To appear as "Biological Information" in The Philosophy of Science: An Encyclopedia, J.Pfieffer and S.Sarkar (Eds.). New York: Routledge

Biological Information.

Stephen M. Downes
Philosophy Department
University of Utah.

0. Introduction.

Information is invoked by biologists in numerous contexts. Animal behaviorists examine the signaling between two organisms or attempt to delimit the structure of the internal map that guides an organism's migration. Neurobiologists refer to the information passed along neurons and across synapses in brains and nervous systems. The way in which information terminology is used in these contexts is not the main critical focus of philosophers. Philosophers of mind discuss animals' representation systems, such as bees' internal maps, and also focus on the way in which brains operate with a view to shedding light on traditional problems in the philosophy of mind. In contrast, the focus of much heated discussion in philosophy of biology is the notion of information invoked in explaining heredity and development: genetic information. The focus in this article will be on this latter form of biological information.

The ideas that genes are bearers of information or contain programs that guide organisms' development are pervasive ones. These ideas are adhered to explicitly and implicitly by biologists and popularizers of biology. The ideas are so pervasive in biology they may seem hardly worth examining or questioning. Consulting any biology textbook will reveal that genes contain information in the form of DNA sequences and that this information provides instructions for the production of phenotypes. In contrast, an examination of the philosophical literature on genetic information reveals that there are very few philosophers of biology who promote unqualified versions of either of these ideas. To understand how this situation has arisen requires first looking at the role informational concepts play in biology. To do this some concepts in molecular biology are briefly introduced and a brief history of the use of information concepts in biology illustrates their pervasiveness. Second, important aspects of the philosophical discussion of information concepts in biology is presented.

1. A Brief Look at some of the Relevant Biology.

Biologists are interested in two important processes: evolution and development. Evolutionary biologists attempt to account for the process of evolutionary change, including speciation and changes to organisms through time within a species. There was much progress in conceptualizing evolutionary change when it was characterized in terms of changing gene frequencies in the 1930's and 1940's. Many evolutionary biologists discuss evolution entirely from a genetic perspective. After genes were established as the relevant heritable material the next step was to conceptualize the relevant heritable material in terms of molecular structure. In 1953 the structure of DNA was discovered and with this discovery came a mechanism for accounting for the duplication of heritable material and its transmission from one generation to the next. What the discovery of the structure of DNA also ushered in was a research focus for the developing field of molecular biology. An important part of this field is directed at uncovering aspects of organisms' development.

Developmental biology is the study of the development of organisms from fertilized egg to adulthood. Thought in developmental biology has often diverged from thought in evolutionary biology. Developmental biologists have periodically challenged views and approaches in evolutionary biology, including evolutionary biologists' focus on the gene. With the new techniques in molecular biology came the hope for a unified approach to evolution and development. On this approach, molecular evolutionary biology and molecular developmental biology would work consistently side by side. Evolution includes the passing on of genetic material form one generation to the next and development is the process moving from genetic material to somatic material. These processes can be understood from a molecular perspective if the component of heredity in evolution is understood to be the passing on of DNA from one generation to the next and development to be the production of proteins from DNA. In this picture, genes are discrete strands of DNA that are passed faithfully from one generation to the next and each gene is responsible for the production of a particular protein or polypeptide. This is a definition of the gene that reflects the story of the union of developmental and evolutionary biology via molecular biology.

Understanding more about the nature of DNA and RNA reveals a role that a concept of information can play in understanding heredity and development. The bases in DNA and RNA can be helpfully construed as letters in an alphabet and the relation between the triplets of letters in the RNA and the resulting polypeptide chain can be construed as a coding relation. So, the DNA contains the code for the polypeptide. Rather than causing the production of the relevant protein, the DNA sequence contains the code for it.

The definition of the gene introduced above can now be reconsidered. Rather than genes being discrete strands of DNA passed on from one generation to the next, genes can now be characterized as containing information that is passed on from one generation to the next and that information is the code for a particular protein or polypeptide. What is relevantly passed on from one generation to the next is the information in the DNA, encoded in the unique sequence of bases. Development can now be conceptualized as the faithful transmission of information from DNA to RNA, via the complementary base patterns, and then the passing on of that information into the linear structure of the protein, via the coding relation between triplets of base pairs and specific amino acids. Molecular biologists have introduced terminology that is consistent with this approach: the information in DNA is replicated in cell division, "transcribed" from DNA to RNA and "translated" from RNA to proteins.

Although the process of development includes every part of the life cycle of any particular organism leading to the whole collection of the organism's phenotypic traits, the discussion that follows focuses on the part of the developmental process operating within cells that starts with the separation of DNA strands and concludes with the production of a protein. In some discussions, genetic information is presented as containing instructions for the production for phenotypic traits such as eyes but these extensions of the concept present many additional problems to those reviewed below (Godfrey-Smith 2000).

2. The Pervasive Information Gene Concept: History and Current Practice.

In his provocative What Is Life, of 1944, the physicist Erwin Shrodinger said "these chromosomes … contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state" (Schrodinger 1944, 20). He went on to explain his terminology: "In calling the structure of the chromosome fibers a code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether the egg would develop, under suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a rhododendron, a beetle, a mouse or a woman" (Schrodinger 1944, 20-21). As Morange puts it, Shrodinger saw "genes merely as containers of information, as a code that determines the formation of the individual" (Morange 1998, 75). Shrodinger's proposals were made before the discovery of the structure of DNA. What is important that Shrodinger's words were read by many of those who were instrumental in the development of molecular biology.

As Sarkar (1996) points out, Watson and Crick are the first to use the term "information" in the context of discussions of the genetic code: "The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information" (Watson and Crick 1980/1953, 244). The French geneticists Jacob and Monod also played roles in sustaining Shrodinger's language of the code helping to reinforce the use of information language in molecular biology (Fox Keller 2000). By the early 1960's this terminology was established in the new field of molecular biology.

The information gene concept is pervasive in the work of theoretical evolutionary biologists. Perhaps the most influential formulation of the concept of heredity in terms of information was that of the evolutionary theorist George Williams. In his widely read <u>Adaptation and Natural Selection</u> [, 1966 #22] he says: "In evolutionary theory, a gene could be defined as any hereditary information for which there is a favorable or unfavorable selection bias equal to several or many times the rate of endogenous change" (Williams 1966, 25). And later: "A gene is not a DNA molecule; it is the transcribable information coded by the molecule" (Williams 1992, 11).

Defining the gene in terms of information or as a code is also the standard usage in biology text books. One of the standard molecular biology texts introduces the topic as follows: "Today the idea that DNA carries genetic information in its long chain of nucleotides is so fundamental to biological thought that it is sometimes difficult to realize the enormous intellectual gap it filled" (Alberts and al. 1994, 98). One of the definitions of the gene in the same text is: "a gene [is] any DNA sequence that is transcribed as a single unit and encodes one set of closely related polypeptide chains (protein isoforms)" (ibid., 457). Textbook introductions to molecular biology are filled with the terminology of transcription, translation and coding, terminology that implies the transfer of information.

It should now be clear that information terminology is pervasive in biology and also at least somewhat clear why this is the case. There were some historical reasons for adopting the terminology and there is some utility to the informational concepts. There are however some problems associated with construing genes informationally. Many of these problems have been introduced by philosophers of biology but there has also been much discussion of the information gene concept in biology. There have even been productive discussions between philosophers and biologists, a rare situation in philosophy of science but a highly desirable one.

3. Problems for the Information Gene Concept.

In several of his recent writings evolutionary biologist Maynard Smith has invited philosophers to join the discussion about the informational gene concept. For example, he says that "given the role that ideas drawn from a study of human communication have played, and continue to play, in biology it is strange that so little attention has been paid to them by philosophers of biology. I think that it is a topic that would reward serious study" (Maynard Smith 2000a, 192). While not addressing the concept of genetic information directly, philosophers of biology have been attending to these issues indirectly for some time in working on central problems in philosophy of biology. For example, the notion of genes as information has played an important role in discussions of reductionism, units of selection, the replicator/interactor distinction, gene/ environment interaction, nativism, and the developmental systems theorists' program. Recently, philosophers' focus has turned more explicitly to the informational gene concept. Several philosophers are now engaged in the project of developing a general notion of information that fits best with biologists' aims when they invoke genetic information.

The informational definition of the gene introduced above in Section 1. says that genes contain information that is passed on from one generation to the next and that information codes for a particular protein or polypeptide. As Sterelny and Griffiths put it: "The classical molecular gene concept is a stretch of DNA that codes for a single polypeptide chain" (1999, 132). Genes, on this view, contain information about the phenotype, the protein that is expressed. While most biologists believe that genes contain information about the relevant phenotype, none believe that the information in the genes is sufficient to produce the relevant phenotypes. Even those most routinely chastised for being genetic determinists understand that the information in the gene is only expressed with the aid of a whole host of cellular machinery. As a result the standard view is that genes contain the relevant or important information guiding the development of the organism. All other cellular machinery merely assists in the expression of the information. One way to put this idea is that genes introduce information to the developmental process while all other mechanisms make merely a causal contribution to development.

One move that philosophers (and some biologists) have made is to characterize the process of passing on the information in the gene by using terms from information theory. Information theory holds that "an event carries information about another event to the extent that it is causally related to it in a systematic fashion. Information is thus said to be conveyed over a "channel" connecting the "sender" [or "signal"] with the "receiver" when a change in the receiver is causally related to a change in the sender" (Gray 2001, 190). On this view information is reduced to causal covariance or systematic causal dependence. Philosophers of biology refer to this characterization of genetic information as the "causal" view. Sterelny and Griffiths (1999) illustrate how the causal information concept could work in the context of molecular biology: "The idea of information as systematic causal dependence can be used to explain how genes convey developmental information. The genome is the signal and the rest of the developmental matrix provides channel conditions under which the life cycle of the organism contains (receives) information about the genome" (Sterelny and Griffiths 1999, p.102).

Several have argued that the causal view suffers from serious problems. Sterelny and Griffiths (1999) point out that "it is a fundamental fact of information theory that the role of signal source and channel condition can be reversed" (p. 102) as the signal/channel distinction is simply a matter of causal covariance. Further, the signal/channel distinction is a function of observers' interests. For example, we could choose to hold the developmental history of an organism constant and from this perspective the organism's phenotype would carry information about their genotype. But if we choose to "hold all developmental factors other than (say) nutrient quantity constant, the amount of nutrition available to the organism will covary with, and hence also carry information about its phenotype" (p. 102). The causal information concept is lacking, because it cannot distinguish the genes as the singular bearers of important or relevant information. Rather, on this view, genes are just one source of information; aspects of the organism's environment and cellular material also contain information. This position is called the "parity thesis" (Griffiths and Gray 1994). The parity thesis exposes the need for another information concept that elevates genes alone to the status of information bearers.

Alternative concepts of information have been examined in attempts to respond to this situation; one is referred to variously as intentional, semantic or teleosemantic information (the term "teleosemantic" is used in what follows). This notion of information has been defended most forcefully recently by Maynard Smith but versions of it are defended by philosophers including Daniel Dennett (Dennett 1995) and Kim Sterelny (Sterelny, Smith et al. 1996; Sterelny 2000). The term "teleosemantics" is borrowed from "the philosophical program of reducing meaning to biological function (teleology) and then reducing biological function to" natural selection. Ruth Millikan and Karen Neander are proponents of this program (A good survey of relations between philosophy of mind and genetic information concepts is provided in Godfrey-Smith 1999). This view is articulated in the philosophy of mind as the thesis that a mental state token, such as a sentence, has the biological function of representing a particular state of the world and that function arose as result of selection.

Applying this view to the current problem results in the following: “a gene contains information about the developmental outcomes that it was selected to produce” (Sterelny and Griffiths 1999, 105). Maynard Smith puts the view as follows: "DNA contains information that has been programmed by natural selection" (Maynard Smith 2000a, 190). Here the information in the gene is analogous to a sentence in the head. The gene contains information not just as a result of relevantly causally co-varying with the phenotype, but as a result of having the function of producing the relevant phenotype. Defenders of this view, claim that this allows for the information to stay the same even if the channel conditions change; if the channel conditions change, the information in the gene has simply been misinterpreted. This concept could solve the problem of rendering the genes the sole information bearers, as “if other developmental causes do not contain [teleosemantic] information and genes do, then genes do indeed play a unique role in development” (Sterelny and Griffiths 1999, 104).

Although the teleosemantic view shows promise, the debate has not ended here. The teleosemantic view opens up a possibility: if a developmental cause, part of the cellular machinery for example, is found to be heritable and performs the function of producing a particular developmental outcome, then by definition, it also contains teleosemantic information. Many, including Sarkar (1996; 2000), Griffiths (1994), Gray (2001), Fox Keller (Fox Keller 2000), Sterelny (1996; 2000), have argued that indeed there are such mechanisms. These authors draw various conclusions from the demonstrated presence of mechanisms that are not genes, are heritable and perform the function of producing a specific developmental outcome. Developmental systems theorists such as Giffiths and Gray take these findings to show that teleosemantic information succumbs to the parity thesis also. They go on to argue that no concept of information will distinguish genes as a special contributor to development. Genes are just fellow travelers alongside cellular machinery and the environment in shaping developmental outcomes. Others such as Sarkar and Fox Keller are more cautious and hold out for a concept of information that can distinguish genes as a distinct kind of information bearer. On the other side, Maynard Smith and others have attempted to refine the notion of teleosemantic information to preserve a biological distinction that seems to be important: "The most fundamental distinction in biology is between nucleic acids, with their role as carriers of information, and proteins, which generate the phenotype" (Maynard Smith and Szathmary 1995, 61).

Three coherent options present themselves to answer the question “where is biological information found?”:

1. Information is present in DNA and other nucleotide sequences. Other cellular mechanisms contain no information.

2. Information is present in DNA, other nucleotide sequences and other cellular mechanisms, for example cytoplasmic or extra-cellular proteins; and in many other media, for example, the embryonic environment or components of an organism's wider environment.

3. DNA and other nucleotide sequences do not contain information, nor do any other cellular mechanisms.

These options can be read either ontologically or heuristically (An introduction to heuristics is provided by Richardson 1999). The ontological reading of 1. is that there is a certain kind of information that is only present in DNA and other nucleotide sequences. As a result any workable concept of information is constrained. The concept adopted cannot be consistent with information of the relevant sort existing in any other media that are causally responsible for an organism's development. The heuristic reading of 1. is that viewing information as present in DNA and other nucleotides is the most reliable guide to good answers in research in developmental molecular biology. The philosophical discussion presented above focuses on developing or challenging accounts of information that are consistent with an ontological reading of 1.. For example, Maynard Smith, and others such as Dennett, are defenders of a version of 1.. Many assume that 2. only makes sense ontologically if one adopts a causal information concept but some of the discussion already referred to indicates that other developmentally relevant media can be construed as containing teleosemantic information. Defenders of the developmental systems theory approach hold a version of 2. as does Sarkar (Sarkar 1996). To my knowledge only one philosopher, Ken Waters (Waters 2000), has provided a sustained defense of option 3.. Maynard Smith argues that to construe all processes of development in causal terms without recourse to the concept of genetic information is to relegate them to the hopelessly complex and implicitly to argue that no systematic explanations will be forthcoming (See e.g. Maynard Smith 1998, 5-6). Waters begs to differ, arguing that information talk in biology is misleading and can all be coherently be substituted for by causal talk. Waters also believes that it is most practicing biologists' intent to provide a causal account of development rather than one that invokes information.

This overview reveals that philosophers are actively cooperating with theoretical biologists to develop fruitful concepts of information that help us make sense of the information terminology widely used in biology. These discussions are still inconclusive and as a result this is a potentially productive area for future philosophical work.

References.

Alberts, B. and e. al. (1994). The Molecular Biology of the Cell. New York, Garland.

Dennett, D. C. (1995). Darwin's Dangerous Idea. New York, Simon and Schuster.

Fox Keller, E. (2000). Decoding the Genetic Program: Or, some circular logic in the logic of circularity. The Concept of the Gene in Development and Evolution. P. J. Beurton, R. Falk and H. Rheinberger. Cambridge, Cambridge University Press**:** 159-177.

Godfrey-Smith, P. (1999). Genes and Codes: Lessons from the philosophy of mind? Where Biology Meets Psychology: Philosophical Essays. V. G. Hardcastle. Cambridge, MIT Press**:** 305-332.

Godfrey-Smith, P. (2000). "On the Theoretical Role of "Genetic Coding"." Philosophy of Science **67**: 26-44.

Gray, R. D. (2001). Selfish Genes or Developmental Systems? Thinking About Evolution: Historical, Philosophical, and Political Perspectives. R. S. Singh, C. B. Krimbas, D. B. Paul and J. Beatty. Cambridge, Cambridge University Press**:** 184-207.

Griffiths, P. E. and R. D. Gray (1994). "Developmental Systems and Evolutionary Explanation." Journal of Philosophy **91**: 277-304.

Maynard Smith, J. (1998). Shaping Life. New Haven, Yale University Press.

Maynard Smith, J. (2000a). "The Concept of Information in Biology." Philosophy of Science **67**: 177-194.

Maynard Smith, J. and E. Szathmary (1995). The Major Transitions in Evolution. Oxford, Oxford University Press.

Morange, M. (1998). A History of Molecular Biology. Cambridge, Harvard University Press.

Richardson, R. (1999). Heuristics and Satisficing. A Companion to Cognitive Science. W. Bechtel and e. al. Oxford, Blackwell**:** 566-575.

Sarkar, S. (1996). Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology. The Philosophy and History of Molecular Biology: New Perspectives. S. Sarkar. Dordrecht, Kluwer**:** 187-231.

Sarkar, S. (2000). "Information in Genetics and Developmental Biology." Philosophy of Science **67**: 208-213.

Schrodinger, E. (1944). What is Life? The Physical Aspects of the Living Cell. Cambridge, Cambridge University Press.

Sterelny, K. (2000). "The "Genetic Program" Program: A Commentary on Maynard Smith on Information in Biology." Philosophy of Science **67**: 195-201.

Sterelny, K. and P. E. Griffiths (1999). Sex and Death. Chicago, University of Chicago Press.

Sterelny, K., K. Smith, et al. (1996). "The Extended Replicator." Biology and Philosophy **11**: 377-403.

Waters, K. (2000). "Molecules Made Biological." Revue Internationale de Philosophie: 539-564.

Williams, G. C. (1966). Adaptation and Natural Selection. Princeton, Princeton University Press.

Williams, G. C. (1992). Natural Selection: Domains, Levels and Challenges. New York, Oxford University Press.