# On the Theoretical Role of "Genetic Coding"

**Peter Godfrey-Smith**

Appears in *Philosophy of Science* 67 (2000): 26-44.

**Abstract**

The role played by the concept of genetic coding in biology is discussed. I argue that this concept makes a real contribution to solving a specific problem in cell biology. But attempts to make the idea of genetic coding do theoretical work elsewhere in biology, and in philosophy of biology, are probably mistaken. In particular, the concept of genetic coding should not be used (as it often is) to express a distinction between the traits of whole organisms that are coded for in the genes, and the traits that are not.

## 1. Introduction

The concept of genetic coding appears to be a central theoretical idea in contemporary biology, one of the keystones of our understanding of metabolism, development, inheritance, and evolution. A standard developmental biology textbook tells us that "the inherited information needed for development and metabolism is encoded in the DNA sequences of the chromosomes" (Gilbert 1997, 5). Current textbooks in cell biology and evolutionary biology make similar claims; our causal knowledge of biological processes is routinely presented as organized around the concept of genetic coding.[1]

The concept of genetic coding is also used to express a distinction between traits of organisms; some traits are coded for in the genes and others are not. A range of programs of empirical investigation are guided by the goal of categorizing various interesting traits, such as intelligence and sexual orientation, according to this distinction.

Frank Sulloway, for example, discussing work in evolutionary psychology, claims:

> [N]o one has identified any genes that code for altruistic behavior. Such genes are nevertheless believed to exist because certain aspects of personality that underlie cooperative behavior -- for example, empathy, sociability, and even altruism itself -- are moderately heritable. (Sulloway 1998, 34)

A trait is heritable, in a given population, if there is a certain statistical tendency for individuals with similar genotypes to be phenotypically similar with respect to that trait. Heritability is a subtle concept, associated with famous pitfalls and fallacies, but these complications do not matter here.[2] The important point is that Sulloway is saying that a statistical association between a psychological trait and genetic factors is evidence for the hypothesis that there are genes that <u>code</u> <u>for</u> that psychological trait. Does Sulloway mean, when he says genes "code for" a psychological trait, merely that there are genes that <u>cause</u> the trait? No, more than that must be meant (although the causal inferences raise their own problems). For suppose that research had found that cooperative tendencies have low heritability, and instead are strongly associated with certain environmental conditions. In the language Sulloway is speaking here, that would <u>not</u> suggest that there are environmental conditions that "code for" cooperative tendencies. According to the standard framework, both genes and environmental conditions <u>cause</u> traits, but only genes <u>code</u> <u>for</u> them.[3]

So whatever coding for traits is supposed to be in current thinking, it is seen as something that genes can do and environmental factors cannot do. This is one reason why the concept of <u>information</u>, often invoked in this context, cannot provide a complete solution to the coding problem. Information is sometimes a useful concept in biology, but on any standard concept of information, both genetic and non-genetic factors can carry information about traits.[4] The concept of genetic coding is apparently meant to pick out a <u>difference</u> between the causal paths leading from genes to traits, and the causal paths leading from environmental factors (or non-genetic factors more generally) and traits.

In understanding what this difference is supposed to be, we should not get sidetracked by crude genetic determinist views. The idea that genes code for traits need not be associated with the idea that genetic causation is inflexible and inevitable; that "genes are destiny." Talk of genetic coding does undoubtedly encourage such ideas, and that is one reason for us to scrutinize such talk carefully. But crude determinist views are not part of mainstream biological thinking. According to the standard picture, the "expression" of the genetic message is a contingent process, but one with key differences from other processes involved in development and metabolism.

Given the philosophical questions raised by all semantic and symbolic properties, given the methodological uncertainties illustrated by the Sulloway passage above, and given the lingering associations between genetic coding and genetic determinism, it is not surprising that some writers have objected to the whole concept of genetic coding. Sarkar (1996) opposes it because he thinks that as more details of molecular biological mechanisms are discovered, the conceptual framework associated with genetic coding becomes less and less appropriate, especially for complex multicellular organisms. Advocates of "developmental systems theory" are suspicious of genetic coding because they are suspicious of all "dichotomous" views of development, which sharply distinguish the causal roles of genetic and non-genetic factors (Oyama 1985, Griffiths and Gray 1994).[5] Philip Kitcher (forthcoming) has responded to some of this pressure by claiming that the concept of genetic coding, despite appearances, carries no explanatory weight; for Kitcher it is nothing more than a picturesque mode of talk. If Kitcher is right, we need not worry much about the language of coding in genetics, as it could be dropped without changing anything in biological theory itself.

At the other end of the spectrum, some philosophers hope to give a precise analysis of the semantic properties found in genes, and to use these properties in philosophical treatments of other problems. For example, Sterelny, Smith and Dickison (1996) appeal to the semantic properties of genes (and of certain other factors) to help resolve debates about the units of selection, via a new analysis of the concept of a "replicator."[6]

So the questions addressed here are as follows: Does the concept of genetic coding make a real contribution to our understanding of biological systems? And if so, what is the nature of this contribution?

I will argue for two set of claims. First, I claim that the concept of genetic coding does make a theoretical contribution to solving a specific, important problem about how cells work. So I disagree with those, like Kitcher, who see the concept of genetic coding as intended to bear no explanatory weight, and also those, like Sarkar, who think the concept is intended to carry explanatory weight but cannot. But the idea of genetic coding has diffused out from this original theoretical context, and has insinuated itself into many other descriptions of biological processes. The concept of genetic coding plays no single role in all the new contexts in which it is invoked; it is associated with a number of different ideas and is part of a diverse range of semantic concepts now routinely applied to genes. My second claim in this paper is that when the concept of genetic coding is found outside its original home, it does not make a contribution to solving any problems. More strongly, it is questionable whether the concept makes <u>any</u> positive contribution to our thinking about biological processes outside its original theoretical context. With caution, I take this conclusion to apply not only to coding but to other semantic concepts applied to genes. As a consequence, I am skeptical about projects of the type exemplified by Sterelny, Smith and Dickison 1996. These more general negative claims are hard to establish, however, and I do not claim the points I will make are decisive.

One way to approach questions about the semantic properties of genes is to ask whether genes meet the criteria laid down by philosophical semantic theories. The present paper does not take that approach. (It is discussed in a companion paper to this one.)[7] Instead, here I focus specifically on the contribution that the idea of genetic coding makes to the solution of problems in biology. I first go back and look at the original problem solved by the idea that DNA sequences comprise a genetic code. Then I try to argue that the theoretical role of genetic coding is restricted to this original context.

**2. Background to the Original Problem**

In this section I outline some historical background, intended to focus attention on a specific part of the problem of explaining gene action.

In early and mid twentieth century discussions of gene action, an important part of background knowledge was the fact that there are many complex molecules within living cells that are not supplied directly from outside. In particular, it was often discovered that each step of a cellular processes requires a specific enzyme to make it occur. All known enzymes were proteins, and it was thought that this might be true in general. Early in the century it had been discovered that proteins are long chains of individual amino acids, and it was hypothesized (correctly) that a given kind of protein molecule is made up of a specific, characteristic sequence of amino acids. Though they were believed to be simple chains at one level of description, proteins evidently formed more complex three-dimensional shapes as well.[8]

On the genetic side, by the end of the 1930s there had been several decades of work on specific mutations, their transmission across generations, and their effects on organisms, but not much progress on the molecular nature of genes. Genes were, however, confidently believed to be located (mostly or always) on chromosomes, in the cell nucleus. One hypothesis was that genes, on the chromosomes, affect what goes on in the cell by somehow controlling the synthesis of specific enzymes.[9]

So, if genes control the production of enzymes, enzymes are proteins, and proteins are linear chains of specific amino acids, then one important problem is understanding how a gene could determine the linear order of amino acids in a protein.

In the first half of the twentieth century it was not clear just <u>how</u> central this problem of ordering amino acids is to an understanding of gene action. In part, this is because it was not clear how amino acid sequence is related to other properties of protein molecules. Writers differed about both the role of amino acid sequence, and about which properties of proteins are under genetic influence. Only in retrospect can we say that the amino acid ordering problem turned out to be <u>the</u> pivotal problem, because although there is more to a protein than its amino acid sequence, solving the ordering problem was the key to a general understanding of how genes have their effects.[10]

### 3. The Problem and Solution

We are focusing on the problem: how can a gene control the exact sequence of a long chain of amino acids strung together by the cell? One strand of investigation from the 1930's through the 1950's speculated about various possible ordering mechanisms.

Given the general importance of enzymes, a natural thought was that proteins themselves, acting as enzymes, might put amino acids in the right order to make other proteins. However, what orders the amino acids in <u>those</u> enzymes?[11] We have an infinite regress, unless proteins are somehow linked in a complex causal web where specific proteins can join particular amino acids together, given the prior state of the amino acid chain, and every protein can be put together by some set of proteins acting as enzymes. This requires a network of multiple specificities and functions of proteins, which seemed unlikely.

A different approach is provided by the concept of a <u>template</u>. What is needed for each protein is a linear structure, of the right size, which can attract to it specific amino acids, in such a way that the amino acids line themselves up in the right order on this surface. Once placed in order, the amino acids can be bonded together. These templates might also be able to replicate, again by acting as templates. But which molecules have the right size and attraction properties to act as templates for proteins?

Here is an economical solution: an amino acid chain might serve as template in its <u>own</u> synthesis. The problem would remain of explaining how this system got started, but all the possible mechanisms have a problem of this kind; there is no vicious regress as in the first suggestion above.

Because of their chemical properties, there seemed no way for amino acid sequences to act as their own templates spontaneously, although some exotic physical mechanisms were entertained.[12] There might, however, be a set of two-sided "connector" molecules that mediate the interaction. Each connector molecule would attach to a specific amino acid on one side, and line up another amino acid of the same kind on the connector's other side.

This idea did not, of course, turn out to be right. For a variety of theoretical and empirical reasons, attention turned to the nucleic acids, and during the 1940's it was demonstrated that DNA is the molecule that somehow does the trick. The last stages of the race to solve the problem are well known (Olby 1994, Judson 1996). DNA was shown to be a double helix, and the solution to its structure by Watson and Crick showed immediately how DNA could be replicated, and led in time to an explanation of how DNA acts as a template in the production of proteins.

I assume that the reader is familiar with the main features of our current understanding of the solution to the problem – the distinction between transcription and translation, the roles of mRNA, tRNA and so on. The important features of this mechanism for our purposes are the role of template mechanisms in both transcription and translation, and the way that tRNA molecules act to associate RNA base triplets with specific amino acids during translation. The "genetic code" is, strictly speaking, the rule linking RNA base triplets with amino acids. This "interpretation" of the RNA determines the "interpretation" of the DNA from which the mRNA was derived.

### 4. Again: What was the Problem and How was it Solved?

The initial problem was the need to explain how cells put together complex protein molecules. Given what proteins are, a key part of this problem is explaining cells can string amino acids together in the right order. Cells do this with a pair of molecules, DNA and mRNA, which act as template surfaces. Some distinctive features of this solution motivate the contemporary concept of genetic coding.

First, the template used in making a protein is not the same protein itself, but another kind of molecule which is not a protein at all. These template molecules contain recurring, ordered elements (base triplets). As these elements are chemically different from the constituents of protein (amino acids) there must be a non-trivial rule of specificity linking the two.

Second, the specification of proteins by these templates is combinatorially structured. This structure exists at two levels. Most obviously, the elements of the templates which are specific for particular amino acids are base triplets. And in addition, in the context of translation a given triplet specifies the same amino acid regardless of its neighboring triplets; the interpretation of a long sequence of bases is a simple and fixed function of the interpretation of its component triplets.[13]

Third, the rule linking base triplets with amino acids is believed to be largely "arbitrary," although good deal of controversy surrounds this point. By "arbitrary," I mean that nothing about the chemistry of a particular amino acid is responsible for it corresponding to a particular base triplet. Contingent features of the tRNA molecules, and the enzymes which attach the amino acids to tRNAs, determine which triplets go with which amino acids.[14]

Thus the problem is solved. It is solved with the idea of a template as an ordering mechanism for proteins, and this solution features a relation between nucleic acid bases and proteins which can be described as a combinatorial, chemically arbitrary rule of causal specificity. These are the features of the relation between genes and proteins picked out by the idea of genetic coding.

These features were not what people had in mind when "coding" talk was first introduced into genetics, and not what everyone has had in mind since. According to Judson (1996), the first influential talk of "coding" in this context was by Schrödinger (1944/1992). Schrödinger said that when he talked of a "code-script" in the genes, what he meant was that a Laplacean "all-penetrating mind" could predict the organism from knowledge of the genes (p. 21). (As for the role of environment, Schrödinger gestured towards a requirement of "suitable conditions" for the development of the fertilized egg.) This official sense of "code-script" is a very weak one; later in his book Schrödinger made it clear that he was thinking about systems that have combinatorial features (p. 61). But the concept of genetic coding was not fixed by these initial discussions; thinking about what sorts of relationships are involved in biological coding evolved during the 1950's and 1960's alongside thinking about the code's actual structure. My argument in this paper does not hang on the issue of what people had in mind in the earliest discussions of coding. My argument concerns the role played by the idea of a genetic code in our current biological understanding.

In this discussion of the features of protein synthesis that motivate the idea of genetic coding I have not discussed the role of evolutionary history, or made use of strong concepts of biological function that depend on evolutionary history. In philosophy of mind some have argued that a concept of function based in evolutionary history is the key to a general explanation of semantic properties; perhaps such a concept of function is at least part of what is involved in genetic coding?[15] Against this suggestion it is important that the solution of the problem of protein synthesis made possible by the concept of genetic coding does not require or directly involve any hypotheses about evolutionary history. Genes would code in the sense relevant to solving the protein synthesis problem no matter what the evolutionary facts were.

## 5. The Restricted Theoretical Role of the Concept of Genetic Coding

The concept of genetic coding has its primary application in a solution to the problem posed by the cell's ability to order chains of amino acids. Given this theoretical role, what does the idea of genetic coding tell us more generally about how genetic and non-genetic causation are related? It tells us that genetic causation does have peculiarities, because at one specific place <u>within</u> the causal chains linking genes and traits, we have some unusual causal processes – the processes described above. Environmental conditions never have that particular kind of causal role, though they can have specific and distinctive roles. But these unusual features of genetic causation concern <u>how</u> genes manage to have some of their <u>immediate</u> effects. These features of genetic causation do not extend beyond the local process in which the protein in question is being made.

The symbolic perspective on genes has a role in solving a problem in cell biology. But once attributed coding properties, DNA seems to awake, to leap into new life and activity. The symbolic perspective on genes is so suggestive and so seemingly powerful that it affects how people talk about quite other aspects of genetics and gene action. The idea of genetic coding now affects how people think about the general distinction between genetic and non-genetic causal factors, and the relationships between genetic and non-genetic causal chains. Thus we encounter the view that among all the causal paths leading to the development of an adult organism, some of these causal paths are distinctive because they involve the expression of a genetically encoded message. That is, the concept of genetic coding is now used to describe and distinguish the entire causal paths in which genes are involved. This use of the concept of genetic coding has, I claim, no empirical basis and makes no contribution to our understanding.

Consequently, it is a mistake to use the idea of genetic coding to mark out a distinction between the traits of whole organisms that are coded for, and the traits that are not. There are good reasons for claiming that proteins are made by being coded for, and hence that a specific gene codes for a particular protein. But once we consider the complex traits of whole organisms, such as camouflage or cooperative behavioral tendencies, none of these traits are coded for by the genes.

More specifically, my view is that if the coding properties we should postulate here are those that have a positive theoretical role, then nothing can be coded for except the primary structure (the amino acid sequence) of a protein molecule. Not even the three-dimensional folded shape of a protein should be seen as coded for. There might be disagreement about the details of where coding properties can reach. But once we reach complex traits whose construction involves a mass of causal interactions, then the message is clear: such traits cannot be coded for.

To make this claim is not to deny that at least some causal relations are transitive, and so to deny that genes can causally affect complex traits of whole organisms. Genes can certainly have a causal role in the production of camouflage, or cooperation, via having a causal role in the synthesis of proteins. The long causal reach of genes is not at issue in this paper. What is at issue is the relation of "coding for...," and these semantic relationships have special properties. A case from everyday life illustrates the point. Suppose you know that if you order the extra-large pizza, that will have the consequence that the delivery arrives late. This fact does not imply that when you order the extra-large pizza you are also ordering them to make the delivery late. The likely or inevitable effects of a message are not all part of the content of the message. Similarly, genes can have a causal role which extends beyond the production of proteins, but proteins are all a gene can code for. So genetic coding is one part of a huge range of causal stories; coding is part of the processes underlying normal development of the human skeleton, and part of the story of how we come to be able to speak and sing. But the role for coding stops once each protein is produced.

Given all the problems arising from more extended and ambitious claims about genetic coding, it is reasonable to wonder why we should continue discussing genes in these terms, even within the specific context of understanding proteins synthesis. After all, what genes do is act as templates; the concept of coding does not add anything extra to our understanding here. When theorizing about the use of semantic concepts in describing complex, artificial computing devices, it is often said that our use of the language of rules and representations enables us to describe higher-level patterns in an otherwise unwieldy mass of physical processes. The same has been claimed about our habit of using semantic properties to discuss natural computing devices (brains) as well (Dennett 1987). No analogous argument can be made in the genetic case, as the use of semantic concepts to describe genes in the original theoretical context does not bring us any benefits of abstraction. Coding talk within discussions of how proteins are made is carried on at about the same level as ordinary description of the biochemical processes. Abstraction from that level is often desirable, but higher-level causal description is available for this task; the "coding" concept is not associated with any useful abstractness that is otherwise unavailable.

As a consequence, it might be thought advisable to drop the language of genetic coding altogether, as more trouble than it is worth. In another paper (forthcoming) I discuss various ways to link the semantic properties attributed to genes with philosophical theories of meaning and representation. If genes were to pass the tests imposed by the true semantic theory, then they would have coding properties whether this fact is useful or not. Sadly though, almost everything in philosophical semantics is controversial, and in addition, it turns out to be hard to handle genetic coding within many naturalistic semantic theories.

The coding properties of genes are probably, from the point of view of philosophical semantics, rather marginal and sui generis; similar in some ways to other representational properties but very different as well. So in strict and philosophical discussions, it is always possible to insist that genes do not code, represent, or instruct; the genetic code is no more than a rule of causal specificity, which need not be viewed as a rule of interpretation. However, the genetic code is a special kind of rule of causal specificity, and in this discussion I will continue to accept that the peculiar features of the templating properties of genes (discussed in section 4) do motivate description of genes as coding for proteins. Genetic coding for proteins is, however sui generis from a philosophical standpoint, nonetheless a well-grounded concept that solves a hard biological problem.

So there is one theoretical context in which the attribution of semantic properties to genes does play a positive role. I am skeptical about whether there are others, and I suggest that the attribution of semantic properties to genes should be restricted to cases in which there is such a positive role. Of course, there is a great range of possible descriptions of genes in semantic or quasi-semantic terms; there are lots of ways to say "by 'genetic message' I only mean to say X," where X is something unobjectionable that exhibits at least a faint analogy with paradigmatic cases of messages and symbols. The more difficult task is showing, as can be shown in the case of explaining protein synthesis, that something is gained by such a description.

**6. A Thought-Experiment**

In the previous section I outlined a restricted view of the theoretical role of genetic coding. In this section I will use a thought-experiment to support these claims.

To ask how essential the idea of genetic coding is to our understanding, one approach we can take is to ask what the world would have to be like if there was no genetic coding. If we suppose there was no genetic coding, does that require us to imagine a world in which biological development and evolution are very different from how they actually are? When one tries to imagine away genetic coding without massively changing basic biological facts, does one inevitably imagine a situation in which there is just a different kind of genetic coding? Or are there possible worlds with no genetic coding in which not much else is different?

I claim there is a possible scenario with no genetic coding, in which development and evolution are much the same as they are in actuality, and where the suppositions needed to entertain this scenario are reasonable ones. Removing genetic coding from the world need not change much else, and this gives support to my claim that we should only think of coding as part of an explanation of how cells achieve the specific task of putting amino acids in the right order.

My thought experiment will be motivated by using again some of the historical story sketched earlier. As outlined in section 3, there was a period before DNA had been shown to be the genetic material, when there was speculation about various possible ordering mechanisms for amino acids.

One suggestion was that proteins could act as templates for themselves, probably with the aid of a set of 20 two-sided "connector" molecules. The connectors might each attach to a specific amino acid in a chain, and then bind on their other side to a single molecule of the same amino acid, floating around the cell. Such speculation vanished when DNA became the focus of research, but suppose this alternative theory had turned out to be true. Suppose it had turned out that genes are made of proteins which act as their own templates.[16] These protein molecules would have to be strung out linearly, losing their folded shape, in order to act as templates – the chromosome somehow holds the protein open in this way, giving the connector molecules access to each amino acid in the chain. I don't know if an arrangement like this is really chemically possible, but I stress again that the idea was seriously entertained.

Suppose genes were protein molecules of this kind. In such a situation, DNA sequences certainly would not contain a genetic code. But further, in such a situation there would be no such thing as genetic coding. A gene in that case would not contain any kind of representation of a protein. Instead, a gene would be a <u>sample</u> of the protein. The nucleus of a cell would contain a sample of every protein it needs. To make its enzymes, the cell generates copies of each protein from its store of samples. To make copies of these protein genes during cell division, the same sort of mechanism could be used.

Though we would have no genetic coding in this situation, little else need be different. In the causal stories which could be told about the development of individual organisms, <u>one</u> part of the stories would always differ from actuality; the way in which genes order amino acids to make proteins would be different. But the causal stories <u>from</u> the point where each protein molecule has been produced would be exactly or nearly the same. Once a protein is produced, it makes no difference whether it was coded for or run off from a sample. Proteins would still catalyze the cell's key reactions, and proteins could still affect which genes are active in which cells of the organism. Roughly the same cascades of developmental processes could occur. We could still have some traits that recur in a given species with great regularity in every generation, and other traits which differ dramatically across slightly different environments. We could still have the same mixture of biological characteristics that resist interference, and characteristics that permit easy modification. These patterns and processes in development would be just as explicable without genetic coding as they are with genetic coding. Removing genetic coding from the world does not entail a different set of general relationships between genetic and environmental causation.

Beside development and metabolism, another context where genetic coding is often seen as important is the explanation of inheritance. Genes, it is often said, carry information across generations, and do so via their coding properties.

When describing this role for genetic coding, it is common to stress the narrowness of the "bottleneck" bridging the generations. Views of inheritance which overstress or overstate the narrowness of this bottleneck have recently been criticized on both empirical and conceptual grounds (Jablonka and Lamb 1995, Oyama 1985, Gray 1992), but those controversies will not matter to my point here. For consider the bottleneck as narrowly as you like, and then consider again the thought-experiment. We imagine protein genes which act as templates in their own synthesis. They are replicated in cell division, and passed via sex cells into the next generation. In this situation the inheritance of complex traits would not be explained in terms of the passing of a coded message across the bottleneck. Instead, what passes though the bottleneck is a sample of the proteins used in the next generation, and the beginnings of the dynamic processes that utilize these proteins. But other than this detail, the processes of development and inheritance could be basically the same as they are in the actual world. And in the scenario with protein genes, there is less temptation to see these genes as reaching out and determining the entire adult structure. Successful passage through the bottleneck does not endow these genes with special powers over their eventual causal products. But if protein genes would have no such special role, then neither do our actual genes, made of DNA. The difference between the situation in the thought experiment and the situation in actuality is merely a difference in the local chemical detail.

In this section, I have taken it to be clear that genes comprising samples of the protein molecules used in the cell do not code for anything. However, it is possible for someone to insist that in this scenario "each protein codes for itself."[17] My reply is that someone could indeed insist on this – could insist that the trivial relation taking each protein molecule to itself is still a coding relation. As I said earlier, there are lots of ways to stipulate special senses of "coding" or "representation." My point is that there would be no biological reason to posit a coding relation in this case.

Similarly, I suggested that the thought experiment describes a situation in which the temptations toward genetic determinism are reduced. But the scenario is one where it is still possible to overstate the causal importance of genes, and endow protein genes with much of the "mystique" enjoyed by DNA genes.[18] I suggest, however, that the over-use of semantic concepts in genetic discussion is a real factor in generating simplistic views about genetic causation, even though mainstream views in biology are not causally simplistic. The thought experiment is useful in this context as it shows that genes without their usual semantic properties have not gained or lost any causal powers.

Arguments made with the aid of thought experiments are always controversial. But no one thinks that all thought experiments are illegitimate (lest even Maxwell's demon be demonized). The argument of this paper, I claim, uses a legitimate thought experiment, and on this point I stress again the basis for the thought experiment in professional speculation about genes in the years before Watson and Crick.

**7. What About Gene Regulation?**

In presenting the thought-experiment above I made simplifications. Most importantly, I omitted the complex processes of gene regulation. But gene regulation is another aspect, apparently, of the problem of genetic coding.

To deal with the issues raised for my arguments by gene regulation, we should distinguish two different, striking properties that living things have. The first is the capacity to make complex molecules from simpler raw materials. The second is the fact that organisms are able to follow reliable, orderly paths in developmental processes.

The first property motivates the idea that the cell contains a <u>representation</u> of the structure of complex proteins. The second property generates the idea that the cell contains something like a <u>program</u> to direct it along its course. These two properties are often discussed together, but they pose distinct problems. It is a mistake to think of the idea of genetic coding and the idea of a genetic program as a package deal.

This separation can be illustrated by means of the thought experiment. Could the protein genes introduced in the previous section have the same role in regulatory processes as actual DNA genes? The answer is that the protein genes could have a similar <u>kind</u> of role. In particular, the phenomena which lead to the positing of a "program" could exist even with genes that do not code for proteins.

The thought experiment becomes more difficult here, as there are so many mechanisms of gene regulation, some of which depend on details of the DNA/RNA chemistry which is being imagined away. But it is standard to hold that the single most important site of gene regulation is the initiation of transcription. Transcription is initiated when an RNA polymerase binds to the DNA and begins formation of what will become an mRNA molecule. In complex multicellular organisms, the initiation of transcription is controlled largely by (i) protein transcription factors which bind to the DNA, and (ii) folding and looping of the DNA molecule itself, in conjunction with histone proteins, forming structures which seem to prevent DNA sequences being exposed to RNA polymerase. These mechanisms are themselves controlled by a web of causal processes.

Although there would have to be many changes to the details, if genes were made of protein and contained samples rather than coded messages, regulatory processes roughly like these could still go on. There could be transcription factors and folding of chromosomes. The sequences of events through which genes become active and inactive could be as complex as they are in the actual world; the paths taken by cell differentiation and development could be as reliable as they actually are, and so on. There could, in the situation described in the thought experiment, be just the same reasons to see the cell as akin to a "programmed device," running through an orderly sequence of events in accordance with evolutionary design. The fact that actual DNA genes contain a code does not make much difference to the issue of whether genes, or anything else in the cell, should be described as a containing a program.

So the property of coding for proteins, and the property of containing a developmental program, are distinct properties. In this paper I am only concerned with coding. The processes of gene regulation do not cause trouble for my use of the thought experiment to justify claims about genetic coding.

Having made this separation of the issues, I leave the problems raised by the concept of a genetic program for another day.

**8. What about Evolution?**

Genetic coding also has no great importance for our understanding of evolution. Evolutionary biology does not provide a context in which the more extended and ambitious attributions of genetic coding properties have explanatory value.

Genetic coding is sometimes taken to be essential to the approximate truth of various versions of Weismannism, the conceptual framework which distinguishes the "germ line" which passes through generations intact, and the "soma" which lives, dies, develops, and interacts with the environment. Genetic coding was embedded by Crick and others within a view which asserts a one-way "flow of information" from DNA to RNA and then to proteins -- the "Central Dogma" of molecular biology. As it is sometimes put, the Central Dogma expresses molecular Weismannism.[19] So it might appear that the idea of genetic coding has become embedded in our theoretical understanding of evolution.

In reply, I point one more time to the thought experiment. With genes made of protein, we could have the approximate truth of various Weismannist doctrines. We could still have a germ line, in both the original sense of a sequestered lineage of cells (in some organisms) for production of gametes, and in the molecular sense of a lineage of molecules that persists in the cell nucleus through developmental changes. A world with non-coding genes need not be much different with respect to the core features of evolution. For inheritance to be possible with the patterns necessary for Darwinian evolution, there does have to be a kind of compression of biological structure, bridging the generations. But this compression need not be via a message or description of the biological structure. It could be a compression in the form of samples of proteins.

When one imagines genes being made of protein, the genetic material is still shut inside the nucleus, but the same kind of molecule is also outside, interacting with the environment in familiar ways. Consequently, protein genes are much less likely to appear like little diamonds (which are forever), or like pure, immaterial information, removed from the ordinary causal give-and-take. Changing the scenario from protein genes back to DNA genes does not turn genes into little diamonds either. All the switch back to DNA does is change the genes' chemical properties, and how they act as templates in protein synthesis. Switching in and out of the thought experiment is, I suggest, a good exercise for keeping the causal properties of genes in perspective.

**9. Conclusion**

One way to ask about the theoretical contribution of the idea of genetic coding is to ask what things would have to be like if there was no genetic code. This might initially seem difficult, as coding talk has insinuated itself into so many areas in biology. Genetic coding looks so central that one might expect that imagining it away would be like imagining away gravity, or natural selection. But it is easy to imagine a situation in which there is no genetic coding, and this is illustrated by consideration of some ideas pursued in the years before Watson and Crick. Once inspected in this light, it is apparent that coding is part of the solution to one specific problem, the problem of ordering amino acids in proteins synthesis. To say this is not to deny that genetic coding is important – the problem and solution are of enormous scientific importance. But genetic coding, properly understood, is part of the local molecular detail. Once the amino acids are placed in order, the role for genetic coding is over. After that, it is up to the proteins to do whatever they can do.

So I suggest that the idea of genetic coding makes no essential contribution to the larger theoretical projects with which it is often associated, the projects of achieving a more general understanding of development and evolution. In particular, the concept of genetic coding should not be used to express a distinction between traits of whole organisms that are coded for in the genes (and hence are evolutionarily relevant) and traits that are not.

To reduce the theoretical role of genetic coding is not to solve or dissolve many of the problems which people currently use the idea of genetic coding to address, however. The problems remain, and should be addressed with different tools.

Of particular importance is the problem of understanding the relation between "genetic" or "innate" traits, on the one hand, and "environmental," "acquired" or "learned" traits on the other. Genetic coding is not the right concept for understanding this distinction, but that does not show that the distinction being gestured towards is misguided or incoherent. It <u>might</u> be misguided, but it might not be. In this paper I have left open the question of whether the "genetic program" is a useful concept; some will rely on that framework to address the issue. I suggest, however, that these problems are better addressed just using causal concepts. Indeed, it is possible that the original "genetic/environmental" distinction, even when construed as a continuum, should be replaced with a number of other distinctions, using the concept of causation along with some more technical concepts such as canalization, homeostasis, and the norm of reaction.[20]

Even within a more complex framework for addressing these questions, it might turn out to be useful to retain a vaguer, catch-all distinction between the "genetic" and the "environmental." That may be, but when the theoretical chips are down, any such distinction should not be understood in terms of genetic coding.

*　*　*　*　*

---

**Footnotes**

**Acknowledgment:** A version of this paper was presented at the 1998 Australasian Association of Philosophy meetings in Sydney. I am grateful for those present for helpful criticisms. Thanks to Richard Francis, David Hull, Philip Kitcher, John Maynard Smith, Susan Oyama, Kim Sterelny, Patrick Suppes, Johan van Bentham, and many students in my philosophy of biology seminar at Stanford, for discussions and correspondence. Two anonymous referees also made valuable comments.

[1] For cell biology, see Alberts et al. 1994, 102: "Organisms differ because their respective DNA molecules carry different nucleotide sequences and therefore different biological messages." For evolutionary biology see Futuyma 1986, 43: "The genotype is a blueprint for the organism, the set of instructions for development received from its parents." In these quotes (and elsewhere) we find a number of different semantic, intentional, or symbolic concepts used to describe DNA. Some of these are discussed in the present paper, but others are discussed in Godfrey-Smith forthcoming.

[2] Lewontin 1974 is a classic discussion, and see also Block 1995. For example, heritability is taken to require that the correlation found between genes and phenotype not be due to a covariance between genetic and environmental causes. But distinguishing the effects of gene/environment covariance from "ordinary" gene action is difficult both conceptually and practically.

[3] Might he mean that genes cause the trait and have the <u>function</u> of causing that trait? This option is discussed in Godfrey-Smith forthcoming and Kaplan and Pigliucci forthcoming, and will be briefly discussed in section 4.

[4] On this point, see Griffiths and Gray 1994 and Godfrey-Smith forthcoming.

[5] See also Moss 1992 and Francis forthcoming for other criticisms of the application of semantic concepts to genes.

[6] For more standard analyses of replicators, see Dawkins 1982 and Hull 1980.

[7] This direct approach, discussed in Godfrey-Smith forthcoming, seems to me likely to remain inconclusive.

[8] This folding of the amino acid chain is in general spontaneous, and the resulting shape is thought to be primarily a consequence of the amino acid sequence itself. But there are exceptions, and a good deal of continuing controversy about how important other factors are to protein folding. See, for example, Frydman and Hartl 1996 on the role of "chaperone" molecules. In the period under discussion there was a great deal of uncertainty about the relation between amino acid sequence and three dimensional structure. See Crick 1958 and Olby 1994, especially pp. 434-35.

[9] This idea was made fully explicit by Beadle and Tatum in the 1940's, but had been in the air since Garrod's work soon after the turn of the century. See Olby 1994, Chapters 8 and 9.

[10] Sources differ about how much consensus there was about the centrality of this "ordering problem" in the decades before Watson and Crick's model of DNA and the work on gene action which followed it. A textbook presentation in Watson et al. 1987 gives the impression of great consensus: "From the very start of serious speculation, the simplest hypothesis [about gene action] was that genetic information within genes determines the order of the 20 amino acids within the polypeptide chains of proteins" (p. 65). But Olby finds only a few clear statements along these lines in the 1930's, and stresses that "it was not an obvious step to make the connexion between the linear sequence of the genes and that of the amino acids in a polypeptide chain" (1994, 116, emphasis in original). Maybe Watson et al. are right that a stress on ordering mechanisms was always the "simplest" idea, but that did not make it seem the most promising idea to everyone at the time. Crick by 1958 certainly saw the problem in these terms: "It is... the problem of 'sequentialization' [of amino acids] which is the crux of the matter" (Crick 1958, 144). See also Muller 1947 for an earlier survey of ideas about both how amino acid order might be determined, and whether understanding the ordering of amino acids is all there is to understanding gene action.

[11] See Crick 1958 and Judson 1996, 253.

[12] See, for example, Pauling and Delbrück 1940, and Judson 1996, 117-18.

[13] Here I assume that the reading frame is fixed. I also make no mention of some exceptions to these principles, some of which are discussed in Sarkar 1996, Moss 1992 and Godfrey-Smith forthcoming.

[14] Note, for example, that the enzymes attaching amino acids to tRNAs sometimes do not recognize the tRNA by its anticodon but by other parts of the tRNA. This <u>chemical</u> arbitrariness of the amino acid/codon assignments is compatible with a denial of what we can call <u>functional</u> arbitrariness; there are some systematic features <u>within</u> the structure of the code that might be products of natural selection (for example, to reduce the harmful effect of mutations). But those systematic features are compatible with many different sets of assignments of codons to amino acids.

The suggestion that the genetic code originates as a chemically arbitrary "frozen accident" is due to Crick (1968). The idea has been challenged; see Maynard Smith and Szathmáry 1995, Chapter 6, for an interesting discussion.

[15] For an influential application of the concept of function to problems in the philosophy of mind see Millikan 1984. Kaplan and Pigliucci forthcoming develop a view of the "gene for..." concept based on an evolutionary concept of function, but unlike Sterelny, Smith and Dickison (1996), Kaplan and Pigliucci do not present their "gene for..." concept as a <u>semantic</u> concept. Consequently, a view like theirs is compatible (as they note) with the treatment of the semantic issues advocated in the present paper.

[16] For discussion of models of this kind by Astbury, Delbrück, Muller and others, in the context both of gene action and gene duplication, see Muller 1947 and Olby 1994 Chapters 5 and 7.

[17] Olby quotes an early discussion of genes, by the biochemist Albrecht Kossel in 1911, in which genes made of protein are said to be able to "contain, to a certain extent, a complete description of the species and even of the individual" (1994, 77). So I cannot claim that no scientist in the field would ever use semantic terminology about genes made of protein. But I do claim that this would be, from the standpoint of our present view, a very unusual way of using such terminology.

[18] A referee stressed this possibility to me.

[19] For a relevant discussion of Weismannism, both original and molecular, and its relation to evolutionary theory, see Griesemer forthcoming.

[20] See Wimsatt's work (eg., 1986) for subtle discussions of some of these issues, especially questions about the developmental interlocking of different traits. See also Cowie 1999 for a different line of argument against some traditional distinctions between the "innate" and the "learned."

**References**


Alberts, Bruce, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson (1994), Molecular Biology of the Cell, 3rd edition. New York: Garland.

Block, Ned. (1995), "How Heritability Misleads About Race", Cognition 56: 99-128.

Cowie, Fiona. (1999), What's Within? Oxford: Oxford University Press.

Crick, Francis. (1958), "On Protein Synthesis", Symposia of the Society for Experimental Biology 12: 138-163.

Crick, Francis. (1968), "The Origin of the Genetic Code", Journal of Molecular Biology 38: 367-79.

Dawkins, Richard. (1982), The Extended Phenotype. Oxford: Freeman.

Dennett, Daniel C. (1987), The Intentional Stance. Cambridge MA: MIT Press.

Francis, Richard. (forthcoming), Genes, Brains, and Sex in the Information Age. Princeton University Press.

Futuyma, Douglas. (1986), Evolutionary Biology. Sunderland MA: Sinauer.

Frydman, Judith and F. U. Hartl (1996). "Principles of Chaperone-Assisted Protein Folding: Differences Between in Vitro and in Vivo Mechanisms", Science 272: 1497-1502.

Gilbert, Scott F. (1997), Developmental Biology. Fifth edition. Sunderland MA: Sinauer.

Godfrey-Smith, Peter. (forthcoming), "Genes and Codes: Lessons from the Philosophy of Mind?", to appear in V. Hardcastle (ed.), Biology Meets Psychology: Constraints, Conjectures, Connections. Cambridge MA: MIT Press.

Gray, Russell. (1992), "Death of the Gene: Developmental Systems Strike Back", in Paul Griffiths (ed.) Trees of Life: Essays in Philosophy of Biology. Dordrecht: Kluwer.

Griesemer, James. (forthcoming), "The Informational Gene and the Substantial Body: On the Generalization of Evolutionary Theory by Abstraction", in

Nancy Cartwright and Martin Jones (eds.), Varieties of Idealization. Amsterdam: Rodopi.

Griffiths, Paul and Russell Gray. (1994), "Developmental Systems and Evolutionary Explanation", Journal of Philosophy 91: 277-304.

Hull, David L. (1980), "Individuality and Selection", Annual Review of Ecology and Systematics 11: 311-32.

Jablonka, Eva. and Marion Lamb. (1995), Epigenetic Inheritance and Evolution. Oxford: Oxford University Press.

Judson, Horace. F. (1996), The Eighth Day of Creation (expanded edition). Plainview: Cold Spring Harbor Laboratory Press.

Kaplan, Jon. and Massimo Pigliucci. (forthcoming), "Genes 'for' Phenotypes: A Modern History View."

Kitcher, Philip. S. (forthcoming), "Battling the Undead: How (and How Not) to Resist Genetic Determinism", to appear in Rama Singh, Costas Krimbas, Diane Paul and John Beatty (eds.), Thinking About Evolution: Historical, Philosophical and Political Perspectives. Cambridge: Cambridge University Press.

Lewontin, Richard C. (1974), "The Analysis of Variance and the Analysis of Cause", reprinted in Richard Levins and Richard C. Lewontin The Dialectical Biologist. Cambridge MA: Harvard University Press, 1985.

Lodish, Harvey., David Baltimore, Arnold Berk, S. Lawrence Zipursky, Paul Matsudaira and James Darnell (1995), Molecular Cell Biology, 3rd edition. New York: Freeman.

Maynard Smith, John. and Eörs Szathmáry (1995). The Major Transitions in Evolution. Oxford: Oxford University Press.

Millikan, Ruth G. (1984), Language, Thought and Other Biological Categories. Cambridge MA: MIT Press.

Moss, Lenny. (1992), "A Kernel of Truth? On the Reality of the Genetic Program", PSA 1992, Volume 1: 335-348.

Muller, Hermann. J. (1947), "Genes", Proceedings of the Royal Society of London, B 134: 1-37.

Olby, Robert. (1994), The Path to the Double Helix (Enlarged edition). New York: Dover.

Oyama, Susan. (1985), The Ontogeny of Information. Cambridge: Cambridge University Press.

Pauling, Linus. and Max Delbrück (1940), "The Nature of the Intermolecular Forces Operative in Biological Systems", Science 92: 77-79.

Sarkar, Sahotra. (1996), "Decoding "Coding" -- Information and DNA", BioScience 46: 857-864.

Schrödinger, Erwin. (1944/1992), What is Life? Cambridge: Cambridge University Press.

Sterelny, Kim, Kelly Smith and Michael Dickison (1996), "The Extended Replicator", Biology and Philosophy 11: 377-403.

Sulloway, Frank. (1998), "Darwinian Virtues", New York Review of Books 45 (6): 34-40. [April 9, 1998]

Watson, James, Nancy Hopkins, Jeffrey W. Roberts, Joan A. Steitz and Alan Weiner (1987), The Molecular Biology of the Gene, 4th edition. Menlo Park: Benjamin/Cummins.

Wimsatt, William C. (1986), "Developmental Constraints, Generative Entrenchment, and the Innate-Acquired Distinction", in P. William Bechtel. (ed.), Integrating Scientific Disciplines. Dordrecht: Martinus-Nijhoff. pp. 185-208.