PHILOSOPHICAL
TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Cite this article: Barbieri M. 2016 A new theory of development: the generation of complexity in ontogenesis. *Phil. Trans. R. Soc. A* **374**: 20150148.
http://dx.doi.org/10.1098/rsta.2015.0148

Accepted: 1 August 2015

One contribution of 18 to a theme issue 'DNA as information'.

Subject Areas:
algorithmic information theory, bioinformatics, computational biology

Keywords:
development, complexity, organic codes, organic memories, reconstruction methods

Author for correspondence:
Marcello Barbieri
e-mail: brr@unife.it

A new theory of development: the generation of complexity in ontogenesis

Marcello Barbieri

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

MB, 0000-0001-7091-0876

Today there is a very wide consensus on the idea that embryonic development is the result of a genetic programme and of epigenetic processes. Many models have been proposed in this theoretical framework to account for the various aspects of development, and virtually all of them have one thing in common: they do not acknowledge the presence of organic codes (codes between organic molecules) in ontogenesis. Here it is argued instead that embryonic development is a convergent increase in complexity that *necessarily* requires organic codes and organic memories, and a few examples of such codes are described. This is the *code theory of development*, a theory that was originally inspired by an algorithm that is capable of reconstructing structures *from incomplete information*, an algorithm that here is briefly summarized because it makes it intuitively appealing how a convergent increase in complexity can be achieved. The main thesis of the new theory is that the presence of organic codes in ontogenesis is not only a theoretical necessity but, first and foremost, an idea that can be tested and that has already been found to be in agreement with the evidence.

1. Epigenesis and design

The most elegant and probably the most influential experiment in embryology was performed by Aristotle some 2400 years ago on chick embryos. Aristotle removed the shell from eggs that had been incubated for different periods of time, and carefully described what he saw: the white spot on the yolk that marks, at the beginning, the position where the embryo is going to appear; the tiny brown lump that begins pulsating on the third day; the protruding bulbs that gradually turn into eyes; the network

THE ROYAL SOCIETY
PUBLISHING

of red vessels that descend into the yolk like the roots of a tree; and the various membranes that wrap up the growing chick [1].

On the basis of these observations, Aristotle concluded that embryonic development is a step-by-step generation of new structures. It is not a single genesis but an *epigenesis*, a sequence of many geneses one after the other. An embryo, in other words, is a system that grows not only in size but also in *number of parts* and therefore in *complexity* (the term complexity is used here in the standard sense as the shortest number of instructions that are necessary to describe a system).

Aristotle underlined that there are *two* distinct features in epigenesis. One is the increase in complexity, the other is that such an increase goes in a very specific direction. A frog embryo, for example, can only develop into a member of the frog species and nothing else. Epigenesis, in other words, is the idea that embryonic development is not just an increase in complexity, it is a *convergent* increase in complexity, and we need therefore to understand how it can be 'channelled' towards a specific result.

To this end, Aristotle proposed a theory that has become the object of countless controversies ever since: the theory that in every living creature there is a *design*. Take a house, for example. When you build one, said Aristotle, you must follow the laws of Nature—those eternal rules that the old philosophers called *necessity*—because nobody, not even the gods, can change them. In order to build a house, however, you need more than that. You need a project, a design: this is the *final* cause of your enterprise, the goal towards which you move in full agreement with the laws of Nature.

The great difference between building a house and building an embryo is that the design of a house comes from a human architect, and therefore *from without* the system, whereas the design of an embryo comes *from within*. Nature, in other words, does not build as we do. She is rational, because she works according to plans, but her designs are *immanent* in her creatures. They do not come from another world, as Plato said; they are an integral part of this world, just as matter is.

The contribution of Aristotle to embryology consists therefore in two great ideas: the *theory of epigenesis*, the doctrine that development is a convergent increase in complexity, and the *theory of design*, the idea that embryonic development is guided by a project that exists in the embryo since the beginning of conception.

The *theory of epigenesis* was officially accepted in biology in 1828, when Karl Ernst von Baer published *On the Development of Animals*, a monumental treatise of comparative embryology that put an end to any version of preformationism by showing that there is a very early stage in the development of all animals where the entire embryo consists in a few sheets, or *germinal layers*, of organic matter [2]. That made it clear that the development of an embryo is not only a process of growth, but also a step-by-step emergence of new structures that increase the complexity of the developing creature.

The theory of design took much longer to be accepted and that happened when Monod [3], Delbrück [4], Mayr [5] and many others pointed out that the immanent design of Aristotle is virtually equivalent to the *genetic programme*.

2. Do we understand development?

To a large extent it is legitimate to say that modern embryology has rediscovered the theories of epigenesis and design proposed by Aristotle. Today, in fact, there is a virtually universal consensus on the idea that *embryonic development is the result of a genetic programme and of epigenetic processes*.

This may give the impression that we now understand, at least in principle, the *logic* of embryonic development, and all that remains to be done is to fill in the details. In *The Problems of Biology*, however, Maynard Smith [6, pp. 99–100] lucidly sounded a note of caution against this attitude:

> It is popular nowadays to say that morphogenesis (that is the development of form) is programmed by the genes. I think that this statement, although in a sense true, is unhelpful.

> Unless we understand how the program works, the statement gives us a false impression that we understand something when we do not . . . . One reason why we find it so hard to understand the development of form may be that we do not make machines that develop: often we understand biological phenomena only when we have invented machines with similar properties . . . and we do not make 'embryo' machines.

In fact, 'understanding' something means explaining it with a model that we are familiar with, and a machine gives us an immediate sense of familiarity. When we see it working before our eyes, we feel that we 'know' it. Actually, we do not even need to build a machine to get this feeling. A description is enough, and so a machine is often a *model*, or an *algorithm*. One of the most famous machines of all times was built by Turing with just pencil and paper.

Maynard Smith's point, in other words, amounts to saying that to understand embryonic development we need a model or an algorithm that allows us to understand, at least in principle, how a system can increase its own complexity in a *convergent* way.

Note that we do have a mathematical model that explains how a *divergent* increase in complexity can take place. That model is provided by the equations of population genetics in the presence of natural selection, and accounts for the fact that in the course of evolution living systems gave origin to the increasingly diverging forms that make up the *tree of life*. But embryonic development is a convergent, not a divergent, increase in complexity, and this is the key point that any theory of development must explain.

As a matter of fact, we do have such a theory. It was proposed in *The Organic Codes* [7], and here it is briefly summarized with a two-step procedure. The first step is to study a phenomenon that has nothing to do with embryonic development but where we have a mathematical model which proves that a convergent increase in complexity does take place. The second step is to look for the most general features of this model and see if we can recognize at least some of them in embryonic development.

3. The reconstruction model

The images produced by X-rays in radiology or in computerized tomography are projections of three-dimensional bodies on two-dimensional surfaces. From a mathematical point of view, the projections are described by integrals and in order to reconstruct the original structure it is necessary to perform the reverse operation from a plurality of projections obtained at different angles. The algorithms that achieve this goal can be either iterative or non-iterative techniques. The non-iterative algorithms perform a single operation on all projection data and belong to different classes because the reconstructions can be achieved with algebraic methods, with deconvolutions in real space, with Fourier transforms, etc.

The iterative algorithms have been introduced in order to get a satisfactory result with simpler procedures. In these cases, a reconstruction is obtained in stages, each of which generates an approximate structure, and it is therefore necessary to repeat the process many times in order to get progressively closer to the original. Iterative methods are less precise than single-application ones, but their great advantage is that they introduce the *time dimension* in the process, and this makes them particularly suitable to simulate biological processes.

A fundamental issue in all reconstruction methods is the *minimum* number of projections that must be collected in order to perform a reconstruction. This number is obtained from basic theorems and has a precise value because the projections must contain (in a condensed form) all the information that was present in the original structure. With algebraic methods, for example, the reconstruction of a structure described by an $n \times n$ matrix amounts to solving a set of equations in $n \times n$ unknowns, and a solution is possible only if the number of (linearly independent) equations is equal to the number of the unknowns.

The number of equations, on the other hand, is obtained from the number of projections, and this allows us to formulate a completely new problem: is a reconstruction possible when the number of equations is *inferior* to the theoretical minimum?

This is the problem of *the reconstruction of structures from incomplete projections*, where the projections are said to be *incomplete* if their number is at least *one order of magnitude* inferior to the theoretical minimum that is required for a complete reconstruction. This minimum, as we have seen, is the number of projections that collectively contain as much information as the original structure, and our problem is therefore equivalent to asking whether it is possible to make a reconstruction with an initial amount of information that is much lower than that of the original structure.

With non-iterative algorithms this is mathematically impossible, but in the case of iterative techniques the situation *could* be different. When a reconstruction is performed in many successive stages we could learn from what takes place during the process and discover new properties that may allow us to compensate for the incomplete information of the initial data.

It will be noted that this is a mathematical version of the problem that we face in embryonic development. The genetic information of the embryo remains constant in the course of ontogenesis (just as the information of the projection data remains constant during a reconstruction) but the overall complexity of the system does increase because of the step-by-step addition of new features. This means that the fertilized egg is less complex than the adult organism and embryonic development is truly a convergent increase in complexity. The reconstruction of structures from incomplete data, on the other hand, is another example, albeit very different, of such an increase, and its study could help us understand the *logic* of what goes on in embryonic development.

4. A memory-based reconstruction method

The mathematical description of a family of iterative algorithms that are capable of reconstructing structures from incomplete information has been given in *The Organic Codes* [7] and will not be repeated here. This paper aims instead at describing the *logic* of these algorithms and will try to make it intuitively appealing with one specific example.

An iterative reconstruction algorithm produces a sequence of pictures that are increasingly more accurate approximations of the original structure. These pictures, on the other hand, are inevitably affected by errors, and in general there is no way of knowing where the errors are falling, except for two cases: the values that are lower than the minimum (0) or greater than the maximum (1) are clearly *illegal*, and it is possible to correct them by setting to the minimum or to the maximum all values which are, respectively, below or above the legal limits. This operation does improve the results and so it is normally applied at regular intervals. In addition to correcting these illegal values, however, we can do something else. We can *keep a record* of them. This can be done by using a *memory matrix* where we store the values of the illegal points before applying the constraint. In this case we need a more complex algorithm, because we must perform in parallel two different reconstructions: one for the structure matrix and one for the memory matrix.

But what is the point of keeping a memory of the errors? The point is that we can study how their patterns change, and this turns out to have unexpected features. As the errors are random, one would expect a completely statistical distribution but this is not always the case. It is true that in many points the errors are totally random, but invariably there are other points where this does not happen. In these points the illegal values keep reappearing without interruptions, and always with the same sign, which explains why they have been called *vortices*. The interesting thing is that they give us a new type of information. When an illegal value keeps reappearing in the same point for a long enough number of times (we can choose 5, 10 or any other convenient number), we can reasonably conclude that that point is either a minimum or a maximum. We can therefore *fix* its value, and this means that the total number of unknowns is reduced by one. By repeating this operation, the number of the unknowns becomes progressively smaller and when it reaches the number of equations a complete reconstruction is possible. That is the result we were looking for. The memory matrix allows us to collect information that was not present in the initial data and which is gradually emerging during the reconstruction process.

Let us now take a closer look at the memory matrix. The information of the vortices appears in the *memory* space, but we use that information in the *structure* space, because it is here that we reduce the number of the unknowns. We are in fact transferring information from the memory space to the structure space with a *conventional rule* of the type 'If a vortex appears in the memory space, fix the corresponding point in the structure space to a minimum or a maximum'.

A reconstruction from incomplete information, in short, does not require only a memory. It requires *memory and codes*. The reconstruction memory is where new information appears. The reconstruction codes are the tools that transfer information from the memory space to the structure space.

5. The code theory of development

Embryonic development is *not* a reconstruction of structures from projections and it is *not* achieved *by* computations, but it is nevertheless a reconstruction *from incomplete information*, and there may well be general principles that apply to all such cases.

When only a fraction of the information is present at the beginning, a reconstruction is possible only if we collect the missing information in a *related but autonomous* space, and the memory space is precisely a world where a system can find the extra information that allows it to increase its own complexity. The memory-based reconstruction method, in other words, leads to a general concept: to the principle that *there cannot be a convergent increase in complexity without memory*.

Another fundamental concept is that information can be transferred from the memory space to the real space only by *conventions.* In order to decrease the number of the unknowns in real space, it is necessary to give a *meaning* to the structures that appear in the memory space, and this too is a conclusion whose validity is absolutely general. Real space and memory space are independent worlds, because if they were equivalent (like real space and Fourier space, for example) they would transport the same information and no increase in complexity would be possible. But between two independent worlds there is no necessary link, and no information can be transferred automatically from one to the other. The only bridge that can establish a link between them is a convention or a code. This amounts to another general principle: *there cannot be a convergent increase in complexity without codes*.

The reconstruction model, in conclusion, gives us four general principles about development:

(1) Epigenesis is a process that allows a system to increase its own complexity in a convergent way.
(2) A convergent increase in complexity is equivalent to a reconstruction from incomplete information.
(3) In *organic systems* there cannot be a convergent increase in complexity without *organic memories*.
(4) In *organic systems* there cannot be a convergent increase in complexity without *organic codes*.

This is the *code theory of development*, the idea that *embryonic development is a convergent increase in complexity that necessarily requires organic codes and organic memories*. This theory is at variance with virtually all present models of development because they do not contemplate the existence of organic codes in ontogenesis.

The code theory was originally proposed with the name of *semantic theory* [7,8] but this name was later changed as a result of a theoretical move from *semantic biology* first to *biosemiotics* and then to *code biology*, a move that was described in detail in [9,10].

What really matters, for our purposes, is that the theory can be tested. One of its basic requirements is that development is based on *organic codes* (codes between organic molecules) and this can be tested because the existence of organic codes in embryonic processes can be proved by the same experimental evidence (the presence of adaptors) that has proved the existence of the genetic code in protein synthesis.

6. The experimental reality of the organic codes

The code theory of development predicts that many organic codes are at work in living systems, and clearly can be accepted only if there is evidence that these codes exist. In 2003 such evidence was already available, but it circulated only in small circles and did not attract the attention of the scientific community at large. More organic codes came to light in the years that followed, again without attracting much attention, and at this stage it may be useful to make a brief summary of what today is known about them.

In 1975, the American biochemist Gordon Tomkins published a paper entitled 'The metabolic code. Biological symbolism and the origin of intercellular communication' [11]. That was the first announcement of a new organic code after the discovery of the genetic code, but tragically Tomkins died that very year and his new world of biological symbolism remained unexplored.

In 1979, David Elder pointed out that the formation of body segments in annelid worms is described by combinatorial rules that in electronics are known as *Gray code*, and proposed that the annelid body plan is based on a biological version of those rules that he referred to as *epigenetic code* [12].

At the end of the 1980s, Edward Trifonov started a life-long campaign in favour of the idea that genomes simultaneously carry several overlapping codes, not just the genetic code, and gave them the collective name of *sequence codes* [13–15].

Finally, in the 1990s and in the early 2000s, a wide variety of new organic codes came to light. Among them: the *Hox* code [16,17], the *adhesive code* [18,19], the *splicing codes* [7,20–26], the *signal transduction codes* [7,21], the *histone code* [27–32], the *sugar code* [33,34], the *compartment codes* [7], the *cytoskeleton codes* [7,35], the *transcriptional code* [36–39], a *universal neural code* [40,41], a *neural code for taste* [42,43], an *odorant receptor code* [44,45], a *space code* in the hippocampus [46–50], the *apoptosis code* [51,52], the *tubulin code* [53], the *nuclear signalling code* [54], the *ubiquitin code* [55], the *bioelectric code* [56,57], the *glycomic code* [58] and the *acoustic codes* [59].

It has to be underlined that codes have been defined in different ways, a problem that is not uncommon in biology, but in our case this is not a major obstacle because there is an operative definition that can be applied to all codes [7]. This definition, furthermore, has been instrumental in the development of mathematical models that can examine any natural system and tell us whether or not it contains organic codes [60–62].

Today, in conclusion, there is ample evidence that organic codes are an integral part of the living world and are involved in a wide variety of biological processes. It remains to be seen, however, whether they are also involved in embryonic development given that most of the present models of ontogenesis do not contemplate any such role for them.

7. Three classical cases

Protein synthesis is an illuminating example of the interplay that exists between genetic and epigenetic processes. The information of a gene is used to generate a linear chain of amino acids, and then a process of folding transforms this one-dimensional chain into a three-dimensional protein. Note that the linear order in a chain of 100 amino acids, for example, is specified by 100 coordinates, whereas their three-dimensional arrangement in a protein requires 300 coordinates (three for each amino acid). Protein folding amounts therefore to adding the 200 missing coordinates to the 100 provided by the gene, and is therefore a true epigenetic process first because it is not written in the genes and second because it generates a genuine *increase in complexity*.

In embryonic development, the distinction between genetic and epigenetic processes is more problematic because we need to prove (a) that there are genes uniquely involved in development and (b) that there are true epigenetic processes in ontogenesis. The great achievement of modern embryology has precisely been the ability to prove these two points.

(a) The pioneering work of Lewis [63,64], Sander [65], Garcia-Bellido [66], Nüsslein-Volhard & Wieschaus [67], Scott & Weiner [68], McGinnis *et al*. [69], Carroll [70], Gehring [71] and many others has shown that there are genes exclusively dedicated to building the body plan in the first stages of development, genes that have been collectively referred to as *the genetic toolkit* [72].

(b) The transformations of ontogenesis are of many different types, but in many cases it has been shown that they are the result of true epigenetic processes. Let us briefly examine three classical examples.

   1. Morphogenesis is a coordinated set of cell movements (migrations, invaginations and tubulations of cell sheets) that achieve in an embryo what folding achieves in a protein: a specific three-dimensional form. In both cases, the final form is the result of physical forces [73,74] whose actions are not written in the genes. According to Beloussov [75], in particular, mechanical stresses are indispensable for organized cell movements and represent the key to morphogenesis.
   2. The classic experiments of Spemann [76] have shown that early embryonic cells can change their histological fate (skin cells, for example, can become nerve cells) if they are transplanted *before* a critical period, but are totally unable to do so if the transplant takes place *after* that period. This means that there is a crucial period of development in which *something* happens that decides the histological fate of the cell and of all its descendants. That something was called *determination* and represents a *cell memory* for generations to come. The key point is that the internal organization of the cell is completely dismantled at each cell division, and yet the daughter cells acquire again the determination state of the mother cell without any new genetic intervention, which means that the state of determination, or cell memory, is transmitted to the descendants by epigenetic means.
   3. During the development of the nervous systems, some neurons send out 'tentacles' that leave the neural tube and move into the body in search of organs that require nerve connections. The neurons do not have any information about their targets but this is compensated by an overproduction of cells, which ensures that some of them will actually reach the targets. At this point another process comes into play. The organs that need to be innervated send off particular molecules, known as *nerve growth factors*, that literally save the neurons from certain death. More precisely, neurons are programmed to commit suicide—i.e. to activate the genes of cell death, or *apoptosis*—at the end of a predetermined period, and nerve growth factors are the only molecules that can switch off this self-destruction mechanism. The result is that the neurons that reach the right places survive, and the others disappear [77–79]. The neural wiring of the body, in other words, is obtained with genetic and epigenetic processes whose final result is the creation of a three-dimensional network whose geographical information is not written in the genes.

8. Organic codes in ontogenesis

Morphogenesis, cell determination and the development of the nervous system have been studied for more than a century now, and, of course, the classical papers could not mention organic codes because these had not yet been discovered. Today, however, we can take a new look at those classical cases.

(a) Morphogenesis

Cells react to mechanical forces and mechanical stress with the standard strategy of signal transduction that they adopt for all signals from the environment. This strategy was discovered when it became clear that the external signals (known as *first messengers*) never reach the genes. They are invariably transformed by molecular receptors into internal signals (called *second messengers*) and only these, or their derivatives, reach the genes. What is still largely ignored is that signal transduction is based on organic codes because we find in it the three essential features that *define* a code: (i) two independent worlds of molecules (first messengers and second messengers), (ii) a set of adaptors (the signal receptors) that create a mapping between them, and (iii) the proof that the mapping is arbitrary because its rules can be implemented in countless different ways [7]. The effects that external signals have on cells, in short, do not depend on what they actually are but on the way in which the cells decode them. Acetylcholine, for example, is a neurotransmitter in the brain, but in the heart it induces relaxation, whereas in skeletal muscles it produces contraction and in the pancreas it has the effect of secreting enzymes [80]. This and countless other examples can be accounted for only by the idea that all signals from the environment, including physical forces and mechanical stress, are decoded by the cell with sets of rules that have been referred to as *signal transduction codes* [7].

(b) Cell determination

In eukaryotes, the DNA filament is folded around groups of eight histone proteins called *nucleosomes*, then it is further folded into spirals of nucleosomes called *solenoids*, and finally into the fibre of the *chromosome*. These multiple foldings allow the eukaryotic cells to pack their chromosomes into the tiny space of the nucleus, and for this reason it was initially assumed that histones have a purely packaging role. The experimental data, however, have shown that the 'tails' of the histones (the parts that protrude from the surface of the nucleosomes) are subject to a wide variety of post-translational modifications (in particular acetylation, methylation and phosphorylation) that have dynamic roles and are involved in the activation or the repression of gene activity [81,82].

Another crucial breakthrough in this field was the discovery that the post-translational modifications of the histones do not act *individually* but in *combinatorial groups*. This led David Allis and colleagues to propose that the histone marks operate like letters and words of a molecular 'language' that was referred to as the *histone code* [27,29].

The same concept was independently proposed by Turner [30,31], who also pointed out that the histone marks have both short- and long-term effects. The short-term modifications change rapidly in response to external signals and represent a mechanism by which the genome quickly responds to the environment [83]. The existence of long-term effects was revealed by the discovery that many histone modifications survive the trauma of mitosis and are transmitted to the daughter cells. They provide in this way a mechanism of *cell memory*, in the sense that they enable the cells to 'remember' their specific pattern of gene expression for many generations [84–87]. It has been shown, for example, that the expression of *Hox* genes in embryonic development is regulated by histone modifications [88,89].

Today, in short, there is good evidence that cell determination and cell memory are based on the combinatorial rules of the histone code [32].

(c) Development of the nervous system

Active cell suicide (apoptosis) is a universal mechanism of embryonic development, one that is used to shape virtually *all* organs of the body. The key point is that suicide genes are present in all cells and the signalling molecules that switch them on and off are of countless different types. This means that there are no necessary connections between the recognition of a signalling molecule and the activation of the suicide genes. They are two independent processes, and the only realistic solution is that the link between them is established by the conventional rules of an *apoptosis code*, i.e. a code that determines which signalling molecules switch on the apoptosis genes in which tissue [51,52].

Another essential contribution to the development of the brain is provided by molecules that bridge the synaptic cleft and decide which neurons are connected and which are not. They function both as synaptic recognizers and synaptic glue, and recently it has been shown that cadherins and protocadherins are good candidates for these roles. Protocadherins, in particular, have an enormous potential for diversification because their genes contain variable and constant regions like the genes of the immunoglobulins. They could therefore provide the building blocks of a neural system that is capable of learning and memorizing, and, like the immune system, can cope with virtually everything, even the unexpected [90]. This strongly suggests the presence of a code because only a code can generate an enormous diversity with a limited number of rules, and this is why various authors have proposed that the wiring of the nervous system is based on an *adhesive code* [18,19].

Today, in conclusion, there is good evidence that organic codes play a very active part in the development of the nervous system.

9. Two fundamental mechanisms

Organic codes are mechanisms that operate in living systems, and in order to properly appreciate their role we must temporarily leave the field of ontogenesis and take a look at the other mechanisms of life.

The mechanisms of evolution, in particular, have been one of the most controversial issues in biology and the great debate about them culminated, in the 1930s and 1940s, in the *modern synthesis*, the theoretical framework where natural selection is regarded as virtually the sole mechanism of evolutionary change. But where does natural selection come from?

Let us start from the fact that the copying of the genes is the elementary act that leads to *heredity*. When the process of copying is repeated indefinitely, however, another phenomenon comes into being. Copying mistakes becomes inevitable and in a world of limited resources not all changes can be implemented, which means that a process of selection is bound to take place. Molecular copying, in short, leads to heredity, and the indefinite repetition of molecular copying in a world of limited resources leads to *natural selection*.

That is how natural selection came into existence. Molecular copying started it and molecular copying has perpetuated it ever since. Which means that *natural selection would be the sole mechanism of evolution if molecular copying were the sole basic mechanism of life*.

The discovery of the genetic code, however, has proved that there are *two* distinct molecular mechanisms at the basis of life, the *copying* of the genes and the *coding* of proteins. The discovery of other organic codes, furthermore, allows us to generalize this conclusion because it proves that coding is not limited to protein synthesis. Life, in other words, is not based on copying alone. It is based on copying and coding, and these two molecular mechanisms give origin to two distinct mechanisms of evolution because an evolutionary mechanism is but the long-term result of a fundamental molecular mechanism.

More precisely, the existence of copying and coding at the molecular level means that there are two distinct types of evolutionary change: *evolution by natural selection*, based on copying, and *evolution by natural conventions*, based on coding [7,8].

This conclusion is reinforced by the fact that there is a very close relationship between organic codes and the great events of macroevolution. The genetic code was a precondition for the origin of life, the signal transduction codes divided the first cells into three primary kingdoms (Archaea, Bacteria and Eukarya), the splicing codes were instrumental to the origin of the nucleus, the histone code provided the working rules of chromatin, and the cytoskeleton codes allowed the Eukarya to perform internal movements, including those of mitosis and meiosis [7]. All these great events were associated with the appearance of new organic codes, and we can understand why. The reason is that a new code brings genuine novelties into existence because it can create arbitrary mappings and generate relationships that have never existed before in the Universe.

The organic codes, in short, have appeared throughout the history of life and have profoundly shaped it, so it is no wonder that they also had a key role in embryonic development.

10. Conclusion

The discovery that there are many other organic codes in life in addition to the genetic code has not received much attention, so far, and, therefore, it has not modified the traditional view of biology. As a result, we still have a theoretical framework that contemplates only two codes in Nature: the genetic code that appeared at the origin of life and the codes of culture that arrived almost 4 billion years later. Which amounts to saying that there have been no other codes in between, and therefore that codes are *extraordinary exceptions*, not *normal components* of life. Such a situation can perhaps be illustrated with an example from the history of science.

In the eighteenth century, gravity was the only known force in the Universe and electricity was confined to the extravagant phenomena that were observed when some materials were rubbed together or when a kite was sent up into the sky during a storm to catch lightning. It took a long time to change this view, but eventually it became clear that gravitation and electromagnetism are both universal forces and fundamental aspects of reality.

In biology, we are today at the stage at which physics was at the time of Benjamin Franklin, when it became clear that electricity is everywhere in Nature and yet it was considered an extravagance. In a similar way, the dominant view today is that molecular copying is the sole fundamental mechanism of life and the genetic code is an extraordinary singularity.

What is coming to light, instead, is that molecular copying and molecular coding (natural selection and natural conventions) are equally fundamental mechanisms. The world of molecular coding, in other words, is the other side of life that so far has remained hidden from sight just as electromagnetism was in the times of Newton and Franklin.

It should be added that mathematical models did have a role in the rise of electromagnetism, and it was in fact those models that *illuminated* the experimental data. The model of the reconstruction of structures from incomplete information has nothing to do with the flesh and blood of embryonic development, and yet it is the only one, so far, that makes us *understand* how a convergent increase in complexity can take place.

Competing interests. I declare that I have no competing interests.

Funding. I received no funding for this study.

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

References

1. Aristotle [384–322 BC]. 1965 *De generatione animalium* (transl. by Arthur Platt). Oxford, UK: Clarendon Press.
2. Von Baer KE. 1828 *Über Entwickelungsgeschichte der Thiere: Beobachtung und Reflexion*. Königsberg, Germany: Gebrüder Bornträger.
3. Monod J. 1970 *Le hasard et la necessité*. Paris, France: Seuil. (Transl. 1971 *Chance and necessity*. New York, NY: A Knopf.)
4. Delbrück M. 1971 Aristotle-totle-totle. In *Of microbes and life* (eds J Monod, E Borek). New York, NY: Columbia University Press.
5. Mayr E. 1982 *The growth of biological thought*. Cambridge, MA: Harvard University Press.
6. Maynard Smith J. 1986 *The problems of biology*. Oxford, UK: Oxford University Press.
7. Barbieri M. 2003 *The organic codes. An introduction to semantic biology*. Cambridge, UK: Cambridge University Press.
8. Barbieri M. 1985 *The semantic theory of evolution*. London, UK: Harwood Academic Publishers.
9. Barbieri M. 2014 From biosemiotics to code biology. *Biol. Theory* **9**, 239–249. (doi:10.1007/s13752-013-0155-6)

10. Barbieri M. 2015 *Code biology. A new science of life*. Dordrecht, The Netherlands: Springer.
11. Tomkins MG. 1975 The metabolic code. Biological symbolism and the origin of intercellular communication. *Science* **189**, 760–763. (doi:10.1126/science.169570)
12. Elder D. 1979 An epigenetic code. *Differentiation* **14**, 119–122. (doi:10.1111/j.1432-0436.1979.tb01020.x)
13. Trifonov EN. 1987 Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16s rRNA nucleotide sequence. *J. Mol. Biol.* **194**, 643–652. (doi:10.1016/0022-2836(87)90241-5)
14. Trifonov EN. 1989 The multiple codes of nucleotide sequences. *Bull. Math. Biol.* **51**, 417–432. (doi:10.1007/BF02460081)
15. Trifonov EN. 1999 Elucidating sequence codes: three codes for evolution. *Ann. NY Acad. Sci.* **870**, 330–338. (doi:10.1111/j.1749-6632.1999.tb08894.x)
16. Hunt P *et al*. 1991 The branchial *Hox* code and its implications for gene regulation, patterning of the nervous system and head evolution. *Development* **2**, 63–77.
17. Kessel M, Gruss P. 1991 Homeotic transformation of murine vertebrae and concomitant alteration of *Hox* codes induced by retinoic acid. *Cell* **67**, 89–104. (doi:10.1016/0092-8674(91)90574-I)
18. Readies C, Takeichi M. 1996 Cadherins in the developing central nervous system: an adhesive code for segmental and functional subdivisions. *Dev. Biol.* **180**, 413–423. (doi:10.1006/dbio.1996.0315)
19. Shapiro L, Colman DR. 1999 The diversity of cadherins and implications for a synaptic adhesive code in the CNS. *Neuron* **23**, 427–430. (doi:10.1016/S0896-6273(00)80796-5)
20. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencow BJ, Frey BJ. 2010 Deciphering the splicing code. *Nature* **465**, 53–59. (doi:10.1038/nature09000)
21. Barbieri M. 1998 The organic codes. The basic mechanism of macroevolution. *Riv. Biol.* **91**, 481–514.
22. Dhir A, Buratti E, van Santen MA, Lührmann R, Baralle FE. 2010 The intronic splicing code: multiple factors involved in ATM pseudoexon definition. *EMBO J.* **29**, 749–760. (doi:10.1038/emboj.2009.397)
23. Fu XD. 2004 Towards a splicing code. *Cell* **119**, 736–738. (doi:10.1016/j.cell.2004.11.039)
24. Matlin A, Clark F, Smith C. 2005 Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386–398. (doi:10.1038/nrm1645)
25. Pertea M, Mount SM, Salzberg SL. 2007 A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC Bioinform.* **8**, 159. (doi:10.1186/1471-2105-8-159)
26. Wang Z, Burge C. 2008 Splicing regulation: from a part list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813. (doi:10.1261/rna.876308)
27. Jenuwein T, Allis CD. 2001 Translating the histone code. *Science* **293**, 1074–1080. (doi:10.1126/science.1063127)
28. Kühn S, Hofmeyr J-HS. 2014 Is the 'histone code' an organic code? *Biosemiotics* **7**, 203–222. (doi:10.1007/s12304-014-9211-2)
29. Strahl BD, Allis CD. 2000 The language of covalent histone modifications. *Nature* **403**, 41–45. (doi:10.1038/47412)
30. Turner BM. 2000 Histone acetylation and an epigenetic code. *BioEssays* **22**, 836–845. (doi:10.1002/1521-1878(200009)22:9<836::AID-BIES9>3.0.CO;2-X)
31. Turner BM. 2002 Cellular memory and the histone code. *Cell* **111**, 285–291. (doi:10.1016/S0092-8674(02)01080-2)
32. Turner BM. 2007 Defining an epigenetic code. *Nat. Cell Biol.* **9**, 2–6. (doi:10.1038/ncb0107-2)
33. Gabius H-J. 2000 Biological information transfer beyond the genetic code: the sugar code. *Naturwissenschaften* **87**, 108–121. (doi:10.1007/s001140050687)
34. Gabius H-J. 2009 *The sugar code. Fundamentals of glycosciences*. Hoboken, NJ: Wiley-Blackwell.
35. Gimona M. 2008 Protein linguistics and the modular code of the cytoskeleton. In *The codes of life: the rules of macroevolution* (ed. M Barbieri), pp. 189–206. Dordrecht, The Netherlands: Springer.
36. Flames N, Pla R, Gelman DM, Rubenstein JLR, Puelles L, Marìn O. 2007 Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695. (doi:10.1523/JNEUROSCI.2750-07.2007)
37. Jessell TM. 2000 Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Genet.* **1**, 20–29. (doi:10.1038/35049541)

38. Marquard T, Pfaff SL. 2001 Cracking the transcriptional code for cell specification in the neural tube. *Cell* **106**, 651–654. (doi:10.1016/S0092-8674(01)00499-8)
39. Ruiz i Altaba A, Nguien V, Palma V. 2003 The emergent design of the neural tube: prepattern, SHH morphogen and GLI code. *Curr. Opin. Genet. Dev.* **13**, 513–521. (doi:10.1016/j.gde.2003.08.005)
40. Nicolelis M. 2011 *Beyond boundaries: the new neuroscience of connecting brains with machines and how it will change our lives*. New York, NY: Times Books.
41. Nicolelis M, Ribeiro S. 2006 Seeking the neural code. *Sci. Am.* **295**, 70–77. (doi:10.1038/scientificamerican1206-70)
42. Di Lorenzo PM. 2000 The neural code for taste in the brain stem: response profiles. *Physiol. Behav.* **69**, 87–96. (doi:10.1016/S0031-9384(00)00191-8)
43. Hallock RM, Di Lorenzo PM. 2006 Temporal coding in the gustatory system. *Neurosci. Behav. Rev.* **30**, 1145–1160. (doi:10.1016/j.neubiorev.2006.07.005)
44. Dudai Y. 1999 The smell of representations. *Neuron* **23**, 633–635. (doi:10.1016/S0896-6273(01)80020-9)
45. Ray A, van der Goes van Naters W, Shiraiwa T, Carlson JR. 2006 Mechanisms of odor receptor gene choice in *Drosophila*. *Neuron* **53**, 353–369. (doi:10.1016/j.neuron.2006.12.010)
46. Brandon MP, Hasselmo ME. 2009 Sources of the spatial code within the hippocampus. *Biol. Rep.* **1**, 3–7. (doi:10.3410/b1-3)
47. Hafting T, Fyhn M, Molden S, Moser MB, Moser EI. 2005 Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806. (doi:10.1038/nature03721)
48. O'Keefe J, Burgess N. 1996 Geometric determinants of the place fields of hippocampal neurons. *Nature* **381**, 425–428. (doi:10.1038/381425a0)
49. O'Keefe J, Burgess N. 2005 Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* **15**, 853–866. (doi:10.1002/hipo.20115)
50. Papoutsi M, de Zwart JA, Jansma JM, Pickering MJ, Bednar JA, Horwitz B. 2009 From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cereb. Cortex* **19**, 2156–2165. (doi:10.1093/cercor/bhn239)
51. Basañez G, Hardwick JM. 2008 Unravelling the Bcl-2 apoptosis code with a simple model system. *PLoS Biol.* **6**, e154. (doi:10.137/journal.pbio.0060154)
52. Füllgrabe J, Hajji N, Joseph B. 2010 Cracking the death code: apoptosis-related histone modifications. *Cell Death Differ.* **17**, 1238–1243. (doi:10.1038/cdd.2010.58)
53. Verhey KJ, Gaertig J. 2007 The tubulin code. *Cell Cycle* **6**, 2152–2160. (doi:10.4161/cc.6.17.4633)
54. Maraldi NM. 2008 A lipid-based code in nuclear signalling. In *The codes of life: the rules of macroevolution* (ed. M Barbieri), pp. 207–221. Dordrecht, The Netherlands: Springer.
55. Komander D, Rape M. 2012 The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229. (doi:10.1146/annurev-biochem-060310-170328)
56. Levin M. 2014 Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. *J. Physiol.* **592**, 2295–2305. (doi:10.1113/jphysiol.2014.271940)
57. Tseng AS, Levin M. 2013 Cracking the bioelectric code. Probing endogenous ionic controls of pattern formation. *Commun. Integr. Biol.* **6**, 1–8. (doi:10.4161/cib.22595)
58. Buckeridge MS, De Souza AP. 2014 Breaking the 'glycomic code' of cell wall polysaccharides may improve second-generation bioenergy production from biomass. *Bioenergy Res.* **7**, 1065–1073. (doi:10.1007/s12155-014-9460-6)
59. Farina A, Pieretti N. 2014 Acoustic codes in action in a soundscape context. *Biosemiotics* **7**, 321–328. (doi:10.1007/s12304-014-9213-0)
60. De Beule J, Hovig E, Benson M. 2011 Introducing dynamics into the field of biosemiotics. *Biosemiotics* **4**, 5–24. (doi:10.1007/s12304-010-9101-1)
61. Görlich D, Artmann S, Dittrich P. 2011 Cells as semantic systems. *Biochim. Biophys. Acta* **1810**, 914–923. (doi:10.1016/j.bbagen.2011.04.004)
62. Görlich D, Dittrich P. 2013 Molecular codes in biological and chemical reaction networks. *PLoS ONE* **8**, e54694. (doi:10.1371/journal.pone.0054694)
63. Lewis EB. 1963 Genes and developmental pathways. *Am. Zool.* **3**, 33–56. (doi:10.1093/icb/3.1.33)
64. Lewis EB. 1978 A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570. (doi:10.1038/276565a0)

65. Sander K. 1975 Pattern specification in the insect embryo. In *Cell patterning* (eds R Porter, J Rivers). CIBA Foundation Symp., vol. 29, pp. 241–263. Amsterdam, The Netherlands: Elsevier.
66. Garcia-Bellido A. 1975 A genetic control of wing disc development in Drosophila. In *Cell patterning* (eds R Porter, J Rivers). CIBA Foundation Symp., vol. 29, pp. 161–178. Amsterdam, The Netherlands: Elsevier.
67. Nüsslein-Volhard C, Wieschaus E. 1980 Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801. (doi:10.1038/287795a0)
68. Scott MP, Weiner AJ. 1984 Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci in Drosophila. *Proc. Natl Acad. Sci. USA* **81**, 4115–4119. (doi:10.1073/pnas.81.13.4115)
69. McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ. 1984 A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and Bithorax complexes. *Nature* **308**, 428–433. (doi:10.1038/308428a0)
70. Carroll SB. 1995 Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485. (doi:10.1038/376479a0)
71. Gehring WJ. 1996 The master control gene for morphogenesis and evolution of the eye. *Genes Cells* **1**, 11–15. (doi:10.1046/j.1365-2443.1996.11011.x)
72. Carroll SB, Grenier JK, Weatherbee SD. 2001 *From DNA to diversity. Molecular genetics and the evolution of animal design*. Malden, MA: Blackwell.
73. Beloussov LV. 1998 *The dynamic architecture of a developing organism*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
74. Forgacs G, Newman SA. 2005 *Biological physics of the developing embryo*. Cambridge, UK: Cambridge University Press.
75. Beloussov LV. 2008 Mechanically based generative laws of morphogenesis. *Phys. Biol.* **5**, 015009. (doi:10.1088/1478-3975/5/1/015009)
76. Spemann H. 1938 *Embryonic development and induction*. New Haven, CT: Yale University Press.
77. Changeaux J-P. 1983 *L'Homme neuronal*. Paris, France: Libraire Arthème Fayard.
78. Levi-Montalcini R. 1975 NGF: an uncharted route. In *The neurosciences—paths of discoveries* (ed. FG Worden). Cambridge, MA: MIT Press.
79. Levi-Montalcini R. 1987 The nerve growth factor 35 years later. *Science* **237**, 1154–1162. (doi:10.1126/science.3306916)
80. Sporn MB, Roberts AB. 1988 Peptide growth factors are multifunctional. *Nature* **332**, 217–219. (doi:10.1038/332217a0)
81. Kornberg RD, Lorch Y. 1999 Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294. (doi:10.1016/S0092-8674(00)81958-3)
82. Wu J, Grunstein M. 2000 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* **25**, 619–623. (doi:10.1016/S0968-0004(00)01718-7)
83. Schreiber SL, Bernstein BE. 2002 Signaling network model of chromatin. *Cell* **111**, 771–778. (doi:10.1016/S0092-8674(02)01196-0)
84. Jeppesen P. 1997 Histone acetylation: a possible mechanism for the inheritance of cell memory at mitosis. *BioEssays* **19**, 67–74. (doi:10.1002/bies.950190111)
85. Lyko F, Paro R. 1999 Chromosomal elements conferring epigenetic inheritance. *BioEssays* **21**, 824–832. (doi:10.1002/(SICI)1521-1878(199910)21:10<824::AID-BIES4>3.0.CO;2-U)
86. Margueron R, Trojer P, Reinberg D. 2005 The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176. (doi:10.1016/j.gde.2005.01.005)
87. Xin L, Zhou G-L, Song W, Wu X-S, Wei G-H, Hao D-L, Lv X, Liu D-P, Liang C-C. 2007 Exploring cellular memory molecules marking competent and active transcriptions. *BioMed Central Mol. Biol.* **8**, 31–40. (doi:10.1186/1471-2199-8-31)
88. Wang GG, Cai L, Pasillas MP, Kamps MP. 2007 NUP98–NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.* **9**, 804–812. (doi:10.1038/ncb1608)
89. Zhu B, Zheng Y, Pham A-D, Mandal SS, Erdjument-Bromage H, Tempst P, Reinberg D. 2005 Monoubiquitination of human histone H2B: the factors involved and their roles in *HOX* gene regulation. *Mol. Cell* **20**, 601–611. (doi:10.1016/j.molcel.2005.09.025)
90. Hilschmann N, Barnikol HU, Barnikol-Watanabe S, Götz H, Kratzin H, Thinnes FP. 2001 The immunoglobulin-like genetic predetermination of the brain: the protocadherins, blueprint of the neuronal network. *Naturwissenschaften* **88**, 2–12. (doi:10.1007/s001140000190)