



Review

Cite this article: Barbieri M. 2016 What is information? *Phil. Trans. R. Soc. A* **374**: 20150060.

<http://dx.doi.org/10.1098/rsta.2015.0060>

Accepted: 1 July 2015

One contribution of 21 to a theme issue
'DNA as information'.

Subject Areas:

algorithmic information theory,
bioinformatics, computational biology,
systems theory

Keywords:

information, meaning, codes, artefacts,
observables, ontology

Author for correspondence:

Marcello Barbieri
e-mail: brr@unife.it

What is information?

Marcello Barbieri

Dipartimento di Morfologia ed Embriologia, Via Fossato di Mortara
64a, Ferrara 44121, Italy

MB, 0000-0001-7091-0876

Molecular biology is based on two great discoveries: the first is that genes carry hereditary information in the form of linear sequences of nucleotides; the second is that in protein synthesis a sequence of nucleotides is translated into a sequence of amino acids, a process that amounts to a transfer of information from genes to proteins. These discoveries have shown that the *information* of genes and proteins is the specific linear order of their sequences. This is a clear definition of information and there is no doubt that it reflects an experimental reality. What is not clear, however, is the *ontological* status of information, and the result is that today we have two conflicting paradigms in biology. One is the 'chemical paradigm', the idea that 'life is chemistry', or, more precisely, that 'life is an extremely complex form of chemistry'. The other is the 'information paradigm', the view that chemistry is not enough, that 'life is chemistry plus information'. This implies that there is an ontological difference between information and chemistry, a difference which is often expressed by saying that information-based processes like heredity and natural selection simply do not exist in the world of chemistry. Against this conclusion, the supporters of the chemical paradigm have argued that the concept of information is only a linguistic metaphor, a word that summarizes the result of countless underlying chemical reactions. The supporters of the information paradigm insist that information is a real and fundamental component of the living world, but have not been able to *prove* this point. As a result, the chemical view has not been abandoned and the two paradigms both coexist today. Here, it is shown that a solution to the ontological problem of information does exist. It comes from the idea that *life is artefact-making*, that genes and proteins are molecular artefacts manufactured by molecular machines and that artefacts necessarily require sequences and coding rules in addition to the quantities of physics and chemistry. More precisely, it is shown that the production of artefacts requires

new observables that are referred to as *nominable* entities because they can be described only by *naming* their components in their natural order. From an ontological point of view, in conclusion, information is a nominable entity, a fundamental but not-computable observable.

1. The chemical paradigm

The view that ‘life is chemistry’ was proposed for the first time by van Helmont [1] and has been re-proposed countless times ever since. One of the most recent formulations has been given by Wächtershäuser [2, p. 483] in these terms: ‘If we could ever trace the historic process backwards far enough in time, we would wind up with an origin of life in purely chemical processes’.

In 1953, Watson & Crick [3,4] pointed out that the nucleotides of the genes are arranged in a linear order in the double helix of DNA. A few years later, the mechanism of protein synthesis was discovered and it was found that the sequence of nucleotides in genes determines the sequence of amino acids in proteins, with a process that amounts to a transfer of linear order from genes to proteins [5]. This led to the idea that *biological information* is the specific linear order in which the subunits of a polymer are arranged [6]. These discoveries gave origin to the *information paradigm*, the idea that living systems are *information-processing machines*, and that life is based not only on chemistry (energy and matter) but also, and above all, on *information* [7].

Wächtershäuser [2, p. 492] claimed instead that ‘information is a teleological concept’, and gave a specific example of the conflict between chemistry and teleology: ‘On the level of nucleic acid sequences it is convenient to use the information metaphor ... and apply teleological notions such as “function” or “information” ... but in the course of the process of retrodiction the teleological notions, whence we started, fade away. And what remains is purely chemical mechanism’. This amounts to saying that biological information, the most basic concept of molecular biology, does not *really* belong to science.

This is the ontological claim of the chemical paradigm, the idea that all natural processes are completely described, in principle, by physical quantities. This view is also known as *physicalism*, and it is based on the fact that biological information is not a physical quantity. So, what is it? A similar problem arises with the rules of the genetic code: they cannot be measured and cannot be reduced to physical quantities, so what are they?

According to physicalism, biological information and the genetic code are mere *metaphors*. They are like those computer programs that allow us to write our instructions in English, thus saving us the trouble of writing them in the binary digits of the machine language. Ultimately, however, there are only binary digits in the machine language of the computer, and in the same way, it is argued, there are only physical quantities at the most fundamental level of Nature.

This conclusion, known as the *physicalist thesis*, has been proposed in various ways by a number of scientists and philosophers [8–14] and is equivalent to the thesis that ‘life is chemistry’.

This is one of the most deeply dividing issues of modern science. Many biologists are convinced that biological information and the genetic code are real and fundamental components of life, but physicalists insist that they are real only in a very superficial sense and that there is nothing fundamental about them because they must be reducible, in principle, to physical quantities.

2. The information paradigm

The idea that life evolved naturally on the primitive Earth suggests that the first cells came into being by spontaneous chemical reactions, and this is equivalent to saying that there is no *fundamental* divide between life and matter. This is the *chemical paradigm*, a view that is very popular today and that is often considered in agreement with the Darwinian paradigm, but this is not the case. The reason is that natural selection, the cornerstone of Darwinian evolution, does not exist in inanimate matter. In the 1950s and 1960s, furthermore, molecular biology uncovered two

fundamental components of life—biological information and the genetic code—that are totally absent in the inorganic world, which means that information is present only in living systems, that chemistry alone is not enough and that a deep divide does exist between life and matter. This is the *information paradigm*, the idea that ‘life is chemistry plus information’.

Ernst Mayr, one of the architects of the modern synthesis, has been one of the most outspoken supporters of the view that life is *fundamentally* different from inanimate matter. In *The growth of biological thought* [15, p. 124], he made this point in no uncertain terms: ‘... The discovery of the genetic code was a breakthrough of the first order. It showed why organisms are fundamentally different from any kind of nonliving material. There is nothing in the inanimate world that has a genetic program which stores information with a history of three thousand million years!’

The idea that ‘life is chemistry plus information’ implies that information is *ontologically* different from chemistry, but can we prove it? Perhaps the strongest argument in support of this claim has come from Hubert Yockey, one of the organizers of the first congress dedicated to the introduction of Shannon’s information in biology [16]. In a long series of articles and books, Yockey [17–19] has underlined that heredity is transmitted by factors that are ‘segregated, linear and digital’ whereas the compounds of chemistry are ‘blended, three-dimensional and analogue’.

Yockey underlined that: ‘Chemical reactions in non-living systems are not controlled by a message ... There is nothing in the physico-chemical world that remotely resembles reactions being determined by a sequence and codes between sequences’ [18, p. 105].

Yockey has tirelessly pointed out that no amount of chemical evolution can cross the barrier that divides the *analogue world of chemistry* from the *digital world of life*, and concluded from this that the origin of life cannot have been the result of chemical evolution. This is therefore, according to Yockey, what divides life from matter: information is ontologically different from chemistry because linear and digital sequences cannot be generated by the analogue reactions of chemistry.

At this point, one would expect to hear from Yockey how did linear and digital sequences appear on Earth, but he did not face that issue. He claimed instead that the origin of life is *unknowable*, in the same sense that there are propositions of logic that are *undecidable*. This amounts to saying that we do not know how linear and digital entities came into being; all we can say is that they were not the result of spontaneous chemical reactions. The information paradigm, in other words, has not been able to prove its ontological claim, and that is why the chemical paradigm has not been abandoned.

3. The idea that ‘life is artefact-making’

The physicalist thesis would be correct if genes and proteins were spontaneous molecules, because there is no doubt that all spontaneous reactions are completely accounted for by physical quantities. This, however, is precisely the point that molecular biology has proved wrong. Genes and proteins are *not* produced by spontaneous processes in living systems. They are produced by molecular machines that physically stick their subunits together and are therefore *manufactured molecules*, i.e. *molecular artefacts*. This in turn means that all biological structures are manufactured, and therefore that the whole of life is *artefact-making* [20–22].

Spontaneous genes and spontaneous proteins did appear on the primitive Earth but did not evolve into the first cells because they did not have biological *specificity*. They gave origin instead to *molecular machines* and it was these machines and their products that evolved into the first cells.

The simplest molecular machines that appeared spontaneously on the primitive Earth were *bondmakers*, molecules that could bind monomers together at random and produce statistical polymers. It has been shown, for example, that the ribosomal RNAs are among the most conserved molecules in evolution [23,24] and contain regions which have the ability to form peptide bonds [25]. This means that pieces of ribosomal RNAs appeared very early on the primitive Earth and that some of them could stick amino acids together into random proteins. Some bondmakers, furthermore, evolved the ability to join monomers no longer at random but in the order provided by *templates*. Those bondmakers were making copies of the templates

and became *copymakers*, the first molecular machines that started populating the Earth with a potentially unlimited number of nucleic acids.

The origin of protein life, on the other hand, was a much more complex affair, because proteins cannot be copied and their reproduction required molecular machines that employ a *code*, machines that have been referred to as *codemakers*. The evolution of the molecular machines, in short, started with *bondmakers*, went on to *copymakers* and finally gave origin to *codemakers*.

Molecular biology, in short, has discovered that genes and proteins require not only catalysts but also *templates*. The catalysts join the subunits together, and the templates provide the *order* in which the subunits are assembled. It is that order that determines biological specificity, and in proteins that order comes from messenger RNAs, i.e. from *external* molecules.

This is precisely the characteristic that divides spontaneous molecules from molecular artefacts. In spontaneous molecules, the order of the components comes *from within* the molecules, from *internal* factors, whereas in genes and proteins it comes *from without*, from *external* templates. This amounts to saying that a molecular system becomes a ‘maker’ when it starts using templates in the production of objects. The difference between spontaneous and manufactured molecules, in short, is a reality because it is an *experimental fact* that genes and proteins are *template-dependent* objects.

This makes us realize that the physicalist thesis is wrong because it is only spontaneous processes that are completely described by physical quantities. The same applies to Yockey’s argument: it is true that linear, digital and specific properties do not exist in spontaneous processes, but they do exist in all manufacturing processes, including those that are produced by molecular machines at the molecular level.

When a copymaker scans a nucleic acid and makes a copy of that molecule, what happens is precisely an operation that brings into existence a linear and digital copy of a pre-existing molecule. It was molecular copying—the simplest form of artefact-making—that started manufacturing biological objects and set in motion the odyssey of life on the primitive Earth.

What is particularly important, to our purposes, is that the concept of artefact-making explains how it is possible that life evolved from inanimate matter and yet it is fundamentally different from it. The divide between life and matter is real because *matter is made of spontaneous objects* whereas *life is made of manufactured objects*.

It must be pointed out that similar ideas have been proposed independently by various authors, in particular by Küppers [26] and Roederer [27].

4. The five properties of sequences

Both the sequence of nucleotides in a gene and the sequence of letters in a book are carrying *information*: hereditary information in genes and syntactic information in language. In both cases, the information is *digital* (because it is made of discrete units) and *linear* (because the units are arranged in a linear order). This is also true in many other cases. A photograph, for example, can be represented by a two-dimensional matrix of pixels, but the rows of the matrix can also be arranged one after the other in a one-dimensional string, and the photograph becomes in this way a linear sequence of pixels. The same can be done with any piece of music by arranging its notes in a line. Finally, we can represent letters, numbers, pixels, musical notes and many other symbols with the characters of computer language, and any sequence can be represented by a sequence of *bytes*. A sequence, in short, is *any collection of a finite number of digital units that are arranged in a linear order*, and the order of the units is the *information* carried by the sequence.

Sequences, on the other hand, have other properties in addition to the linear order of their units, and each of them tells us something different.

The first of these properties was introduced by Claude Shannon in 1948 and represents the *probability* of a sequence. Shannon pointed out that when we receive a message we pass from a state of ignorance to one of knowledge, as we do when we toss a coin and pass from a state of equal probabilities to one of certainty. This is why he proposed that a message is characterized by the *probability* of choosing that particular message from the set of all potential alternatives [28], a

probability that turned out to have an entropy-like formula and for this reason it has been referred to as *statistical information*.

A second important property of a sequence was introduced independently by Kolmogorov [29] and Chaitin [30] and is the *algorithmic complexity* of a sequence measured by its *length* in bytes. More precisely, the *algorithmic complexity* of a sequence is defined as the *shortest binary message* that instructs a computer to generate it, and the *length* of this message is taken as a measure of the overall complexity of the sequence, whatever is the linear order of its units.

The third property of a sequence is its *distance* from another sequence, a quantity that measures the degree of *compatibility*, or *relatedness*, that exists between them. It is possible, for example, to count the number of nucleotides that a gene has in common with a reference gene, and obtain in this way an estimate of the similarity that exists between them. In the literature, the same technique allows us to measure the distance between two books or two languages by counting the number of words that they have in common.

Another outstanding property of a sequence is that it *can* have a *meaning*. The linear order, and therefore the *information*, of the word 'ape', for example, is the same in all languages, but in English it means 'tailless monkey', whereas in Italian it means 'bee' and in French it has no meaning.

There are, in conclusion, five distinct characteristics in a sequence. Three of them (probability, length and relative distance) can be measured whereas the other two (information and meaning) cannot. All we can do is to *name* their components in their natural order, and this gives us the problem of understanding the role that they actually have in living systems.

5. Organic information

In genes and proteins, as we have seen, organic information is the specific sequence of their subunits. This definition however is not entirely satisfactory because it gives the impression that information is a *static* property, something that molecules have simply because they have a sequence. In reality, there are countless molecules which have a sequence but only in a few cases this becomes information. That happens only when a copymaker uses it as a guideline for copying. Without such guidelines, on the other hand, copymakers would produce only *random* sequences, not specific ones. Sequences alone or copymakers alone, in other words, have nothing to do with information. It is only when a sequence provides a guideline to a copymaker that it becomes information for it. It is only an act of copying, in other words, that brings organic information into existence.

This tells us that organic information is not just the specific sequence of a molecule, but *the specific sequence produced by a copying process*. This definition underlines the fact that organic information is not a thing or a property, but the result of a process. It is, more precisely, an *operative* definition, because information is defined by the process that brings it into existence. We realize in this way that organic information is as real as the copying process that generates it. This in turn means that organic information is essential to describe genes and proteins. To this purpose, in fact, it is no less essential than the physical quantities, and this means that organic information *has the same scientific status as a physical quantity*. They both belong to the class of objective entities that allow us to describe the world.

This conclusion, however, raises immediately a new problem, because there are two distinct groups of physical quantities: a small group of *fundamental* quantities (space, time, mass, charge and temperature) and a much larger group of *derived* quantities. That distinction applies to all objective entities, so we need to find out whether organic information belongs to the first or to the second group.

Luckily, this problem has a straightforward solution because the sequences of genes and proteins have two very special characteristics. One is that *a change in a single component of a biological sequence may produce a sequence which has entirely new properties*. This means that although a biological sequence can be said to have 'components', it is at the same time a single indivisible whole. The second outstanding feature is that *from the knowledge of n elements of a biological sequence*

we cannot predict the element $(n + 1)$. This is equivalent to saying that a specific sequence cannot be described by anything simpler than itself, so it cannot be a derived entity.

We conclude that organic information does not have the status of a derived physical quantity because it cannot be expressed by anything simpler than itself. This means that organic information has the same scientific status as the fundamental quantities of physics.

Again it has to be underlined that a similar idea has been proposed by Küppers [26] in respect to the concept of ‘pragmatic information’.

6. Organic meaning

A code is a set of rules that establish a correspondence between the objects of two independent worlds. The Morse code, for example, is a correspondence between groups of dots and dashes with the letters of the alphabet, and the genetic code is a correspondence between groups of nucleotides and amino acids. Let us note now that establishing a correspondence between, say, object 1 and object 2 is equivalent to saying that object 2 is the *meaning* of object 1. In the Morse code, for example, the rule that ‘dot-dash’ corresponds to the letter ‘A’ is equivalent to saying that letter ‘A’ is the meaning of ‘dot-dash’. In the code of the English language, the sound ‘apple’ is associated with the fruit ‘apple’, and this is equivalent to saying that that fruit is the meaning of that sound.

By the same token, the rule of the genetic code that a group of three nucleotides (a codon) corresponds to an amino acid is equivalent to saying that that amino acid is the *organic meaning* of that codon. Anywhere there is a code, be it in the mental or in the organic world, there is meaning. We can say, therefore, that *meaning is an entity which is related to another entity by a code*, and that organic meaning exists whenever an organic code exists [31–33].

The existence of meaning in the organic world may seem strange, at first, but in reality it is no more strange than the existence of a code, because they are the two sides of the same coin. To say that a code establishes a correspondence between two entities is equivalent to saying that one entity is the meaning of the other, so we cannot have codes without meaning or meaning without codes. All we need to keep in mind is that *meaning is a mental entity when the code is between mental objects, but it is an organic entity when the code is between organic molecules*.

Modern biology has readily accepted the concept of information but has carefully avoided the concept of meaning, and yet they are parallel processes. We have seen that organic information cannot be measured, and the same is true for organic meaning. We have seen that organic information is an objective entity, because any sequence is the same for any number of observers, and that is also true for organic meaning, which consists in coding rules that are the same for all observers. Finally, we have seen that organic information is a non-reducible entity, because it cannot be described by anything simpler than its sequence, and the same is true for organic meaning, which cannot be defined by anything simpler than its coding rules.

Organic information and organic meaning, in short, belong to the same class of natural entities because they have the same defining characteristics, but what exactly are they? What is it that they have in common and what is it that differentiates them from the other entities of Nature?

7. The new observables

Physical theory starts with the definition of natural entities, or *observables* (space, time, mass, etc.), and then looks for relationships between them, which are referred to as *laws* and *constraints*. The choice of the observables is the very first step in the description of the world. The movements of planets and stars, for example, can be described with only two observables—space and time—and in that case we get either a Ptolemaic model or a Copernican system. By introducing a third observable—mass—we obtain the laws of motion, universal gravitation and the Newtonian model of the world.

The three fundamental observables of classical physics can be combined together in different ways and produce many derived observables (speed, acceleration, force, energy, power,

momentum, etc.), but what defines the whole system is the initial number of fundamental observables. The identity of these observables can be changed (space and time, for example, can be replaced by speed and time, and in that case space becomes a derived entity), but the minimum number of fundamental observables does not change. This number defines a whole world of phenomena, and we can discover new worlds, new aspects of reality, only if we discover new fundamental observables. The world of electromagnetism, for example, required the introduction of new fundamental observables, and so did the world of thermodynamics, the world of nuclear forces and the world of elementary particles.

Classical physics, thermodynamics, electromagnetism and elementary particles, in other words, were all based on the discovery of new fundamental observables, and now we realize that this is also true in biology. Life is based on the copying of genes and on the coding of proteins, and these processes require entities, like sequences and coding rules, that have all the defining characteristics of *new observables*. This is because the role of observables is to allow us to describe the world and we simply cannot describe living systems without sequences and codes. But what kind of entities are these new observables?

According to a long tradition, natural entities are divided into *quantities* and *qualities*. Quantities are objective and can be measured, whereas qualities are subjective and cannot be measured. In the case of organic information and organic meaning, however, this scheme breaks down. Organic information is not a quantity because the specific order of a sequence cannot be measured. But it is not a quality either, because that order is recognized by all observers and is therefore an objective feature of the world, not a subjective one. The same is true for organic meaning. This too cannot be measured, so it is not a quantity, but it is not a quality either because the rules of the genetic code are the same for all observers.

A scheme based on quantities and qualities alone, in short, is not enough to describe the world. In addition to quantities (*objective and measurable*) and qualities (*subjective and not-measurable*) we must recognize the existence in Nature of a third type of entities (*objective but not-measurable*). Organic information and organic meaning belong precisely to this third class. They are a new type of observables, so let us take a closer look and see if we can learn something more about them.

8. Names and nominable entities

Physical theory consists of laws, constraints and observables, but in addition to these three components there is also a fourth one, and that is *names*. Science is always expressed in words and we need therefore to give names to what we observe in Nature. Names (including those that we call ‘numbers’) are necessarily a fourth essential component of physical theory, but are different from the first three because they are language-dependent. This is because names (or *nominal entities*, to use a classical term) in general have nothing to do with the intrinsic features of the named objects, and are therefore mere labels that we attach to them.

The deep divide that exists between ‘names’ and ‘objects’ has been at the centre of many controversies in the past, in particular of the celebrated mediaeval dispute over ‘nominal entities’ versus ‘real entities’. The relationship between names and objects is also a crucial issue in science, but here it has taken on a new form. Let us underline that all names are sequences of characters (alphabetic, numerical or alpha-numerical) and that each sequence is *unique*. Names, in other words, have *specificity*. In general, the specificity of a name has nothing to do with the characteristics of the named object, and in these cases we can truly say that names are mere labels. Science, however, has invented a new type of names where the sequence of characters *does* represent an order that is objectively present in the named objects.

The chemical formula of a molecule, for example, describes an objective sequence of atoms, and any atom can be described by the objective sequence of its quantum numbers. In these cases, the names are no longer arbitrary labels but true *observables* because they describe characteristics that we observe in Nature. This shows that there are two distinct types of names in science: *labels* and *observables*.

In the case of the observables, furthermore, there is another distinction that must be considered. The spontaneous assembly of amino acids generates proteins with random sequences, whereas specific sequences can be obtained only if the amino acids are put together by a molecular machine in the order provided by a template that is *external* to the protein itself. We need therefore to distinguish between two different types of observables.

The sequence of quantum numbers in an atom, or the sequence of atoms in inorganic molecules, is determined *from within*, by internal factors, whereas the sequence of amino acids in a protein is determined *from without*, by external templates. In the first case the sequence is a *physically computable* entity, in the sense that it is the automatic result of physical forces, whereas in the second case it can only be described by ‘naming’ its components, and is therefore a *nominable* entity (this term should not be confused with the classical concept of *nominal* entity, which applies to all names). A *nominable* entity, in other words, is not a label but an observable, and more precisely a *non-computable* observable.

All names, in conclusion, are specific sequences of alpha-numeric characters, and in science they can be divided into two great classes: labels and observables. The observables, in turn, can be divided into *computable* and *nominable* entities. The important point is that physics and chemistry deal exclusively with computable entities (physical quantities), whereas nominable entities (information and coding rules) exist only in living systems. Organic information and organic meaning, in conclusion, are not mere names, as the chemical paradigm has claimed: they are *fundamental nominable observables*.

9. A new paradigm

The idea that life is an extremely complex form of chemistry is still very popular today, and is based on the physicalist thesis that all biological processes can be reduced, in principle, to physical quantities. According to this view, genetic information and the genetic code are metaphorical and teleological terms that we use only because they are intuitively appealing.

We have seen however that the physicalist thesis is valid only in spontaneous systems, whereas genes and proteins are invariably manufactured by molecular machines, and all manufacturing processes require not only physical quantities but also additional entities like sequences and coding rules. The information carried by sequences, in other words, is an *observable* because it is absolutely necessary to the *description* of living systems. We simply cannot describe the transmission of genes or the synthesis of proteins without their sequences, and we cannot replace these sequences with anything else, which means that using information to describe living systems is perfectly equivalent to using space, time and mass to describe physical systems.

We realize in this way that there is no more teleology in organic information and in the genetic code than there is in the quantities of physics and chemistry. Sequences (organic information) and coding rules (organic meaning) are *descriptive* entities and are absolutely essential to the scientific study of life.

The information paradigm, on the other hand, has claimed that information is distinct from chemistry but has not been able to say why. On top of that, the idea that ‘life is chemistry plus information’ is equivalent to saying that genetic information is real but the genetic code is not, again without being able to say why.

We need therefore a paradigm that goes beyond the two present paradigms of biology. A paradigm that fully accepts the implications of the existence of the genetic code. The implication that life is based on copying *and* coding, that both biological sequences (organic information) and biological coding rules (organic meaning) are *fundamental* observables that are as essential to life as the fundamental quantities of physics. This is the *code paradigm*, the idea that ‘life is chemistry plus information plus codes’.

Competing interests. I declare I have no competing interest.

Funding. I received no funding for this study.

Acknowledgements. I am indebted to two anonymous referees whose comments and suggestions have greatly improved the revised version of this manuscript.

References

1. van Helmont JB. 1648 *Ortus medicinae*. Amsterdam, The Netherlands: Apud Ludovicum Elzevirium. (Engl. transl.: ML Gabriel, S Fogel (eds). 1995 *Great experiments in biology*. Englewood Cliffs, NJ: Prentice-Hall.)
2. Wächtershäuser G. 1997 The origin of life and its methodological challenge. *J. Theor. Biol.* **187**, 483–494. (doi:10.1006/jtbi.1996.0383)
3. Watson JD, Crick FHC. 1953 Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738. (doi:10.1038/171737a0)
4. Watson JD, Crick FHC. 1953 Genetical implications of the structure of deoxyribose nucleic acid. *Nature* **171**, 964–967. (doi:10.1038/171964b0)
5. Crick FHC. 1957 Discussion in *The structure of nucleic acids and their role in protein synthesis*, Biochemical Society Symposium, no. 14, London, February, pp. 25–26 (ed. EM Cook). New York, NY: Cambridge University Press.
6. Kay LE. 2000 *Who wrote the book of life? A history of the genetic code*. Stanford, CA: Stanford University Press.
7. Maynard-Smith J. 2000 The concept of information in biology. *Philos. Sci.* **67**, 177–194. (doi:10.1086/392768)
8. Boniolo G. 2003 Biology without information. *Hist. Philos. Life Sci.* **25**, 255–273. (doi:10.1080/03919710312331273055)
9. Chargaff E. 1963 *Essays on nucleic acids*. Amsterdam, The Netherlands: Elsevier.
10. Griffith PE. 2001 Genetic information: a metaphor in search of a theory. *Philos. Sci.* **68**, 394–412. (doi:10.1086/392891)
11. Griffith PE, Knight RD. 1998 What is the developmental challenge? *Philos. Sci.* **65**, 276–288. (doi:10.1086/392639)
12. Mahner M, Bunge M. 1997 *Foundations of biophilosophy*. Berlin, Germany: Springer. (doi:10.1007/978-3-662-03368-5)
13. Sarkar S. 1996 Biological information. A skeptical look at some central dogmas of molecular biology. In *The philosophy and history of biology* (ed. S Sarkar), pp. 187–231. Dordrecht, The Netherlands: Kluwer Academic.
14. Sarkar S. 2000 Information in genetics and developmental biology. *Philos. Sci.* **67**, 208–213. (doi:10.1086/392771)
15. Mayr E. 1982 *The growth of biological thought*. Cambridge, MA: Belknap Press.
16. Yockey HP, Platzman RL, Quastler H. (eds). 1958 *Symposium on information theory in biology*. New York, NY: Pergamon.
17. Yockey HP. 1974 An application of information theory to the central dogma and the sequence hypothesis. *J. Theor. Biol.* **46**, 369–406. (doi:10.1016/0022-5193(74)90005-8)
18. Yockey HP. 2000 Origin of life on Earth and Shannon's theory of communication. *Comput. Chem.* **24**, 105–123. (doi:10.1016/S0097-8485(00)80010-8)
19. Yockey HP. 2005 *Information theory, evolution, and the origin of life*. Cambridge, UK: Cambridge University Press.
20. Barbieri M. 2004 The definitions of information and meaning. Two possible boundaries between physics and biology. *Rivista di Biologia-Biology Forum* **97**, 91–110. See <http://europepmc.org/abstract/MED/15648213>.
21. Barbieri M. 2006 Life and semiosis: the real nature of information and meaning. *Semiotica* **158**, 233–254. (doi:10.1515/sem.2006.007)
22. Barbieri M. 2015 *Code biology. A new science of life*. Dordrecht, The Netherlands: Springer. (doi:10.1007/978-3-319-14535-8)
23. Woese CR. 1987 Bacterial evolution. *Microbiol. Rev.* **51**, 221–271. See <http://mmbr.asm.org/content/51/2/221.short?rss=1&ssource=mfc>.
24. Woese CR. 2000 Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396. (doi:10.1073/pnas.97.15.8392)
25. Nitta I, Kamada Y, Noda H, Ueda T, Watanabe K. 1998 Reconstitution of peptide bond formation. *Science* **281**, 666–669. (doi:10.1126/science.281.5377.666)

26. Küppers B-O. 1990 *Information and the origin of life*. Cambridge, MA: MIT Press.
27. Roederer JG. 2005 *Information and its role in Nature*. Berlin, Germany: Springer. (doi:10.1007/3-540-27698-X)
28. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–424, 623–656. (doi:10.1002/j.1538-7305.1948.tb01338.x)
29. Kolmogorov AN. 1965 Three approaches to the quantitative definition of information. *Probl. Inform. Transm.* **1**, 4–7.
30. Chaitin GJ. 1966 On the length of programs for computing binary sequences. *J. Assoc. Comput. Mach.* **13**, 547–569. (doi:10.1145/321356.321363)
31. Artmann S. 2007 Computing codes versus interpreting life. Two alternative ways of synthesizing biological knowledge through semiotics. In *Introduction to biosemiotics. The new biological synthesis* (ed. M Barbieri), pp. 209–233. Dordrecht, The Netherlands: Springer. (doi:10.1007/1-4020-4814-9_9)
32. Artmann S. 2009 Basic semiosis as code-based control. *Biosemiotics* **2**, 31–38. (doi:10.1007/s12304-008-9037-x)
33. Barbieri M. 2003 *The organic codes. An introduction to semantic biology*. Cambridge, UK: Cambridge University Press.