



The Idea of Information in Biology

Author(s): John Maynard Smith

Source: *The Quarterly Review of Biology*, Vol. 74, No. 4 (Dec., 1999), pp. 395-400

Published by: [The University of Chicago Press](#)

Stable URL: <http://www.jstor.org/stable/2664718>

Accessed: 04/10/2013 09:11

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *The Quarterly Review of Biology*.

<http://www.jstor.org>



THE IDEA OF INFORMATION IN BIOLOGY

JOHN MAYNARD SMITH

*School of Biological Sciences, University of Sussex
Falmer, Brighton BN1 9QG, United Kingdom*

TRANSSCRIPTION, translation, code, library, message, editing, proof-reading—any textbook account of molecular biology will define these terms, and any discussion among professionals will employ them. All are concerned with the transmission and storage of information. They have precise and well understood meanings in biology, and the parallels between the biological and colloquial meanings are rather close. For example, in proof-reading a copy is compared, symbol by symbol, with an original, and if any discrepancy is found the copy is altered so that it corresponds with the original.

In our book, *The Major Transitions in Evolution*, Eors Szathmáry and I (Maynard Smith and Szathmáry 1995) suggested that there have been, in the course of evolution, a series of changes in the way in which information is stored and transmitted—for example, the origin of chromosomes, of the genetic code, of sex, of multicellular organisms, and of language. Each transition was the precondition for the evolution of further complexity. Few biologists, I think, would wish to disagree with this picture, although they might, with good reason, wish to argue with our account of the mechanisms underlying particular transitions.

These two examples show that notions of information are now pervasive in biology. Yet, with a few honourable exceptions (e.g., Sarkar 1996; Godfrey-Smith 1999; Sterelny and Griffiths 1999), philosophers of biology either ignore the concept of information, or argue that it is irrelevant and misleading (Mahner and Bunge 1997). So, if philosophers are unwilling to analyse the notion of information in biology, it will have to be done by biologists. Jablonka

(in press) has recently made such an attempt. In particular, she reviews the different mechanisms of heredity observed in the living world. This essay is my own contribution. In the main, it is a discussion of how the idea has been applied in genetics, evolution and developmental biology: I am not competent to discuss recent applications of Information Theory in neurobiology.

WEISMANN AND THE TRANSMISSION OF INFORMATION

Biologists have never been slow to draw analogies between contemporary engineering and the living body. The twentieth century has seen the development of ever more complicated machines that transduce information, starting with the electric telegraph and ending with our present immersion in the internet. Not surprisingly, biologists have drawn analogies from these developments. The first, and perhaps the most important and illuminating, was drawn by August Weismann towards the end of the last century. Most biologists who have thought about the matter at all believe that he rejected the inheritance of acquired characters because of the segregation of germ line and soma (and because, when he cut the tails off mice, their babies were born with normal tails). Indeed, I learnt this as a schoolboy from Shaw's *Back to Methuselah*, and believed it for many years. But nothing can be further from the truth.

In *The Evolution Theory*, Weismann (1904) devotes two chapters to this topic. The one argument not mentioned is the segregation of germ line and soma; that was important to him

in another context. It would in any case be a poor argument: there is no such segregation in plants, yet they are no more likely than animals to transmit acquired characters through the seed. He presents several lines of evidence that animals evolve adaptations that could never have appeared first as individual modifications during development: for example, i) the cuticle of insects is hardened before use, and so cannot adapt in the lifetime of an individual, and ii) female moths lay their eggs on suitable food plants, but are dead before they hatch, and so cannot learn the plants' suitability. But this only shows that the inheritance of acquired characters is not necessary for adaptive evolution—not that it does not happen. His crucial argument is that there is no plausible mechanism whereby acquired traits can be transmitted. Suppose, for example, that an animal learns a useful behaviour: this could only be converted into a genetically programmed instinct if the changes in neural connections in the brain responsible for learning could alter the germ cells so that, in the next generation, the brain developed with the altered connections already in place. He cannot see how this could happen. To drive home his point, he writes "... which is very like supposing that an English telegram to China is there received in the Chinese language."

This is an extraordinary remark, for two reasons. First, it makes explicit use of the information analogy, so far as I know for the first time in biology. (In fact, he uses the translation analogy twice, in slightly different forms.) He understands that heredity is about the transmission of information, and not just matter and energy. I have to confess that I did not see this clearly until after the relation between DNA and protein had been elucidated.

The second reason why the remark is extraordinary is that, although the information analogy led him to one of the important insights in the history of biology, the argument is in a sense fallacious. If a germ cell can influence the development of a brain, why cannot the reverse be true? Some information-transducing machines are irreversible (one cannot cut a record by shouting at a record player), but others are not (a tape recorder works both ways). Given a distinction between genotype and phenotype, the non-inheritance of ac-

quired characters is a contingent, not a necessary fact. Today we would explain it in terms of the "central dogma" of molecular biology: one can introduce a novel protein into a cell, but this will not cause the appearance of a changed DNA molecule able to code for the novel protein. But it is, I think, possible to conceive of a "reverse translataase" which would do just that: it happens not to exist. Would such a reverse translataase solve Weismann's problem, and make possible the inheritance of acquired characters? I think not. The problem is that an adaptive change in phenotype—for example, a learnt behaviour—is typically not represented by one or a few novel proteins. DNA and protein are connected by a code, which could in principle be decoded. The relation between DNA and phenotype is more complex: I will return to this problem.

QUANTIFYING EVOLUTION

I want now to turn to a less successful use of the information analogy, but one from which I at least learnt something. In the late 1950s, stimulated by the development of a quantitative theory of information by Shannon (1948), I had the idea that one could use the theory to quantify evolution. My idea was that three levels of the evolutionary process—genotype, selection, and phenotype—could be quantified in the same units, bits. The genotype is trivial: two bits per base. The matter is complicated by the presence of non-coding DNA, and by the redundancy of the code, but these can be allowed for: an approximate estimate of the amount of information that can be carried by a genome of a given size can be made.

Selection is a bit more difficult, but not hopeless. How much information, in bits, can one program into the genome by selectively removing part of the population each generation? I reasoned as follows. Suppose that in one generation half the population dies for selective reasons: this can be treated as one "bit of selection." Then, starting from a random population of sequences, one could program one base every two generations. The snag is, of course, that evolution does not start from a random set of sequences. Typically, the starting point will be the duplication of a gene, or group of genes. This adds no information: to say something twice adds nothing. New in-

formation requires the alteration of some, but not all, nucleotides. The fact that only some nucleotides need be altered reduces the “bits of selection” needed. However, to alter one base requires that an initially rare nucleotide, arising by mutation, increase in frequency to fixation, a process requiring many bits. At this stage in my thoughts, I was already beginning to wonder whether a useful quantification would emerge. However, Kimura (1961), who must have been thinking along similar lines, did suggest how the problem of selection might be given a quantitative solution in terms of bits.

The real difficulty, however, lies in quantifying the information in the phenotype. Clearly, it is not necessary to specify the exact nature and position of every atom in the adult organism, because not every atom is genetically determined. Imagine a pair of monovular twins. What is genetically specified is what they have in common. Suppose, then, that photographs, represented as black and white dots, or pixels, are taken of the twins (the fact that the twins are three-dimensional, and need more than two colours to represent, are difficulties, but not insoluble). If the size of the pixels is too small, the two photographs would not be identical. So increase the pixel size until the twins *are* identical. Then the information required is simply the number of pixels.

It is only necessary to describe this process to see what is wrong with it. Consider the information needed to specify two different patterns: a random pattern of dots, and a black circle on a white ground. The random pattern would indeed require a number of bits equal to the number of pixels—a point that has since been used to define the word “random.” The circle requires only the information “If $(x-a)^2 + (y-b)^2 < r^2$, then black; else white,” where r is the radius of the circle, and a , b the coordinates of its centre. In other words, if a pattern can be generated by a simple rule, little information is needed to specify it; in any case, genes could specify a , b and r . One might object that a developing organism knows no coordinate geometry, but that is to miss the point. Given the laws of physics and chemistry, and an appropriate environment, genes can generate forms by specifying, for example, rates of reaction and diffusion (Turing 1953).

Wisely, I never attempted to publish these

musings. However, they did teach me one thing. The genotype is not a description of the adult organism, but a recipe for making one, given an environment, initial conditions, and the laws of physics. Until one knows the physical processes that generate an organism, one has no chance of quantifying the genetic information needed to specify it.

GENOTYPE AND PHENOTYPE

Information enters biology when there is a distinction between genotype and phenotype. When did this arise? It is now widely agreed that the present DNA-protein world was preceded by an “RNA world” (Joyce 1989), in which the same RNA molecules were both replicator and enzyme. Such entities would meet Muller’s (1966) definition of living: they possessed multiplication, variation and heredity, and hence were capable of adaptive evolution. Selection would favour those molecules whose enzymic activity assisted their own replication. But there was no genetic code—one cannot code for oneself—and hence no transmission of information.

In all existing organisms there is a genotype-phenotype distinction, and a division of labour between DNA and protein, linked by the genetic code. Is a genetic code essential for the evolution of complex organisms? Perhaps not. Godfrey-Smith (1999) has imagined a world without nucleic acids, in which proteins replicate themselves directly. He supposes the existence of 20 “linkers,” each with two identical ends binding to the same specific amino acid. Given such linkers, amino acid sequence could be replicated without need for a code. Such an imaginary world enables us to ask questions about the role of information in biology. Clearly, there is no genetic code, but what of development? A complex multicellular organism in “protein world” would, presumably, need to have a “master” set of proteins, corresponding to the genome, in every cell, and some means of making multiple copies of particular proteins in particular places. In other words, although there would be no genetic code, there would be an analogue of the “genetic program” that controls our own development. The snag with the phrase “genetic program” is that it conceals our ignorance. The control of development is not very

like the kind of computer program that I write; for one thing, it involves massively parallel processing. There is, however, one feature of the control of development which closely resembles both a computer program, and verbal instructions. This is the symbolic nature of the process, which I want now to discuss.

THE SYMBOLIC NATURE OF DEVELOPMENTAL REGULATION

The origin of our ideas about how development is controlled lies in the analysis of the *lac* operon by Jacob and Monod (1961). The philosophical implications of this analysis were beautifully explained in Monod's *Chance and Necessity* (1971). The genes of the *lac* operon, which enable the cell to metabolise lactose, are switched on by the presence of lactose in the medium: lactose is therefore referred to as an "inducer." The inducer binds to a regulatory protein and alters its shape, thereby preventing it from binding to a "promoter" (a length of DNA immediately upstream of the lactose-metabolising genes). This in effect "liberates" the *lac* genes, or "switches them on." The essential point is that, because of the allosteric nature of the reaction, it would be possible, by altering the gene that codes for the regulatory protein, to alter that protein so that some other inducer—tryptophan, say—switched on the *lac* genes. Of course, such a change would be disadvantageous. But it is natural selection that is responsible for the fact that the genes are switched on by lactose; there is no chemical necessity about it.

Monod called this arbitrary nature of gene regulation "gratuity." I think his meaning would have been better expressed by saying that gene regulation is "symbolic." A symbol is defined as a signal whose meaning is unrelated to its form. Words (with the exception of a few special cases like "cuckoo") are symbolic in this sense; "horse" means a single-toed domestic ungulate, but this is conventional; it could equally well mean a hinged opening, and "door" could mean the ungulate. This symbolic character is even more obvious of the variables we use in a computer program; we choose what x and y are to mean. It is now clear that development is controlled, as the *lac* operon is controlled, by symbolic signals. I suspect that any system capable of controlling

a system as complex as metazoan development has to be symbolic.

THE NATURE OF THE LIVING WORLD

From the viewpoint of information, the living world today has certain very general features:

- i) There is a distinction between genotype and phenotype, and hence separate processes of replication, whereby information is transmitted between generations, and of development, whereby genetic information specifies adult form.
- ii) Heredity is unlimited: that is, the genetic system is capable of transmitting an indefinitely large number of different messages.
- iii) Heredity is modal but development is global: that is, the genetic information can be altered one nucleotide at a time, and can be replicated in the changed form, but there is no one-to-one mapping between particular nucleotides, codons or genes and particular parts of the body. Note that, in contrast, in a computer-generated picture, there is such a one-to-one mapping between the electric impulses travelling along the wire connecting the computer to the printer, and elements of the picture.
- iv) Heredity is digital and not analog (in the sense of varying continuously). It is interesting that speech is analog, in that it consists of a series of sounds whose frequency and amplitude can vary continuously, but it is interpreted digitally, in that speakers of a given language interpret the sounds they hear as being one of a small number of phonemes.
- v) Development is symbolic. This is true at two stages of the process: the code connecting DNA to protein is arbitrary, and the meanings of regulatory signals are also arbitrary.

It is tempting to argue that these features are necessary for the evolution of complex organisms, and will turn out to be true of any that have evolved elsewhere in the universe. But this may merely reflect the fact that we lack the ability to imagine a world fundamentally different from the one we know.

CONCLUSION

Today, in our own species, information is transmitted between generations, and between individuals, by language as well as by genes. It is often said that cultural inheritance is "Lamarckian," but I think this misses the point. The snag with Lamarckian inheritance as a basis for evolutionary change is not just that it is hard to conceive of a mechanism whereby it could occur; this is a problem, but not I think insuperable. The real snag is that most acquired characters are not adaptive—they are the consequences of injury and ageing. Language plus learning provides a mechanism whereby, not only are acquired characters transmitted, but there is a selective process ensuring that the characters transmitted are an adaptive subset of the characters acquired. There is of course room for argument about the extent to which the cultural traits that we have acquired are really of benefit to us, let alone to the society of which we are members. But they are certainly not a random sample of the traits that our predecessors acquired. The rapidity of cultural as compared to genetic change is not due to the superiority of language over the genetic code as a means of transmitting information; indeed, they have much in common. It is because learning, combined with language, makes possible a kind of "adaptive Lamarckism."

I fear that much of what I have said is obvious, not to say banal. Why, then, has the notion of information in biology been ignored or rejected by philosophers? Partly it is because,

outside neurobiology, biologists have made little use of Shannon's quantitative theory of information; one does not need to be a mathematician to work out that, if one is to specify a sequence of 20 kinds of amino acid with four nucleotides, one needs a triplet code. But the almost-deterministic nature of replication and translation does not make the concept of information irrelevant—it just makes it easy. After all, one cannot assert that a typed message carries no information merely because typing errors are rare.

I think that the essential reason for the rejection of biological information is that, in colloquial speech, a "message" implies an intelligent sender, and the receipt of information implies an intelligent receiver. Where, in genetics, are the intelligent senders and receivers? I have discussed this question at some length elsewhere (Maynard Smith, in press). For biologists, however, the answer is so obvious as hardly to need stating. When we speak of a "selfish gene," we do not imagine a conscious agent with self-interested motives. We imagine a chemical entity which has evolved by natural selection to produce effects that ensure its own replication (or the replication of other entities with the same sequences as itself). In the same way, when we speak of a gene, or RNA molecule, as carrying information, we mean that it has adaptive consequences (for example, it causes the synthesis of a functional protein), and second, that it does so because its sequence is the result of past natural selection. In biology, the notions of meaning and intelligence are replaced by those of function and natural selection.

REFERENCES

- Godfrey-Smith P. In Press. On the theoretical role of "genetic coding." *Philosophy of Science*.
- Jablonka E. In Press. The systems of inheritance. *Cycles of Contingency*, edited by Oyama et al. Cambridge (MA): MIT Press.
- Jacob F, Monod J. 1961. On the regulation of gene activity. *Cold Spring Harbor Symposium on Quantitative Biology* 26:193-211.
- Joyce G G. 1989. RNA evolution and the origins of life. *Nature* 338:217-224.
- Kimura M. 1961. Natural selection as a process of accumulating genetic information in adaptive evolution. *Genetical Research* 2:127-140.
- Mahner M, Bunge M. 1997. *Foundations of Biophilosophy*. New York: Springer.
- Maynard Smith J. In Press. The concept of information in biology.
- Maynard Smith J, Szathmáry E. 1995. *The Major Transitions in Evolution*. Oxford: W. H. Freeman.
- Monod J. 1971. *Chance and Necessity*. New York: Knopf.
- Muller H.J. 1966. The gene material as the initiator and organizing basis of life. *American Naturalist* 100:493-517.
- Sarkar S. 1996. Biological information: a skeptical look at some central dogmas of molecular biology. Pages 187-231 in *The Philosophy and History*

- of Molecular Biology: New Perspectives*, edited by S Sarkar. Boston: Kluwer Academic Publishers.
- Shannon C E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379-423; 623-656.
- Sterelny K, Griffiths P E. 1999. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press.
- Turing A M. 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London Series B*:237:37-72.
- Weismann A. 1904. *The Evolution Theory*, translated by J A Thomson, M R Thomson. London: Edward Arnold.