though the pace of change may have been slower, the generation, storage, transfer, and reception of information among intelligent agents have been the enablers of human civilization, if not of the very existence of our species. It is no wonder that we tend to view information as inextricably linked to intelligence. Today we often refer to our era as the “information era” and marvel at the ease of global information exchange through the internet.

The study of the concept of “information,” known as information theory, has moved into the arena of science and Christian faith largely because of its potential apologetic value. As biochemists unravel the secrets of information

processing within living cells, the similarities of those processes to information processing in our communication and computing systems becomes ever more intriguing. If such information processing could be shown to be necessarily related to intelligent sources, then we might establish a scientific inference toward an intelligent agent as a causal factor in the origin of life. For some, it is a small but obvious leap of faith to connect such an indeterminate intelligent agent with the Creator God whom we as Christians worship. It is essential

Randy Isaac was vice president of science and technology at the IBM Thomas J. Watson Research Center prior to retiring in 2005 and taking his current position as executive director of the ASA. He received his BS degree in physics from Wheaton College in Wheaton, Illinois, in 1972 and his MS and PhD degrees in physics from the University of Illinois at Urbana-Champaign in 1974 and 1977, respectively. In 1977 he joined the IBM Thomas J. Watson Research Center in Yorktown Heights, NY, as a research staff member in silicon technology. He was director of silicon technology before becoming project manager of the joint IBM/Siemens development program of the 64Mb DRAM product. In 1995 he was the founder and director of the IBM Austin Research Laboratory in Austin, Texas, which focuses on high-performance microprocessor design. As vice president of science and technology, he had worldwide responsibility for physical science research, including information theory, and for semiconductor, packaging, and communications technologies.

Article

Information, Intelligence, and the Origins of Life

to have a deep understanding of the nature of information to assess the value of such an apologetic.

Despite its prominence in our society, information continues to be a poorly understood concept. The term is used in so many different ways with so little precision, that confusion and misunderstanding abound. The intent of this article is to explore the various categories of meaning of the term “information,” to discuss the relationship of information to intelligence, and to consider the implications for our understanding of the origins of life.

What Is Information?

Our most common understanding of “information” is an idea, a concept, or an observation which we hold in our minds. But “information” is a much broader and more general concept. In his sweeping history of information, James Gleick attributes one of the earliest articulations of “information” to John Wilkins. He was a vicar and mathematician who later became master of Trinity College in Cambridge, and he was a founder of the Royal Society. In 1641 Wilkins wrote, “For in the general we must note, That whatever is capable of a competent Difference, perceptible to any Sense, may be a sufficient Means whereby to express the Cogitations.”¹ That is to say, information exists wherever something could be different; information does not exist where nothing could be different.

In the broadest possible sense, every elementary particle in the universe could be otherwise. Its properties, such as velocity or spin, could be different, or it could cease to exist or be transformed into energy. In this sense, there are estimates that all the particles in the universe comprise on the order of 10^{90} bits of information.² This type of usage of the term “information” is in a category that might be called *thermodynamics* since many of these properties are involved in thermodynamic considerations such as entropy.

To be useful in conveying conceptual information, it is necessary to restrict consideration to a subset of the vast thermodynamic category. Upon defining and selecting a specific convention for conveying information, a common usage of the term “information” is in the category of *capacity* or the number of bits available. This refers to the number of differences that are possible, such as the number of letters in the alphabet.

A third category of usage of the term “information” is the *syntax*, which is the specific selection made to convey the conceptual information. The term “information” can be used in various ways to explore the sequence in which the selected letters of the alphabet, for example, are arranged.

Finally, there is a category of *semantics* in which the usage of the term “information” refers to the meaning or function of the selected syntax.

When we wish to express or convey conceptual information, we embody it in a particular physical pattern, according to previously established conventions. These conventions could be, for example, the meaning of a sequence of sounds when we speak, or of the series of black shapes on a white background such as that you are now reading. This pattern is part of the category of information called the *syntax*, while the idea embodied by the pattern is in the category of *semantics*. To have meaning (semantics), the physical pattern that carries a specific piece of information must be drawn from a much larger number of possible physical patterns. If only one pattern were possible, it would not convey a distinguishing idea. The total set of all physical patterns that can be utilized for the embodiment of an idea is the *capacity* of information. The relationship among these categories is schematically illustrated in figure 1.

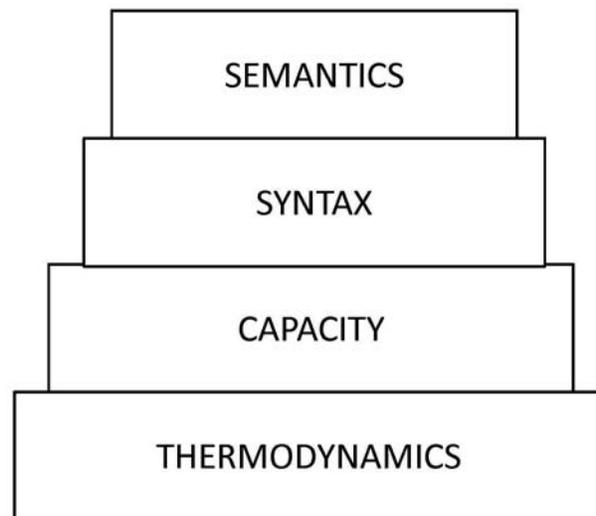


Figure 1. Categories of Information. Our most common concept of information involves semantics which attributes meaning to the syntax of a particular physical pattern, selected from a large capacity of possible patterns, which in turn are a subset of all thermodynamically possible physical states.

As an example, consider the information of the number of items in a group, say, 14. We can think of this number and retain it in our minds, which presumably are correlated in some way with our brain states. When we wish to convey this number to someone else, we express it in a physical pattern with any physical substance that can be shaped to resemble the form of the numerals 1 and 4 adjacent to each other. This conforms to established conventions for expressing numbers. There are alternatives, such as expressing the number in binary form so that the shape 1110 would be understood to signify the same number. This representation works if our convention has been adjusted to interpret the syntax in a binary format. The significance of a particular syntax depends on the set of possibilities that exist. The set of possible physical shapes can be thought of as the capacity of information. For two-digit Arabic numerals, the capacity would be 100, while for a 4-digit binary system, the capacity is 16.

In normal communication among humans, the primary message is conveyed using a common language, which is a convention of meaning assigned to specific physical patterns. Only a tiny fraction of the information available within our chosen medium (sounds, marks on a piece of paper, or electrical states on a silicon chip) is actually used. In the case of "marks made on a piece of paper," there is a much larger capacity (number of possible physical patterns) which is not useful in conveying a message simply because it is not part of our convention for expressing information. For example, variations between two styles of handwriting may not affect the primary message but might convey different information about the identity of the writer.

In our digital world, we find it convenient to express all information in terms of binary digits, or "bits" for short. To designate a bit of information, we need a physical feature that can have two possible states, 0 or 1, as in figure 2. In the terminology used above, the system in figure 2 has a capacity of one bit. In the case of a coin toss, the potential barrier is so high between "heads" and "tails" that no spontaneous transition can occur. In the case of an atom that could be in position 0 or 1, it is possible for thermal activation to occur between 0 and 1. The actual state of the system, whether 0 or 1, is the syntax.

The world around us is permeated with complex physical configurations which can, in principle, be expressed as a large collection of bits, as if figure 2 were replicated many times. Every particle or combination of particles can exist in more than one configuration with multiple variables that can have different values. The amount of information, I , is given by the logarithm of all possible states, N , that can exist: $I = \log_2 N$. The selection of which states to include in this equation depends on the context being used. For thermodynamic discussions of energy, entropy, and conservation laws, all possible microscopic states must be included. For the more common intent of conveying a message from a sender to a receiver, there must first be an established convention known by both the sender and the receiver.

In the case of a coin toss, a thermodynamic discussion of information might entail consideration of the atomic composition of the coins. This is irrelevant when the message is simply "heads" or "tails," in which case the number of states depends only on how many coins are used. The use of different coin types, such as pennies in addition to quarters, would change the result only if the convention in use involved coin types as well as "heads" or "tails." The amount, or capacity, of information therefore depends both on the communication convention being used and on the number of elements, such as coins, that are used.

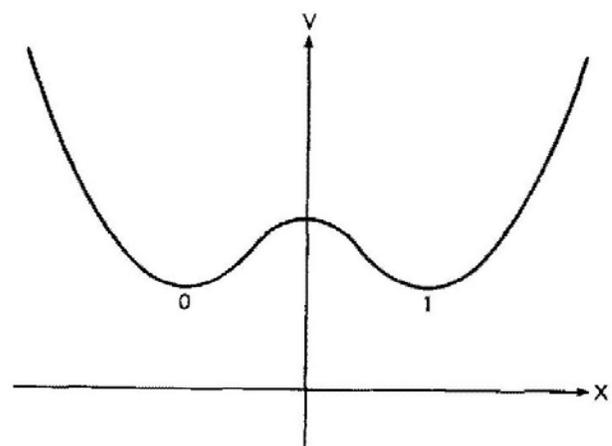


Figure 2. Bistable Potential Well. x is a generalized coordinate representing a quantity which is switched. (From Rolf Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development* 5, no. 3 (1961): 184, used with permission.)

Article

Information, Intelligence, and the Origins of Life

Another example may help elucidate the distinction among these categories. Consider the collection of ink molecules on paper comprising the words “red” and “blue.” The thermodynamic category of information encompasses all possible atomic and molecular states. One might extract information on the source or age of the ink, chemical properties of adhesion, the style of font, etc. The capacity of information in these words depends on the language that is chosen. For English, as opposed to say Chinese, the capacity is limited to an alphabet of 26 characters plus special symbols, or an established vocabulary of more than 100,000 words. The syntax category includes several types of analysis such as word order or grammar, encryption, or abbreviation, but always deals with the actual letters and words selected. The category of semantics deals with the meaning of the words. In various contexts, the meaning, in this case of “red” and “blue,” could refer to a particular wavelength of light, to an emotional state of mind, or to a political party inclination.

From this discussion, we can see that for communication purposes, the capacity, syntax, and semantics are all defined according to the convention known and accepted by the sender and the receiver.

Living organisms contain an immense amount of information in each of the categories listed. The sequence (the syntax) of all (the capacity) the nucleotide base pairs in the DNA molecules comprises coded genetic information that is translated into sequences of amino acids assembled into proteins. These proteins have physical and chemical functions (the semantics). A cell will survive only if these functions carry out the steps for metabolism, reproduction, etc. The information content (sequence of base pairs and/or amino acids) of a cell can and does change through a persistent series of natural reproduction events with change. For this reason, researchers studying the origins of life seek to determine whether such processes might be able to explain not only the continual transformation and development of the building blocks of life, but also the transition from nonlife to life.

This introduction to information has made it clear that information permeates the entire universe. Virtually all physical elements can be expressed in some form similar to figure 2. The capacity, syntax, and

semantics of information depend on the perspective of the sender and the receiver, be it an intelligent agent or a natural environment. We now turn to a more detailed discussion of each of these categories.

Information and Thermodynamics

In this section, we consider the usage of information in what we have called the thermodynamic category. Arguably, the most significant breakthrough in information theory was Rolf Landauer’s observation fifty years ago that energy must be expended to erase information. He showed that the energy required to erase one bit of information is at least $kT \ln 2$.³ Paul Gough points out that

Landauer’s principle applies to all systems in nature so that any system, temperature T , in which information is “erased” by some physical process will output $kT \ln 2$ of heat energy per bit “erased” with a corresponding increase in the information of the environment surrounding that system.⁴

To erase information, as opposed to changing the information, it is necessary to modify the potential wells in figure 2 so that only one state is possible. Paradoxically, there is no minimum energy requirement to generate information.

Information is, therefore, a fundamental physical parameter in the universe, related to energy and entropy. Entropy is a measure of uncertainty which increases as the number of possible states increases. Information is the reduction of uncertainty by the designation of one of those possible states. An increase in the number of possible states increases the uncertainty and consequently increases the amount of information when one of those states is selected. Information and entropy are therefore related and both tend to increase in a closed system. Information changes in a similar way as entropy, and can be transformed from one form to another, like energy, as the universe expands. This information includes all possible variables of all constituents in a closed thermodynamic system. Many, if not most, of these variables are not accessible to us for use in computing or communication or other information processing. For example, there are variables connected with the spin states of each individual electron, proton, or other elementary particles. Other variables are connected with the location of atoms relative to their

lattice sites in solid crystals, such as vacancies and interstitials. These are difficult, if not impossible, for us to detect and modify and store at a pace that is useful. Our discussion of information can be reasonably restricted to those variables that are information bearing, that is, those that are associated with distinguishable physical states that can be readily used to convey a message.

Restricted to these information-bearing variables, information is not inherently conserved. The erasure of a bit involves the transfer of information to a non-information-bearing degree of freedom, usually as energy dissipation to the surrounding environment.

Since information is a universal characteristic of all physical systems, there is no necessary relationship between information and intelligence. One could argue that intelligence is a particular method, but not the only one, of processing information. Natural processes continually transform information in our universe. However, the restriction of consideration to “useful” information-bearing variables is itself an action by intelligent agents and not by nature. That restriction is technology dependent. If, in the future, we were to invent a method of rapidly detecting and modifying the spin state and location of every atom in a solid, the information-processing potential would be extraordinary. Similar observations led Richard Feynman to exclaim more than fifty years ago, “There’s plenty of room at the bottom,”⁵ and there still is.

Our focus in this article is to understand information in order to determine its relationship to intelligence and to obtain clues to the origin of life. In the cosmological scheme of the universe, both life and what we consider to be intelligence seem to have appeared at least once, approximately 10 billion years after the big bang. Transformation of information appears to be a continuing universal process since the beginning of time.

Capacity of Information

At the heart of all information is its physical embodiment. Distinguishable physical states are necessary for information to be generated, stored, transmitted, and received. The capacity of information addresses the question, “How many bits are there?” and is simply the logarithm of the number of possible states.

Though we still deal with the legacy of information being expressed in base 10 or base 12, 24, 60, etc., it is the binary system, base 2, that dominates today’s information processing world. The unit of information is the “bit,” which is a contraction of “binary digit.” As noted above, the amount of information that can be expressed in any physical system is given by $I = \log_2 N$ where N is the number of distinguishable physical states.

Coin tosses are an easy example to illustrate this concept. Tossing four identical coins can result in sixteen distinguishable outcomes, leading to a bit capacity of 4, which we already knew since we had four coins, each of which can have two outcomes. A pair of dice is somewhat more complex since each die can have six outcomes. If the sequence of the dice is distinguished, then there are 36 possible outcomes or 5.17 bits.

Coins can also illustrate the importance of distinguishability. If the four coins mentioned above are all identical, say all quarters, then the order in which the coins are tossed is indistinguishable. If the coins are all different, say a quarter, a nickel, a dime, and a penny, then there are additional distinguishable outcomes. If the sequence is important, then there are 384 possible outcomes, or 8.6 bits of information. On the other hand, if all of the coins are identical and are perfectly smooth so that the two sides of the coins are indistinguishable, then only one outcome is possible and the bit capacity is zero.

Combinatoric information is a key subtype of capacity information that grows exponentially by the number of bits. For example, if each coin in a series of coin tosses is different and if the sequence is important, then the number of possible combinations is vast. Each possibility counts in the magnitude of capacity.

In computer logic and memory applications, physical states are designed for density, speed, and power efficiency in storing and processing bits of information. Typically, a node of a circuit can be held at either a voltage of 0 or of the supply voltage V . Either one can be arbitrarily assigned the symbol “0” and the other is assigned a “1.” With specified constraints on the physical states and their interaction, computers can be designed to generate, process, store, retrieve, and transmit vast amounts of information. Capacity of information is familiar

Article

Information, Intelligence, and the Origins of Life

to us as the capacity of a hard drive (e.g., 250GB) or of computer memory (e.g., 4GB). These values are independent of what, if any, messages are actually stored on those devices.

Communication technology has also grown exponentially, allowing bits to be transmitted at rates that were scarcely dreamed of only a few decades ago. Photons guided through optical fibers are the dominant physical mode of information transfer in our internet world. These photons are constrained according to specifications established by the communication designers. Claude Shannon of Bell Labs wrote the seminal paper on information communication in 1948,⁶ showing how to determine the capacity of information that could be transmitted in a noisy channel.

Distinguishable physical states can be established either through natural causes or by intelligent agents. It is not sufficient to observe distinguishable states to determine evidence of an intelligent source. However, the constraint that these physically distinguishable states must be easily detected, modified, and transmitted puts a significant limitation on what constitutes useful information. It is almost always the case that information useful for intelligent agents involves physical states established by those agents. The clearest way to ascertain an intelligent source is whether the physical states in question conform to the constraints imposed by an intelligent source. In other words, if the physical states meet criteria established by intelligent agents, then the source of those physical states is most likely, though not necessarily, an intelligent agent. The linkage between information and intelligence is derived not from the fact that the physical states represent information, but that they conform to the constraints imposed by the intelligent agent. The connection between information and intelligence is derived from the intelligent source and not from the information per se.

Applying these considerations to a living cell, we can detect a number of information-bearing variables. The best-known one is the DNA molecule, called the genome, containing a sequence of nucleotide base-pairs. There are other information-bearing components in the epigenetic system, and it is possible, even likely, that more such variables will be discovered in the future. For convenience we will focus on the DNA sequence, while recognizing that many other aspects of information may be present.

The genome has a vast capacity for information because of the nature of the physically distinguishable states. Each site along the nuclear DNA can have one of four distinguishable nucleotides. With approximately 3.5 billion sites in the human genome inherited from each parent, the bit capacity is an incredible 7 billion bits while the number of possible combinatoric outcomes is an inconceivable $10^{2,100,000,000}$. In combination with a second copy inherited from the other parent and a large variety of epigenetic factors that influence which genes are expressed to what degree, the information capacity is beyond comprehension.

Genome sequencing in the past decade indicates that only a small fraction of the genome actually codes for genes, and a very large portion of the genome has no apparent function. The capacity for useful combinations of the nucleotide base-pairs that do serve as codons is still so vast that the number of possibilities is countless. This capacity can be modified, either increased or decreased, by numerous mechanisms, ranging from single nucleotide insertion or deletion to relocation or duplication of a large segment of DNA.

Syntax of Information

Another category of usage of the term “information” relates to the actual state, out of all the possible states, in which the system exists. This usage addresses some variation of the question, “What are the bits?” The previous category of capacity was independent of the actual value of any bit, whereas this category deals with the values and relationship of values among the various bits. It basically considers whether any particular bit is a “0” or a “1” and the relative relationship among all of the bits.

Consider again the tossing of four identical coins. The capacity of information is always 4 bits, no matter what the outcome. Syntax is concerned with whether those coin tosses are heads or tails and the relationship between the results of the various coins.

If the outcome of four coin tosses results in all heads, the relationship of the values of the various bits attracts attention. The probability of that outcome is 1 in 16, no different than that of any other particular outcome such as 3 heads and 1 tail. But the outcome is noteworthy because we recognize

a specific relationship among those values. If the number of coins were very large, we would be justified in suspecting a process other than pure random coin tosses. Our clue would be more than the low probability of occurrence of that pattern. It also notes that the pattern of results matches an a priori relationship established by intelligent agents.

In the case of coins, we understand the process of tossing coins, and we can therefore assess probabilities of particular outcomes with reasonably high accuracy. In each toss, the history of previous tosses is effectively erased and has no bearing on the outcome. However, in many cases, the particular physical state is a function of a series of past events, essentially a contingent-history syntax. For example, a sample of rock studied by a geologist would have an atomic concentration, or syntax, that depends on its history. A well-known example is the ratio of radioactive elements in that sample. Understanding the probabilities of any particular concentration depends on a clear knowledge of the process steps that can modify such information over time. In general, when the particular state of information can change over time, an attempt to calculate the probability of occurrence of that state requires detailed understanding of all the various ways in which it could arise. Coupled with the knowledge of possible changes, an information state can lead to a deeper level of information about its history and origin.

We can also note that if all coins are heads, the information content, from the syntactical perspective, is smaller than if there is a mixture of heads and tails. It is of considerable interest to mathematicians and engineers to find algorithms that can express the values of a large number of bits with a much smaller number of bits. The mathematical elegance that can result has been explored by Kolmogorov and Chaitin and the result is known as Kolmogorov-Chaitin information, sometimes referred to as algorithmic entropy or descriptive complexity. This addresses the question of "What is the minimum number of bits required to express a given sequence of bits?" An information system can be called complex if a pattern of bits cannot be expressed algorithmically in a much smaller number of bits. Paradoxically, in this sense, a purely random sequence would be considered to have the maximum information, while a highly ordered sequence would have less information.

Engineers are interested in this category of information to achieve efficient compression techniques. Reducing the number of bits required to describe the actual sequence of bits is a valuable tool to reduce information capacity requirements as well as data transmission times. Video transmission in particular relies on compression where the action is slow or portions of the image are identical.

In a living cell, the syntax is primarily about the sequence of base pairs in the nuclear DNA. That sequence can be seen to have a small probability of changing during a reproduction event or during external stimulation such as radiation. These changes can occur as point mutations or as larger-scale shifting of DNA segments, such as gene duplication or transposons which are rearranged in the genome.

Semantics of Information

The category of meaning of the term "information" that we use the most often is semantics. This category addresses the question "What do the bits mean?" Our primary concept of information is the message that the bits are intended to convey. Paul Revere famously used two lanterns to indicate a powerful message, reducing the British means of transportation to a signal conveyed by one or two lanterns. The bit capacity was small but the semantic meaning was profound.

Information theory does not address semantics. Shannon explicitly excluded meaning from his consideration. James Gleick quotes Shannon as writing,

Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.⁷

Semantic information is not quantifiable in the sense that capacity or syntax can be defined. Mathematical formulations may indicate what physical configurations are useful and might have a meaning in certain circumstances, but do not express the meaning itself.

The semantic meaning may nevertheless be important in determining the capacity, for example, of the information channel. Shannon showed how information is inversely proportional to the probability of occurrence. Accordingly, knowledge of the

Article

Information, Intelligence, and the Origins of Life

frequency of occurrence of a letter of the alphabet or of a combination of letters, or of a word, can be used to determine the probability and thereby optimize the capacity of a communication channel. The semantics of the English language influences the usage, which can be measured and used to optimize capacity. But the meaning itself is not part of the information-engineering calculation.

For some, the term “semantics” assumes the presence of an intelligent agent as a sender and as a receiver of the message.⁸ In this article, the term is used more broadly to indicate the significance of a message, whether or not an intelligent agent is involved. It includes the possibility that the message is a physical effect, a causal factor for a physical or chemical action.

Symbols, commonly in the syntax category of information, are physical representations of meaning. In physical symbolism, the symbol has a physical property which serves as the message. For example, a shiny, smooth metal surface can serve as a symbol of high reflectivity. Or a north pole of a magnet serves as a physical symbol of attraction to a south pole magnet. The meaning or significance of a physical symbol is derived from the physical properties of the symbol itself.

In abstract symbolism, the symbol has a meaning assigned to it which does not necessarily derive from its physical properties. For example, the meaning of the shape of the letter “A” in the English language is assigned to that shape and does not derive from the shape itself. Paul Revere’s message was not derived from the number of lanterns but was assigned to it. Anyone intercepting the message had no way of decoding the message from the physical characteristics of the lanterns without acquiring the knowledge of the abstract relationship assigned by the sender.

Abstract symbolism is a hallmark of intelligence, especially as manifest in language and communication techniques. The ability to associate abstract symbolic significance with a distinguishable physical pattern is a key indicator of intelligence, though not the only factor. Primatologists look for signs of such ability in order to assess the degree of intelligence in primates, for example. Abstract relationships are so important in our daily lives that we often take them for granted. All of our communication technology, computing technology, mathematics,

and virtually any activity involve some degree of abstract thinking. This is a key feature that links intelligence with information.

When abstract relationships are a necessary part of information systems, then an intelligent agent must be involved to generate or interpret or design that system. In computer technology, for example, the criterion for verifying proper design involves testing the output for the right answer. If 2 plus 2 produces an answer of 5, then the physical connections from the input to the output produce an answer that correctly reflects the actual design of the logic components. But an agent with knowledge of arithmetic must be involved to determine whether such connections meet the desired design. It may not be possible to determine if the answer is correct solely from the physical connections themselves. If 2 plus 2 is 4, then the computer meets the test of our abstract concept of arithmetic and the design is pronounced to be correct.

A communication system is tested by comparing the message received with the message intended to be transmitted by the sender. That abstract relationship means that an intelligent agent must be involved in setting up the communication system. A physical test could determine whether the same syntax exists in the received message as the sent message, but an agent would need to decipher any abstract meaning.

For living cells, significance seems to be all physical and chemical. There appears to be no abstract meaning assigned in the operation of the cell. Even the coding of a base pair sequence that translates into a sequence of amino acids to produce a protein is a chemical process and not an abstract one. We can generate an abstract coding table (a “look-up table,” relating any given codon to a corresponding amino acid sequence) to describe what is happening, but the actual translation event occurs physically, independent of the influence of any intelligent agent. A more detailed discussion of the nature of the biochemical information processing in living cells is provided by Jonathan Watts and by Stephen Freeland in other articles in this issue.⁹

We now turn to a closer examination of the information contained in living cells to see what other clues there may be that pertain to the origin of life.

Clues to Life's Origins

Where there is no change, there is no history. Seeking the origins of life involves sifting through the patterns of change in living systems that might provide evidence of the kinds of changes that may have given rise to life. The detailed answer of how life began may never be fully known, but a study of the information in living cells provides tantalizing leads to plausible scenarios.

William Dembski has claimed that information is conserved and can only be generated by intelligent beings.¹⁰ Recognizing that this is not true of all information, he considers what subset of information obeys this type of conservation law and whether DNA information is of that type. Dembski focuses on complex specified information (CSI), a term attributed first to Leslie Orgel,¹¹ as being that subset. The term "complex" refers not only to a large capacity but also to a syntax that is not reducible to a much simpler equivalent formulation. Specificity is essentially the functionality or meaning of the syntax of that information. Specificity does not lend itself to mathematical formulation and is part of the semantic category that is not addressed in the field of information theory as noted above.

Stephen Meyer expands on the concept of CSI and shows how DNA information is part of that subset.¹² He shows that specificity can include functionality as well as meaning and that DNA information is specified information because of its functionality. He then asserts that CSI is habitually generated by intelligent sources and, therefore, the genetic code must have been as well.¹³

The primary objection to this assertion is empirical. Observation of biochemical systems shows that while DNA information meets the definition of complex specificity, new CSI is also generated without involvement of intelligent agents. One example is provided by Craig Story in his discussion of the immune system and the generation of cells that produce antibodies in response to antigens.¹⁴ An original population of cells with identical nuclear DNA produces a population of lymphocytes that have a novel sequence of base pairs in a particular subset of their DNA and which produce antibodies that have high affinity to the antigens. This constitutes specificity through the functioning of the antibody. New CSI information is generated, without involvement

of an intelligent agent, in the production of useful antibodies. Other examples are given by Watts in this issue.¹⁵

In a much broader sense, we observe that the offspring of virtually all sexually reproducing species have a DNA sequence that is similar but new compared to their ancestors. The functionality that meets the criterion for specificity is clear in the survival of the offspring and is subtly different from that in the parents. We can therefore see that the conservation law of CSI does not hold for biological systems and is not universally applicable.

Meyer's argument also falls short theoretically of being compelling. Meyer uses only inductive reasoning, claiming that all known abiotic examples of CSI require an intelligent source, and extrapolates that, therefore, nonliving systems cannot generate life. He points to similarities in examples such as computer programming, language texts, and phone numbers which inherently require an intelligent source. These analogies, while intriguing, are hardly conclusive. Meyer does not present a characteristic of CSI that is necessarily related to intelligent agents.

One possibility that could relate intelligent agents to a subset of CSI is abstract symbolism. With the ability to carry out abstract reasoning as a trait uniquely attributed to intelligent agents, it would follow that abstract specificity would therefore require intelligent agents. Unfortunately, Meyer does not pursue the distinction between physical and abstract specificity. Since the functionality of DNA information resides in its physical-chemical action, no abstract specificity is evident in a living cell.

This discussion still leaves open the possibility that even if biological evolution involves an increase of CSI without intelligent agents, perhaps chemical evolution is restricted. Nonliving information systems are vastly simpler than living systems, and information, even useful information, can be generated without intelligent agents. But could chemical evolution occur? Is it possible for a nonbiological system to increase CSI to the point of becoming a living biosystem? No one has offered a compelling answer to this question. It is the heart of research in the origin of life and is discussed further by Freeland elsewhere in this issue.¹⁶

Article

Information, Intelligence, and the Origins of Life

Information theory does not seem to provide any basis for claiming that such chemical evolution could not happen. A physical information system can be generated from a prior system with less (or more or different) information, corresponding to the thermodynamic, capacity, or syntax categories. Whether such systems can have meaningful semantics is not within the purview of information theory. If there is clear evidence of new abstract specificity, it is reasonable to infer that an intelligent agent was involved. If there is no such clear evidence and only physical symbolism is evident, then such involvement cannot be inferred.

Does a living cell exhibit any form of abstract symbolism? Considering the details of any living cell, the criterion for significance is functionality and contribution toward survival of the cell, usually shown by the cell's ability to reproduce itself. This is a physical criterion that includes no connection to an abstract relationship. Though the existence of information and its structure are fascinating and interesting, particularly in the similarities to information-handling techniques humans have devised in recent decades, no feature of the information content inherently requires an intelligent source. We must take a closer look at the information in order to determine how to invest research activity into the origins of life.

We first note that, from a thermodynamics perspective, living cells are dynamic, open systems that continually exchange energy, entropy, and information with their surrounding environment. For multicellular organisms, that environment is, first of all, a vast collection of cells with nearly identical nuclear DNA, while single-celled organisms interact directly with their ecological system. For example, mitochondria (organelles within most eukaryotic cells) act as power sources that convert a variety of fuels from the environment into usable energy. Thus there is plenty of opportunity for information to be transformed from one variable to another, from various physical states to useful information-bearing variables. Information in a cell is not conserved, just as entropy is not conserved in an open system.

The capacity for information in living cells, as noted earlier, is immense. The sequence of nucleotide bases along the nuclear DNA is the best known, but other variables, such as receptors for various bio-

chemical molecules, can also bear key elements of information. The number of distinguishable physical states possible is not only inconceivably large but it can change as, for example, the length of the DNA increases or decreases. For complex eukaryotic organisms, the capacity for information can change considerably during reproduction through a variety of processes such as gene duplication. In humans, for example, genomic studies indicate that there are approximately 10 to 50 major changes, increases or decreases, in the number of genetic sites between parent and child, with some as large as a million base pairs.¹⁷ Many of these are copy number variants of genes or transposons that have been moved to another region of the DNA. Even larger changes can be seen in terms of chromosomal rearrangements or extra copies of entire chromosomes. This is still miniscule compared to the total number of base pairs in the human genome, but the principle is clear. The DNA information capacity of a cell or organism can and does change through the natural process of reproduction.

The syntax of the DNA information in living systems provides the most intriguing insight into life's origins. Whole organism genomic sequencing has become not only possible but also affordable in the last three decades, opening a treasure trove of insight into the information contained in living cells. Since any particular sequence of DNA is derived from a very similar yet different DNA sequence, the syntax is strongly historical-contingent. A given sequence occurs as a result of a long history of changes. Without a clear understanding of all possible historical paths, no credible probability of occurrence can be determined. Irreducibility, the term used to describe a sequence that could not be derived from any other smaller sequence, cannot be compellingly demonstrated simply due to the vastness of the possible historical pathways. Walter Bradley provides a fairly rigorous treatment of information and entropy but fails to recognize that probabilities and improbabilities cannot be reliably assessed unless all historical pathways and processes are well understood.¹⁸

The semantics of DNA information is the subject of many courses in biochemistry. The significance of the information is the biochemical function that is carried out. The genetic coding is translated in ribosomes into chains of amino acids that form proteins which fold in unique ways to carry out elaborate

functions that contribute to the survival of the organism. The term “genotype” is used to refer to the syntactical information in the portion of the genome that codes for genes. The term “phenotype” is used to refer to the semantic information, or function, of those genes. What concerns us here is that all of these functions are physical or chemical processes without evidence of an abstract symbolic value. Coding in and of itself does not necessitate intelligence unless the coding represents abstract symbolic meaning.

Two primary conclusions can be drawn from detailed studies of genomic sequences. The first conclusion of note is common ancestry of all organisms. Charles Darwin and Alfred Russel Wallace drew on their detailed observation of many species to conclude that all organisms may have descended from a common living form. Using techniques these naturalists could never have imagined, geneticists can now examine the sequences of base pairs in DNA to determine inheritance. Going far beyond paternity suits, the patterns of similarity and differences of DNA sequences reveal information about family ties that go back billions of years. The evidence continually grows stronger: all species seem to have derived from a common source rather than have independent origins.¹⁹ This is a major clue which sharpens our research into life’s origins to the genesis of a simple life form in a primordial environment. It confirms the historical path of incremental changes of DNA information.

The second conclusion is derived from observations about the location where DNA information changes. Comparing the genomes of various individuals within a species as well as with those in other species, it is clear that DNA regions that code for some critical genes change at a far slower rate compared to regions whose function is less critical. This is a consequence of natural selection. If a change occurs in a function necessary for life, the organism will not survive. Those changes will not be seen. Changes in less critical regions of the genome will have no or negligible impact on survival, and these changes may persist. Some of the changes might be beneficial for survival and be adopted rapidly in the population.

While neither chance alone nor deterministic necessity can lead to the diversity of information required for life, the combination of chance and

necessity is a powerful method of designing the proper building blocks of life. The signature we find in the syntax of information in living cells is a process of natural selection which is powerful in enabling efficient derivation of functional configurations. We do not yet know what kind of system could have preceded and generated an initial RNA complex that might have initiated biological evolution. It is fair to extrapolate that processes analogous to reproduction with variation and natural selection, which explain the development of species, may account for such an origin of life from nonbiological sources. No principle from information theory precludes such a scenario. Discoveries in the past few decades of autocatalytic processes, self-assembly, and other analogous processes, give an indication that this research is moving in the right direction.

Though the mysteries of life’s origins have not yet been solved, it seems reasonable to conclude that the inference to the best explanation is not an indeterminate intelligent agent but processes akin to reproduction with variation and natural selection. As Christians, we have faith in the existence of an Intelligent Designer who utilizes the design tools of these natural processes to carry out his creative intent. ↘

Acknowledgment

The author is deeply grateful to the reviewers and many friends and colleagues who provided substantial constructive feedback and ideas for expressing the concepts in this paper.

Notes

¹James Gleick, *The Information: A History, a Theory, a Flood* (New York: Pantheon Books, 2011), 157.

²*Ibid.*, 397.

³Rolf Landauer, “Irreversibility and Heat Generation in the Computing Process,” *IBM Journal of Research and Development* 5, no. 3 (1961): 183–91.

⁴M. Paul Gough, “Information Equation of State,” *Entropy* 10 (2008): 150–9, doi:10.3390/entropy-e10030150.

⁵Richard Feynman, “There’s Plenty of Room at the Bottom,” *Caltech Engineering and Science* 23, no. 5 (February 1960): 22–36, <http://www.zyvex.com/nanotech/feynman.html>.

⁶Claude E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal* 27 (1948): 379–423, 623–56.

⁷Gleick, *The Information: A History, a Theory, a Flood*, 222.

⁸Keith Ward, “God as the Ultimate Informational Principle,” in *Information and the Nature of Reality*, ed. Paul Davies and Niels Henrik Gregersen (New York: Cambridge University Press, 2010), 282–300.

Article

Information, Intelligence, and the Origins of Life

⁹Jonathan K. Watts, "Biological Information, Molecular Structure, and the Origins Debate," *Perspectives on Science and Christian Faith* 63, no. 4 (2011): 231-9; Stephen Freeland, "The Evolutionary Origins of Genetic Information," *Perspectives on Science and Christian Faith* 63, no. 4 (2011): 240-54.

¹⁰William A. Dembski, *The Design Inference: Eliminating Chance through Small Probabilities* (New York: Cambridge University Press, 1998).

¹¹Leslie Orgel, *The Origins of Life: Molecules and Natural Selection* (New York: John Wiley, 1973).

¹²Stephen C. Meyer, *Signature in the Cell* (New York: HarperOne, 2009).

¹³*Ibid.*, 346.

¹⁴Craig Story, "The G.O.D of Immunology," *Perspectives on Science and Christian Faith* 61, no. 4 (2009): 221-32.

¹⁵Watts, "Biological Information, Molecular Structure, and the Origins Debate."

¹⁶Freeland, "The Origin of Genetic Information."

¹⁷Peter H. Sudmant et al., "Diversity of Human Copy Number Variation and Multicopy Genes," *Science* 330 (2010): 641-6.

¹⁸Walter Bradley, "Information, Entropy, and the Origin of Life," in *Debating Design*, ed. Michael Ruse and William A. Dembski (New York: Cambridge University Press, 2004), 331-51.

¹⁹Douglas L. Theobald, "A Formal Test of the Theory of Universal Common Ancestry," *Nature* 465 (2010): 219-23.

The American Scientific Affiliation
seeks submissions for its first annual

STUDENT WRITING COMPETITION

for Poetry, Fiction, and Creative Nonfiction about

Science
&
Faith

Why?
How?
Who are we?

1st
Prize:
\$200

Contest is open to graduate and undergraduate students.
Deadline: January 31, 2012
Visit www.asa3.org for details.