

Biological Information — What is It?

Werner Gitt^{1*}, Robert Compton² and Jorge Fernandez³

*¹German Federal Institute of Physics and Technology, Braunschweig, Germany. Former Head of Information Technology, Director and Professor (retired). ²Department of Veterinary Anatomy and Physiology, College of Veterinary Medicine, Washington State University, Pullman, Washington (retired). ³President, Mastek Inc., Melbourne, Florida. (*Corresponding author: werner.gitt@t-online.de)*

Abstract

Scientific discoveries, especially over the last six decades, have left no doubt that ‘information’ plays a central role in biology. Specialists have thus sought to study the information in biological systems using the same definitions of information as have been traditionally used in engineering, computer science, mathematics and in other disciplines. Unfortunately, all of these traditional definitions lack aspects that even non-specialists recognize as being essential attributes of information — qualities such as meaning and purpose. To remedy that deficiency, we define another type of information — Universal Information — that more accurately embodies the full measure of information. We then examine the DNA/RNA protein synthesizing system with this definition of Universal Information and conclude that Universal Information is indeed present and that it is essential for all biological life. Furthermore, other types of information, such as Mental Imaging Information, also play a key role in life. It thus seems inevitable that the biological sciences (and science in general) must consider other-than-the-traditional definitions of information if we are to answer some of the fundamental questions about life.

Key words: information, codes, Universal Information, biological information, scientific laws, laws of nature, transmitter, receiver

Introduction

The title of this symposium is “Biological Information: New Perspectives”. But what do we mean by the term “biological information”? We suggest that, at present, it cannot be unambiguously defined. Yet, an unambiguous definition would be extremely helpful because multiple levels of communication systems are being researched: from the DNA-coded information in the genome, to the intra-cellular communication networks involving RNA and proteins, to inter-cellular signaling via entities such as hormones, all the way up to and including the nervous system and the brain. Clearly, identifying all of these communication systems and defining the information that is being transferred will be a challenge. It is clear that

there are many subsets (or categories) of biological information, and that many more will be discovered [1]. At some point, perhaps after further research, we will be in a position to more precisely define “biological information”. For the interim, we will offer a description of it as a placeholder until we have enough knowledge to define it with scientific rigor.

We propose that biological information includes all manifestations of information in living organisms. This description has the potential to include all categories of information. From recent scientific studies in genetics it is clear that there are many subsets of biological information (codes), and many more wait to be uncovered within the DNA/RNA systems alone [1]. It is reasonable to believe that progress on biological information will be accelerated if each information subset is unambiguously defined. Towards this goal we should begin by defining a definition: *a description or explanation of a word or thing by its attributes, properties or relations that distinguishes it from all other entities* [2]. Even applying this definition carefully is important, because scientifically rigorous results cannot be achieved when using ambiguous terms. A common example of this is the claim that, “Evolution is a fact.” The validity of such a claim is certainly going to depend on the precise meaning of the term ‘evolution’.

Defining Subsets of Information

This leads us to ask the more general question: *What precisely is information?* Anyone who has studied this field is aware of three working definitions of information:

- **Classical Information Theory:** Shannon (statistical) information [3]; dealing solely with the technical/engineering aspects of communication. This involves analyses including obtaining statistics on the material symbols for data transmission, storage and processing.
- **Algorithmic Information Theory:** Solomonoff/Kolmogorov/Chaitin [4–6]; dealing with the ‘complexity’ (as this term is defined in the theory) of material symbols in data structures and of objects in general.
- **Complex Specified Information (CSI) Theory:** Dembski [7]; roughly the same as Classical Information Theory but adding the important concept of a ‘specification’.

These theories, like modern genetics, focus primarily on the material carriers of the information. On the other hand, American mathematician and National Medal

of Science recipient Norbert Wiener in 1968, made his often quoted statement “*Information is information, neither matter nor energy.*”[8]

Wiener’s statement prompted one of us (W. Gitt) to ask; if information is not matter (mass and energy) then what is it? Gitt therefore started a long quest to define information — at least the information that was most familiar to scientists of that day. As an information scientist, Gitt not only examined the information conveyed within human natural languages, but also the information conveyed within abstract and artificial languages such as machine languages. In his studies he identified five attributes, four of which qualified as *distinguishing* attributes of ‘information’. Before we examine these, let us make it clear that these natural and artificial language systems were first studied because at that time they had already been extensively characterized. We used these human information systems as ‘known systems’, which would most likely be amenable to precise definition.

Distinguishing Attributes of Information

Code plus syntax

At the basic level of information in these languages we find a set of abstract symbols formally known as an alphabet — this set constitutes the code. By abstract we mean that each of these symbols has no resemblance and no inherent physical relationship to the entity that they represent. These symbols have a characteristic two-dimensional configuration that distinguishes them from each other. One way in which this is manifested in the material domain is by inscribing these symbols onto a wide variety of material media and formats.

Examples of this are abundant. For instance, the first five words of Lincoln’s Gettysburg Address — “Fourscore and seven years ago” — may be inscribed on paper with ink, or chiseled onto a block of granite, or on a blackboard with chalk, or in the air with smoke signals or with the vibrations of speech, or on a transmission line with electrical dots and dashes as in Morse code, or on a computer’s hard drive by properly setting magnetic ‘bits’, or in many other ways.

With this we see that the actual information is completely independent of the material medium that serves only to ‘carry’ it. Any one of a multitude of material media and formats may be used to carry exactly the same information. While there is indeed a *correlation* between the material media and format that carries the information, the dictum “correlation does not imply causation” certainly applies here. The material carrier cannot be and is not the *cause* of the information.

Further examination reveals that there is also a set of rules governing what is permissible regarding the arrangements of the symbols — this set of rules constitutes the syntax.

With the combination of abstract code and syntax we are able to generate more complex language structures such as words and sentences. However, at this (formal language) stage *meaning* plays no role. It was at this level *only* that Shannon developed his *Theory of Communication* [3] into the highly useful statistical analyses of the material symbols, solely for the technical purposes of data transmission, storage and processing. Code plus syntax is a necessary distinguishing attribute of all human languages. Let us pause for a moment and reflect on how this all comes about.

In order to develop, learn, or use a code plus syntax system, it requires a high degree of mental effort and intelligence. No one at any time has ever observed this basic attribute of information (i.e., code plus syntax) being established through unguided, purely physicochemical processes. However, we have observed young children learning the alphabet and learning to read, write, and speak words. Also many of us as adults have developed and/or learned machine languages. We may say that people acquired these abilities from their parents and so on down through history. However, this does not in any way explain how the *first* human acquired this ability. If we assume this happened without intelligent guidance, there are only two alternatives: 1) it is an inherent property of matter or, 2) it is possible for these abilities to ‘evolve’ over time. A person may choose to believe in either of these alternatives but that person would have to also accept that this is a *belief* with no hard science to support it.

Meaning

The next level of the distinguishing attributes of information in human languages is meaning. At this level, words that were formed by short sequences of symbols are assigned to represent ‘something’ (where that ‘something’ may be any particular entity or object, event, thought or concept). Additionally, that ‘something’ must be defined and that definition is also represented by words.¹ Higher levels of meaning and information content are constructed using phrases, sentences and paragraphs

¹For example, consider the word ‘cat’. ‘Cat’ is an abstract representation of the actual creature. If we then define this creature as ‘a four-legged mammal that purrs and meows’, we then have other words that are being used to represent both the word ‘cat’ and the creature. We note that the words being used also be defined with other words — this goes on level after level. For instance, the above definition for ‘cat’ included the word ‘mammal’. That word ‘mammal’ must be defined (with other words, of course) and then those other words will in turn need to be defined. Thus, a measure of circularity is ultimately unavoidable.

when the meaning from one word is insufficient. In our example of a ‘cat’ (footnote 1), its definition is a sentence that represents a creature and the word ‘cat’.

Meaning is an absolutely essential attribute of information that is conveyed in language and communication. Words, both written and spoken, can be used to represent entities, events and/or concepts — literally anything. The entities need not be present but words, serving as their placeholders, represent and thereby communicate their reality as if they were present. Unguided, purely physicochemical processes have never been observed creating this ‘substitutionary’ process [10]. We are referring here to natural, unguided, purely physicochemical processes that have no external guiding (control) systems found in information-rich systems. These seem to eliminate all biological information systems as being examples of unguided, purely physicochemical processes.

Expected Action

The examination of sentences or paragraphs in a message reveals an implied request or a command for the receiver of the message to perform some action. These actions start with the receiver reading and understanding the message (this in itself involves very complex actions). From understanding the message, the receiver must decide whether or not he/she will comply fully, partially or not at all with the sender’s expected action. If the decision is to fully comply then the receiver performs whatever action was indicated (purposed!) by the sender’s message. Here we must distinguish between two types of receivers: 1) an intelligent being that possesses the capability of making free choices and is able to determine the meaning of the message, and 2) a machine that does not have these capabilities. In the former, the intelligent being can respond to the request or command in highly variable ways. With the machine the meaning has been programmed into command signals that ‘start’ or initiates the action level — the systems control program guides the machine to automatically perform the action. It must be pointed out that in both cases machines are essential for performing the expected actions [10]. In the case of the intelligent being, the machinery of his body may be sufficient or he may need to utilize external machinery (which may be mechanical/electrical machines, animals, other humans, etc.).

Intended Purpose

Prior to issuing an original written or verbal message there must be an internal thought process that motivates the sender to formulate a message. This thought

process is necessarily complex and involves need, motivation or intent for something to be achieved. If it is not to be performed by the sender, then the thought process must include selecting a particular receiver and determining whether or not that receiver is capable of performing the expected action. If the whole process is completed successfully, then the sender's original purpose is achieved. Thus we see that information's attribute of intended purpose is essential at the very beginning of a message. The achievement of that purpose is the result of the receiver's performance of the desired action. From this we see that the most important attribute of information is the intended purpose and that it is at both ends of a successfully completed message. Purpose may thus be thought of as the 'bookends attribute'.

The Definition of Universal Information

All four attributes described above are necessary to unambiguously distinguish this subset (category) of information. Due to this, the formal definition of Universal Information (UI) stated below incorporates all four of these distinguishing attributes.

A symbolically encoded, abstractly represented message conveying the expected action and the intended purpose.

Now we can appraise the three previously discussed working definitions of information in light of the attributes of Universal Information.

Shannon's classical information theory concerns itself only with statistical relationships of material symbols found within the code of Universal Information. This was because nothing more was necessary in order to address the technical issues of information transmission and storage. While Shannon stated this point clearly in his landmark paper [9], most modern day evolutionary theorists champion his definition primarily because it allows for the creation of 'information' by randomly assembling symbols. This makes creation of biological information trivial, and separates biological information from biological functionality. The attempt to define biological information in this way is clearly ideologically driven and is obviously not sufficient, since no thinking person would exclude meaning and purpose from biological (functional) information.

Algorithmic Information is a measure of the information content of material systems in terms of the degree of 'complexity' (as algorithmic 'complexity' is defined) of the system. Those material systems displaying greater complexity (more aperiodicity) have higher information content than those material systems

displaying less complexity (more periodicity). The four distinguishing attributes of Universal Information are *not* required for algorithmic information.

Complex Specified Information (CSI) exists in all material systems that exhibit a ‘specification’ and this specification is expressed in terms of functionality or purpose. As a result, CSI requires only UI’s distinguishing attribute of purpose. By definition this means that any system exhibiting CSI implies design. Even though all of the distinguishing attributes of UI were necessary during the design and construction phase, these attributes need not be present in the observed complex specified system.

The Nature of Universal Information

Having clearly distinguished Universal Information from other types/definitions of ‘information’, we now proceed to answer (at least for UI) the question [8]: if ‘information’ is not mass and energy, what is it? In the following discussion we will use the term ‘matter’ to include both mass and energy and the term ‘nonmaterial entity’ to refer to all entities outside the material domain.

There are many significant criteria for distinguishing material entities from nonmaterial entities. Perhaps the most simple, direct and scientific criterion is the fact that all material entities can be measured and thereby ‘quantified’ using one or a combination of the seven units of measurement established by the System International. These are the meter, kilogram, ampere, kelvin, mole, candela and second. Any entity within the universe that cannot be measured and described with one or a combination of these units is, by definition, a nonmaterial entity. Another criterion is that a nonmaterial entity does not and cannot originate from unguided, purely physicochemical processes [10, 12]. Finally, a nonmaterial entity does not have any direct physicochemical interaction with matter [10].

Universal Information satisfies all of the above criteria for a nonmaterial entity. A material medium is essential for the storage, transmission and processing of UI but, as described earlier, the quantity and type of matter that is used is highly variable and not correlated at all to the value of the Universal Information; i.e., the UI is completely independent of the material medium.

Additionally, the symbols (code level) that are utilized and physically manifested in the material domain display a vast degree of variation. To illustrate this, Figure 1 depicts the words from ten different languages that have the same meaning even though the individual symbols/letters differ markedly from one another. However, regardless of the symbols used the ‘content’ of the meaning remains essentially the same. Content as used here includes the attributes of meaning, action and purpose.

sixty-one three-letter words, or codons, means/represents/denotes one of the twenty amino acids utilized in polypeptide/protein synthesis. The codon for methionine also denotes or represents a start command. Additionally, the specific sequence of codons in messenger RNA (mRNA) represents the specific sequence of amino acids in the polypeptide precursor to the protein. Despite intensive research, no physicochemical bonding relationship has been found between the codons and the amino acids they represent [10, 12]. Hence, the second distinguishing attribute of UI is present: abstract meaning.

The Expected Action

The messenger RNA (mRNA) is transported out of the cell nucleus into the cytosol to a very complex RNA/protein machine — the ribosome. At the ribosome, beginning with a start codon on the mRNA, this specific mRNA codon is joined with an anticodon at one end of the small transfer RNA (tRNA) molecule. At the other end of tRNA is the amino acid specified by the mRNA codon, in this case methionine. The mRNA is then advanced one codon step and another tRNA anticodon is joined to the mRNA codon. At this stage two amino acids have been brought together and the ribosome, utilizing energy, joins the two amino acids together by forming a peptide bond. This process repeats itself until a stop codon is reached on the mRNA. The polypeptide thus formed is then folded by other protein machines into a functional protein with a highly specific three dimensional configuration. This precise synthesis of a unique functional protein by the ribosome (machine) fulfills the third distinguishing attribute of UI: expected action. However, this is only the first level of action — the proteins themselves have higher-level functions, e.g., the ribosome, which is primarily protein. At a macroscopic level the activity of proteins in muscles of higher animals perform useful work. Figure 2 demonstrates that DNA replication during cellular reproduction requires protein nanomachines such as DNA polymerase and usable energy. Next, transcription to mRNA requires a DNA template, several nanomachines (such as RNA polymerase and spliceosome) and usable energy. Finally, synthesis of all protein nanomachines and protein structural elements require mRNA, tRNA and nanomachines such as ribosomes and chaperonins, and usable energy. This essential closed-loop conundrum has stymied researchers for decades as they have attempted to account for the origin of the first living cell through unguided purely physicochemical processes. Their attempt at ‘protein first’, ‘DNA first’ or ‘RNA first’ models have all failed [10, 12]. As demonstrated in Figure 2, all three must be ‘first’ *simultaneously*.

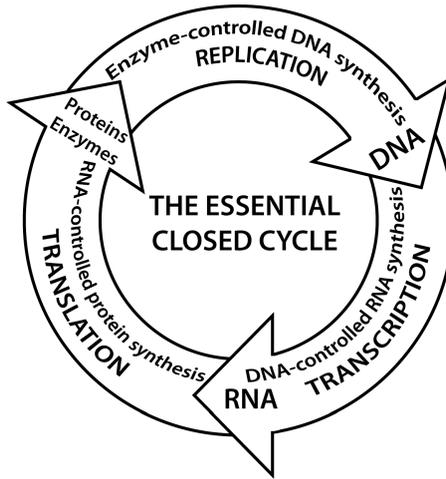


Fig. 2. A simplified representation of a UI-controlled cyclical process in living cells. The translation mechanism (protein synthesis) corresponds to the lowest level of expected action. However, the action of a protein nanomachine (DNA polymerase) is required in the next step of the cycle in DNA replication. The intricate process of mRNA synthesis (transcription) requires the DNA template and nanomachines (RNA polymerase II and spliceosome). Each of these three steps must be present *simultaneously*.

The Intended Purpose

The UI instructions for protein synthesis are stored within the nuclear DNA. The initial purposes of these UI instructions are achieved as the processes of transcription and translation are successfully accomplished. The ultimate physical purpose for the DNA/RNA protein synthesizing system is for the initial creation of organisms, and for their operations, maintenance and reproduction. Undoubtedly the earth's biosphere would not exist if all of the protein components were absent. At the intracellular level, while not identical, the protein requirements are similar in many areas for both plant and animal life. However, in multicellular animals the use of extracellular protein is far more extensive than in multicellular plants. Therefore, the greater diversity of protein in animals than in plants will require more complex amounts of UI stored in the DNA and transcribed into RNA. Further research into this difference as well as comparing the DNA coding for protein with the DNA coding for cellulose synthesis in plants may reveal important features of DNA coding. The multiple purposes achieved by the DNA/RNA protein synthesizing system attests to the fact that the fourth distinguishing attribute of UI (intended purpose) is indeed present.

UI Senders, Transmitters and Receivers

Problems associated with determining the origin and utilization of UI can be somewhat mitigated if we use specific terms to differentiate between the following:

1. An **original sender** is an intelligent agent that creates the original UI message. As demonstrated by Gitt *et al.* [10] this intelligent agent must have a nonmaterial component beyond the embedded UI. This is because even UI-guided purely physicochemical processes wholly constrained by natural laws have never been observed to create *de novo* UI despite all scientific efforts to date [10, 12]. Since humans do create *de novo* UI they qualify as original senders. This is strong evidence that humans have a nonmaterial component beyond their embedded UI [10].
2. **Intermediate transmitters** receive a UI message and simply copy, transmit, display or broadcast the message. Ideally, an intermediate transmitter will not distort the meaning of the original message in any way [10]. Intermediate transmitters can be intelligent agents or machines that are specifically designed to perform the transmitting processes.
3. **Machine receivers** obtain and process the messages and perform the commanded action thereby achieving the purpose intended by the original sender. Machine receivers (either mechanical or biological) do not have the capability to freely interpret the messages. They must be 'pre-programmed' with the capability to receive, then process and then execute the commanded actions *without* requiring that the meaning of the messages be determined. In other words, the programmer must convert the meaning of the messages into a series of preprogrammed executable steps that are initiated by start commands so that the proper actions are performed [10, 11].
4. **Intelligent receivers** possess the capability of determining the meaning of the message and also possess the capability of making free choices. This latter capability allows the intelligent receiver to decide whether to perform the expected action fully, partially or not at all.

When the UI in the DNA/RNA protein synthesizing system is expressed in biological life, it guides the transcription/translation processes to produce a specifically controlled amount of a specified protein. This protein will then perform specific functions within the cell or within the organism. This is an example of number 3 above whereby machines are guided by instructions (namely UI) that

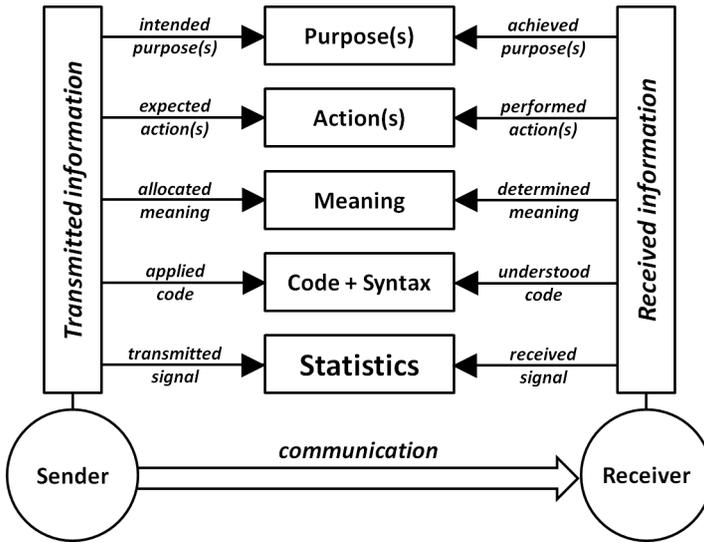


Fig. 3. A comprehensive diagram of the five levels of Universal Information. All five levels are relevant for both the intelligent sender and the intelligent receiver.

was stored in the nuclear DNA of the cell by the original sender of that UI. Figure 3 displays a comprehensive diagram of Universal Information being originated (*de novo*) by an intelligent sender and being received by an intelligent receiver. More complex diagrams that include transmitters and machine receivers can be found in [10].

The Existence, Validity and Significance of Universal Information

While identifying and studying the distinguishing attributes of Universal Information (UI), we discovered and formulated 32 Empirical Statements involving the origin and nature of UI [10]. We have repeatedly verified these Empirical Statements over a 30-year period. Not one of these Empirical Statements has ever been refuted despite wide dissemination of this information and they remain an open challenge to this day.

We then turned our attention to the code discovered in the DNA and the volumes of research describing it. It was easily determined that Universal Information is definitely stored, transmitted and utilized within the DNA/RNA Protein Synthesizing System of all living organisms. In other words, UI is not merely an interesting theoretical concept; UI truly exists. UI is not only a foundational component of human languages and communication, it is also a vital control system found in all biological life on earth.

Undoubtedly the most important activity in science is to utilize factual data and observations to construct reliable and valid conclusions. This goal is achieved via sound, logical arguments that lead to those conclusions. According to Kreeft [13], there are three things that must be in place in order to develop logically sound arguments.

- The significant terms must be unambiguous.
- The premises must be true.
- The conclusion must logically follow from the premises (logically valid).

In order to satisfy these three requirements, we carefully defined all significant terms so that there would be clear, unambiguous formulations of the questions, arguments and conclusions [10]. In addition, the Empirical Statements were continually evaluated by a number of independent individuals in order to ensure clarity of meaning and validity of the statements. With this foundation we then proposed specific Empirical Statements as Scientific Laws and used them, along with verified scientific facts, as premises in our deductions [10]. Therefore, our premises are extracted from the two categories of science — verified facts and scientific laws — that have the highest degree of scientific certainty. Finally, we constructed ten logically-sound deductions that led to ten strong conclusions [10]. By rigorously following this procedure we have minimized investigator bias or interference from our conclusions. This is important for any conclusion in science, but especially so in this case because of the broad significance of these ten conclusions. Also, by minimizing investigator interference these results retain objective validity to the extent that this is possible.

Conclusion

Coming full circle, we return to our original question regarding Biological Information — *What is it?* We have identified an important subset of Biological Information that we call Universal Information that is present in every cell of every living organism.

We the authors of this paper used Universal Information in order to communicate these things to you. This Universal Information was processed through our brains that in turn, controlled our body parts to write the words on this page. These words reach receptors in your visual system that will then send impulses (i.e., messages) to your brain. You then determine the meaning of the words of our message and consider their significance. This too is Universal Information and is also part of Biological Information.

Between our intracellular DNA/RNA systems and our capacity to express thoughts through words there are many levels of highly integrated, organized biological systems which themselves necessarily operate under the control of some type of biological information. At each of these levels there are many structural components and biological machines that perform the required actions. Essentially all of these structures and machines are composed of proteins synthesized by the DNA/RNA protein synthesizing system.

Will we find Biological Information in forms other than Universal Information? We believe that we will. For instance, we are already aware of Mental Image Information (MII). MII is information in which there is meaning, action and purpose but no *abstract* code, syntax or abstract meaning. Recall that two of the distinguishing attributes of Universal Information is an abstract code with syntax and abstract meaning such as that which is manifested in the DNA/RNA protein synthesizing system. We know that MII plays a role in living organisms yet MII does not have an abstract code, syntax or meaning. For example, a ‘spoon and fork’ on a highway sign *directly* (i.e., not abstractly) represents ‘food’ or ‘eating place’ since it resembles the entity that it represents. Another example is the pheromones emitted by certain insects for, say, mating purposes. These pheromones have an inherent physicochemical relationship with the entity they represent. When received, these pheromones convey meaning, expected action and purpose *directly* (i.e., not abstractly) instead of through some intermediate substitute possessing ‘abstract meaning’ expressed via an abstract code with an associated syntax. In other words, the pheromone molecule is not an abstract substitute for the entity, it is the entity itself.

Just as was the case for Complex Specified Information in Intelligent Design Theory, Universal Information and related topics represent a revolutionary departure from the materialistic approach to information. Since UI and its requisite machines have great explanatory power in biology, a search for machines, even without explicit (embedded) UI, operating at various ranges of scale in the inanimate world may also yield results with great explanatory power [10].

References

1. Montañez G, Marks II RJ, Fernandez J, Sanford JC (2013) Multiple overlapping genetic codes profoundly reduce the probability of beneficial mutation. In: Marks II RJ, Behe MJ, Dembski WA, Gordon B, Sanford JC (eds) Biological Information — New Perspectives. World Scientific, Singapore, pp. 139–167.
2. Funk and Wagnalls Standard Dictionary (1962).
3. Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana IL.

4. Solomonoff R (1960) A preliminary report on a general theory of inductive inference. Report V-131, Zator Co, Cambridge, MA.
5. Kolmogorov AN (1965) Three approaches to the quantitative definition of information. *Problems Inform. Transmission* 1 (1): 1–7.
6. Chaitin GJ (1969) On the simplicity and speed of programs for computing infinite sets of natural numbers. *Journal of the ACM* 16: 407.
7. Dembski WA (1998) *The design inference*. Cambridge University Press, United Kingdom.
8. Wiener N (1968) *Cybernetics: or the control and communication in the animal and the machine*. First American Edition (Technology Press, MIT).
9. Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* Vol.27.
10. Gitt W, Compton B, Fernandez J (2011) *Without Excuse*. Creation Book Publishers.
11. Seaman J (2013) DNA.EXE: A sequence comparison between the human genome and computer program code. In: Marks II RJ, Behe MJ, Dembski WA, Gordon B, Sanford JC (eds) *Biological Information — New Perspectives*. World Scientific, Singapore, pp. 385–401.
12. Meyer SC (2009) *Signature in the cell*. Harper One, New York.
13. Kreeft P (1989) *Three philosophies of life*. Ignatius Press, San Francisco, CA.