Information in Biology

Peter Godfrey-Smith

Harvard University

1. Introduction

The concept of information has acquired a strikingly prominent role in contemporary biology. This trend is especially marked within genetics, but it has also become important in other areas, such as evolutionary theory and developmental biology, particularly where these fields border on genetics. The most distinctive biological role for informational concepts, and the one that has generated the most discussion, is in the description of the relations between genes and the various structures and processes that genes play a role in causing. For many biologists, the causal role of genes should be understood in terms of their *carrying information* about their various products. That information might require the cooperation of various environmental factors before it can be "expressed," but the same can be said of other kinds of message.

An initial response might be to think that this mode of description is entirely anchored in a set of well-established facts about the role of DNA and RNA within protein synthesis, summarized in the familiar chart representing the "genetic code," mapping DNA base triplets to amino acids. However, informational enthusiasm in biology pre-dates even a rudimentary understanding of these mechanisms (Schrodinger 1944). And more importantly, current applications of informational concepts extend far beyond anything that can receive an obvious justification in terms of the familiar facts about the specification of protein molecules by DNA. This includes:

(i) The description of whole-organism phenotypic traits (including complex behavioral traits) as specified or coded for by information contained in the genes,

(ii) The treatment of many causal processes within cells, and perhaps of the whole-organism developmental sequence, in terms of the execution of a *program* stored in the genes,

(iii) The idea that genes themselves, for the purpose of evolutionary theorizing, should be seen as, in some sense, "made" of information. From this point of view, information becomes a fundamental ingredient in the biological world.

There is no consensus about the proper form and status of these kinds of description, and the result has been the development of a foundational discussion within both biology and the philosophy of biology. Some have hailed the employment of informational concepts here as a crucial advance (Williams 1992). Others have seen almost every biological application of informational concepts as a serious error, one that distorts our understanding and contributes to lingering genetic determinism (Francis 2003). Most of the possible options between these extreme views have also been defended. These include various arguments that *some*, though not all, of the popular uses of informational concepts in biology are legitimate (Godfrey-Smith 2000, Griffiths 2001). They also include arguments that even the more tendentious uses of these concepts are legitimate so long as the concepts are applied consistently (Sterelny, Smith, and Dickison 1996, Jablonka 2002). Other philosophers and biologists regard the whole matter as a storm in a teacup; they do not think that the development of an informational language for describing genes makes much of a difference to anything, as it is obviously a loose metaphorical usage that carries no real theoretical weight (Kitcher 2001).

The philosophical discussion has developed for two reasons. One is the general philosophical interest in abstract conceptual problems in particular areas of science – an interest in debates that seem resistant to empirical adjudication, but do not seem merely terminological. So some philosophical interest here is akin to more familiar philosophical attention to such biological concepts as fitness, species, and natural selection. But the concept of information is not merely an ordinary theoretical concept within a particular part of science. It is also part of a family of concepts that has been the focus of intense

study in several parts of philosophy, stretching back for centuries. "Information" itself does not have a long history in philosophy, but it is closely related to concepts that do, such as the concept of meaning, which is central to philosophy of language and much philosophy of mind. So philosophers are familiar with the kinds of puzzles that are generated by this family of "semantic" concepts. It is not that philosophers have developed a consensus theory that can be applied, in an off-the-shelf way, to new cases. But philosophers are intimately acquainted with many of the puzzles, twists and turns, red herrings, and trade-offs that arise in this area. So as information (and related concepts) have become more prominent in biology, some philosophers have thought that this is an area where they are qualified to help in the development of useful and coherent biological concepts.

This paper has two main sections. The next section gives an outline of some of the arguments and options developed to date. The third section then develops some more novel ideas, which are presented in a cautious and exploratory way.

Before moving to the survey section, there are two other preliminary points to make. First, the topic of this paper is not the role of the concepts of information and representation in the parts of biology where they are most *obviously* relevant; the paper is not concerned with neuroscience, perception, language-processing, and so on. The topic of this paper is the role of information (and its relatives) in parts of biology where its role is less obvious, such as the description of genes, hormones, and (to some extent) signaling systems at the cellular level. Secondly, in the early part of this discussion I will not put much emphasis on some of the finer distinctions between the concepts of information, representation, meaning, coding, and so on. As the paper goes on, distinctions between concepts within this family will become more important, but some of the subtle distinctions will be backgrounded initially.

2. Outline of the debate

One common way to start organizing the problem is to make a distinction between two senses of "information," or two kinds of application of informational concepts. One of these is a weak or minimal sense, and the other is stronger and more controversial. In the

weaker sense, informational connections between events or variables involve no more than ordinary correlations (or perhaps correlations that are "non-accidental" in some physical sense involving causation or natural laws). This sense of information is associated with Claude Shannon (1948), who showed how the concept of information could be used to *quantify* facts about contingency and correlation in a useful way, initially for communication technology. For Shannon, anything is a *source* of information if it has a number of alternative states that might be realized on a particular occasion. And any other variable *carries information* about the source if its state is correlated with that of the source. This is a matter of degree; a signal carries more information about a source if its state is a better predictor of the source, less information if it is a worse predictor.

This way of thinking about contingency and correlation has turned out to be useful in many areas outside of the original technological applications that Shannon had in mind, and genetics is one example. There are interesting questions that can be asked about this sense of information (Dretske 1981), but the initially important point is that when a biologist introduces information in this sense to a description of gene action or other processes, she is not introducing some new and special *kind* of relation or property. She is just adopting a particular quantitative framework for describing ordinary correlations or causal connections.

Consequently, philosophical discussions have sometimes set the issue up by saying that there is one kind of "information" appealed to in biology, Shannon's kind, that is unproblematic and does not require much philosophical attention. The term "causal" information is sometimes used to refer to this kind, though this term is not ideal. Whatever it is called, this kind of information exists whenever there is ordinary contingency and correlation. So we can say that genes contain information about the proteins they make, and also that genes contain information about the whole-organism phenotype. But when we say that, we are saying no more than what we are saying when we say that there is an informational connection between smoke and fire, or between tree rings and a tree's age. The more contentious question then becomes whether or not biology needs *another*, richer concept of information as well as Shannon's concept. Information in this richer sense is sometimes called "semantic" or "intentional" information.

What is the difference between them, and why might we think that biology needs to employ a richer concept? There is a range of differences between the two. First and most importantly, informational connections in the Shannon sense connect environmental conditions with biological traits in the same way that they connect genes and those traits. With respect to Shannon information, there is what Griffiths and Gray call a "parity" between the roles of environmental and genetic causes (Griffiths and Gray 1994, Griffiths 2001). In addition, information in the Shannon sense "flows" in both directions, as it involves no more than learning about the state of one variable by attending to another. So we can read off something about the phenotype from the state of the genes, but we can also learn something about the genes by attending to the phenotype.

Some talk about information in biology is consistent with these features of Shannon information, but some is not. It is usually thought that at least some applications of informational language to genes is supposed to ascribe to genes a special kind of causal property that is *not* ascribed to environmental conditions, even when they are causally important, and that is also uni-directional.

In addition, a message that carries "semantic information," it is often thought, has the capacity to mis-represent, as well as accurately represent, what it is about. There is a capacity for error. Shannon information does not have that feature; we cannot say that some variable carried false information about another, if we are using the original Shannon sense of the term. But biologists do apparently want to use language of that kind when talking about genes. Genes carry a message that is *supposed* to be expressed, whether or not it actually is expressed.

Once we take the alleged semantic properties of genes as seriously as this, some subtle questions arise. If genes are carrying a message in this sense, the message apparently has a prescriptive or imperative content, as opposed to a descriptive or indicative one. Genes contain instructions, not descriptions. Their "direction of fit" to their effects is such that if genes and phenotype do not match, what we have is a case of unfulfilled instructions rather than inaccurate descriptions.

Several philosophers and biologists have argued that much informational talk about genes uses a richer concept than Shannon's, but this concept can be given a naturalistic analysis. It is not a lapse back into unscientific teleological thinking. One way

to proceed is to make use of a rich concept of biological *function*, in which the function of an entity derives from a history of natural selection (Sterelny, Smith, and Dickison 1996, Maynard Smith 2000, Shea forthcoming). This sort of move is familiar from the philosophy of mind, where similar problems arose in the explanation of the semantic properties of mental states. When an entity has been subject to and shaped by a history of natural selection, this can provide the grounding for a kind of purposive or normative description of the causal capacities of that entity. To use the standard example (Wright 1976), the function of a heart is to pump blood, not to make thumping sounds, because it is the former effect that has led to hearts being favored by natural selection. The hope is that a similar "teleofunctional" strategy might help make sense of the semantic properties of genes, and perhaps other biological structures with semantic properties.

There are several ways in which the details of such an account might be developed (Godfrey-Smith 1999), some focusing on the evolved functions of the genetic machinery as a *whole*, and others on the natural selection of *particular* genetic elements. All versions of this idea offered so far have problems of detail. One problem is that there is no overall connection between biological function and semantic properties; having a function in the rich historical sense is not generally sufficient for having semantic properties. Legs are for walking, but they do not represent walking. Enzymes are for catalyzing reactions, but they do not instruct this activity. There are things that legs and enzymes are *supposed* to do, but this does not make them into information-carriers, in a rich beyond-Shannon sense. Why should it do so for genes?

Sterelny, Smith and Dickison seem to think there is a quite intimate connection between evolutionary function and semantic properties in the case of biological structures that have been selected to play a causal role in *developmental* processes. They argue that genes, in virtue of these functional properties, represent the outcomes they are supposed to produce. They add, however, that any non-genetic factors that have a similar developmental role, and have been selected to play that role, also have semantic properties. So Sterelny, Smith and Dickison want to ascribe very rich semantic properties to genes, but not only to genes. Some non-genetic factors have the same status.

Proposals that appeal to evolutionary design to "enrich" the informational properties of genes have problems of detail, but they also have attractive features. It is

striking that John Maynard Smith, when he came to grapple with the status of his enthusiasm for informational concepts in biology, opted for something along these lines (2000). The resulting overall picture has good structural features. We would have a loose, uncontroversial Shannon sense of information that applies to all sorts of correlations, and an "overlay" of richer semantic properties in cases where we have the right kind of history of natural selection. Genes and a handful of non-genetic factors would have these properties; most environmental features that have a causal role in development would not. The neatness of the resulting picture provides, for some people, good reason to persevere with some account along these lines.

So far in this section I have mostly discussed the concept of information; there has not been much talk of "coding." And the ideas discussed so far do not put any emphasis on the special features of genetic mechanisms themselves, such as the combinatorial structure of the "genetic code." But surely these features of genetic mechanisms provide much of the underlying motivation for the introduction of semantic concepts into biology? It might seem so, but a lot of discussions have in effect treated this as an open question. As noted above, the enthusiasm for semantic characterization of biological structures extends back before the genetic code was discovered. (See Kay 2000 for a detailed historical treatment.) But another line of thought in the literature, overlapping with the ideas above, has focused on the special features of genetic mechanisms, and on the idea of "genetic coding" as a contingent feature of these mechanisms.

Both I (2000) and Griffiths (2001) have argued that there is one highly restricted use of a fairly rich semantic language within genetics that is justified. This is the idea that genes "code for" the amino acid sequence of protein molecules, in virtue of the peculiar and contingent features of the "transcription and translation" mechanisms found within cells. Genes specify amino acid sequence via a templating process, that involves a regular mapping rule between two quite different kinds of molecules (nucleic acid bases and amino acids). This mapping rule is *combinatorial*, and apparently *arbitrary* (in a sense that is hard to make precise – Stegmann 2004).

The argument is that these features make gene expression into a causal process that has significant analogies to various paradigmatic symbolic phenomena, such as the use of natural language. Some have argued that this analogy becomes questionable once

we move from the genetics of simple prokaryotic organisms (bacteria), to those in eukaryotic cells. This has been a theme of Sarkar's work (1996). Mainstream biology tends to regard the complications that arise in the case of eukaryotes as mere details, that do not compromise the basic picture we have of how gene expression works. An example is the editing and "splicing" of mRNA transcripts into a processed mRNA that is used in translation. This is a biologically important process, and it does make the DNA a much less straightforward predictor of amino acid sequence, but it can be argued that this does not much affect the crucial features of gene expression mechanisms that motivate the introduction of a symbolic or semantic mode of description.

So the argument in Godfrey-Smith (2000) and Griffiths (2001) is that there is one kind of informational or semantic property that genes and only genes have: coding for the amino acid sequences of protein molecules. But this relation "reaches" only as far as the amino acid sequence. It does not vindicate the idea that genes code for whole-organism phenotypes, let alone provide a basis for the wholesale use of informational or semantic language in biology. Genes can have a reliable causal role in the production of a whole-organism phenotype, of course. But if this causal relation is to be described in informational terms, then it is a matter of ordinary Shannon information, which applies to environmental factors as well. This restriction of semantic language to the first steps in the causal chain in protein synthesis is essentially along the same lines as Crick's (1958) view.

In this section I have distinguished one line of thought that looks at Shannon information and its "enriched" relatives, and another line of thought that looks at the peculiar features of the mechanisms of gene expression, and the original narrow idea of a "genetic code." But the two lines of thought can be married in various ways. Maynard Smith, in response to problems with his teleo-functional account, appealed at one point to some special features of genetic mechanisms, including the apparent "arbitrariness" of the genetic code. This idea has been popular but is hard to make precise. The key problem is that any causal relation can look "arbitrary" if it operates via many intervening links. There is nothing "arbitrary" about the proximal mechanisms by which a molecular binding event occurs. What makes the genetic code seem "arbitrary" is the fact that the mapping between base triplets and amino acids is mediated by contingent features of the

sequences of tRNA molecules, and of the enzymes that bind amino acids to tRNA molecules. Because we often focus on the "long-distance" connection between DNA and protein, and pay less attention to the intervening mechanisms, the causal relation appears arbitrary. If we picked out and focused on steps in any other biological cascade that are separated by three or four intervening links, the causal relation would look just as "arbitrary." Here it is also significant that the standard genetic code is turning out to have more systematic and non-accidental structure than people had once supposed (Knight, Freeland, and Landweber 1999).

I will discuss three more topics, in a more self-contained way, to finish the survey. The first is the idea that genes contain a *program*, in a sense analogous to that in computer science (Mayr 1961, Jacob and Monod 1961, Moss 1992, Marcus 2004). This idea has not been discussed in such a concerted way by philosophers, though it is seen constantly in biological discussion. Here the focus is more on the control of *processes* by genes, as opposed to the specification of a particular *product*.

The "program" concept seems to be applied in biology in an especially broad and unconstrained way, often guided only by very vague analogies with computers and their workings. First, we might isolate a *very* broad usage, in which talk of programming seems merely aimed at referring to the intricate but orderly and well-coordinated nature of many basic processes in biological systems. Here, the most that talk of "programs" could be doing is indicating the role of evolutionary design. An example might be talk of "programmed cell death" in neuroscience, which is a very important process within neural development that could just as accurately be described as "orderly and adaptive cell-death in accordance with evolutionary design."

Secondly, however, we might isolate a sense in which talk of "programs" in biology is driven by a *close* analogy between some biological process and the *low-level* operation of modern computers. One crucial kind of causal process within cells is cascades of up and down-regulation in genetic networks. One gene will make a product that binds to and hence down-regulates another gene, which is then prevented from making a product that up-regulates another... and so on. What we have here is a cascade of events that can sometimes be described in terms of Boolean relationships between variables. One event might only follow from the conjunction of another two, or from a

disjunction of them. Down-regulation is a kind of negation, and there can be double and triple negations in a network. Gene regulation networks have a rich enough structure of this kind for it to make sense to think of them as engaged in a kind of computation. Computer chip "and-gates," neural "and-gates" and genetic "and-gates" have some genuine similarities. Most other biological processes, though just as much the product of evolutionary design, do not have a structure that motivates this sort of computational description. And once again we find, as in the case of "genetic coding," that the domain in which this computational language is well-motivated, when applied to genes, is confined to the cellular level. Less elaborate cascades of this kind can also be found in the endocrinological (hormone-using) systems within the body. Here too, informational language can seem naturally applicable, and may be justified by a similar line of argument.

The second of the three topics I will discuss to finish this section is the link between informational description and genetic determinism. A number of critics have argued that the informational or semantic perspective on gene action fosters or encourages naive ideas about genetic determinism (Oyama 1985, Griffiths 2001). Others think that genetic determinism, when it is false, is an ordinary error about causal relations that has no particular link to the informational description of those relations. I side with the critics here, who say that there is something definite about informational description of genes that encourages fallacies about genetic causation. The key point has been summarized by Griffiths. He notes that in complex systems, almost all causal factors are context-dependent, and usually it is not hard to remember this. If we think in ordinary causal terms, it is straightforward to note that a genetic cause will only have its normal effects if accompanied by suitable environmental conditions, and an environmental cause will only have its normal effect if accompanied by suitable genetic conditions. (If the sensitivity on either side is high, then talk of "normal" effects itself may be misleading.) But, Griffiths suggests, the informational mode of describing genes (and other factors) fosters the *appearance* of context-independence. "Genetic causation is interpreted deterministically because genes are thought to be a special kind of cause. Genes are instructions – they provide information – whilst other causal factors are merely material.... A gay gene is an instruction to be gay even when [because of other factors]

the person is straight" (2001, pp. 395-96). So the idea is that the inferential habits and associations that tend to go along with the use of informational or semantic concepts lead us to think of genes as having an additional and subtle *kind* of extra causal specificity. These habits can have an effect even when people are willing to overtly accept context-dependence of (most) causes in complex biological systems. Relatedly, the idea of internal genetic messages may also foster a tendency towards a kind of essentialist thinking; the meaning of the internal message tells us what the "true nature" of the organism is, regardless of whether this nature is actually manifested.

My final topic in this survey is the most strong and tendentious employment of informational language for genes, which arises in the context of evolutionary biology. It has been common for some time to say that, in the evolutionary context, we should think of a gene in terms of its sequence, which is preserved over many replication events, and not in terms of particular DNA molecules, which come and go (Dawkins 1976, 1986). The idea that sequence can be preserved across changes in the underlying molecules is certainly reasonable and important. But this message, important as it is, has been expressed in extreme and philosophically mysterious ways by some theorists. G. C. Williams (1992), for example, holds that because of these facts about the preservation of gene sequence across changes in molecules, we should think of information as a kind of fundamental ingredient of the universe, along with mass and energy, that exists in its own "domain." This makes the causal connections between the informational domain and the ordinary physical domain quite mysterious, and Williams himself finds this an important problem. But the appearance of a problem arises only from an unnecessary reification of information. We can instead say that what has been learned from work on the evolutionary features of genes is that various different physical objects can share their informational properties. These informational properties are explicable in terms of the lower-level physical properties of the objects, and the contexts in which the objects are embedded. Such a view does raise some further questions, but it does not introduce the idea of information as a separate "stuff" whose relations to ordinary physical things are tenuous and problematic.

The enthusiasm for a reified treatment of information can lead to a other theoretical problems in biology. Some of the recent advocates of "intelligent design"

creationism have tried to use the special and mysterious properties of information to mount anti-Darwinian arguments (Dembski 2001, criticized in Godfrey-Smith 2001). These arguments have no real force. Indeed, the resulting views tend to *less* plausible than earlier versions of the argument from design, because even routine and low-level forms of evolution by natural selection, such as the evolution of drug-resistance in bacteria, tend to be ruled out as impossible in principle. But the informational terminology in which the arguments are expressed lends them a spurious appearance of rigor.

3. The Next Steps?

In this section I will cautiously introduce some ideas that approach the whole problem somewhat differently. I will motivate the change in tack by asking what looks like an odd question. Is the informational or semantic description of genes *metaphorical*, or not? This should be an easy question to answer, but in fact seems to be surrounded by uncertainty. On the one hand, biologists sometimes say that the introduction of an informational framework was a crucial theoretical advance. This suggests that it is not at all a metaphor. If electrical charge, and entropy, were crucial theoretical advances in their day, it was not by being metaphors. But if one presses hard on what these informational properties are supposed to be, especially once we get beyond the simple idea of a combinatorial mapping from nucleic acids to amino acids, it is common to encounter a retreat to the idea of genetic information as a metaphor. It is not literally true that genes are programming development or representing the whole-organism phenotype, but this is a metaphor that has proven invaluable to biology.

Of course, we have to expect some vagueness here. And we can't expect biologists to be experts on the analysis of literal and non-literal language. But what makes the situation odd is the fact that if someone tried to carefully adjudicate this question, they would run immediately into the fact that in the case of ascriptions of semantic properties, there is no clear and well-understood border between literal and metaphorical. There is not a clear and well-demarcated sense of what the literal domain is, to which metaphorical cases are being compared.

The same problem arises, to some extent, in cognitive science, which is often based on the idea that the mind/brain can be seen as a computer. Does this mean: there is such a thing as computation, and the brain literally does it? Or is the idea of neural computation something more like a metaphor? The abstract theory of computation, within mathematics, is not especially helpful for answering this question (Smith 2002).

In the case of computation in cognitive science, the question can deflected initially by saying that computation is being treated as a "model" for the mind. But the term "model" is so ambiguous that this does not help much. Sometimes "model" means a provisional and cautiously defended theory. This does not help here because caution is not the issue. We want to know whether information-processing, computation, representation (etc.) are real natural kinds that brain activity – and genetic activity – might be literal instances of, or whether some other story about the role of these concepts has to be told.

In the remainder of this section I will sketch one alternative story of this kind. It is designed to contrast with the simpler idea that informational properties are definite but elusive properties that genes either do or do not have. Instead, informational description of genes is motivated by a family of factors, which I will group into three categories. First, it is motivated by some real and uncontroversial features of genes and DNA themselves, that would not alone be sufficient to motivate an elaborate informational description. Second, the use of informational and semantic language introduces into biology a particular "causal schematism," derived from everyday contexts in which symbols are used. The schematism functions as a model, in a sense discussed in some recent philosophy of science.

Thirdly, the informational framework reflects and reinforces a commitment to a way of demarcating the scientifically important features of genes and associated mechanisms. The framework foregrounds one set of properties and backgrounds another, and the properties of genes and other molecules that are being foregrounded are *sequence* properties, as opposed to all their other chemical properties. The result of this analysis is an account of the role of informational language in biology that is more focused on the entire disciplinary role of the informational framework, and less on specific informational properties that might or might not be real.

I will say more about each of the three categories in turn. First, the informational framework is motivated, of course, by some real and uncontroversial features of genes and DNA themselves. Some of these were highlighted in previous section, and they include the combinatorial structure and regularity of the mapping rule from nucleic acids to amino acids. But some motivation may also come from a feature of DNA that is not so often remarked on in this context. This is the passivity, or comparative inertness, of DNA. Here we focus on some facts about what DNA does *not* do, as well as what it does.

The evolution of DNA as a repository of sequence information is often said to be due in part to its chemical stability. Origin of life work emphasizes that fact that RNA is a good initial replicator molecule because it has some enzymatic activity, but DNA is more stable once proteins have been developed for enzymatic work. And in modern cells, DNA does not *do* very much in chemical terms; almost of all its effects go via a particular indirect causal pathway by which DNA sequence is transcribed and translated. (The main exception to this is DNA's direct interaction with transcription factors, in gene regulation.) Proteins, as is always noted, do most of the actual chemical work in the cell. DNA specifies amino acid sequence, and does not do much else. So to call DNA an "informational" molecule, in a modern context, is often a gesture towards what it does *not* do, as well as to what it does.

My suggestion for a second set of motivations is more tendentious. It involves a general analysis of when and why people introduce semantic concepts (including information) into scientific and other explanatory contexts. The suggestion is that the use of these concepts is generally guided – not always consciously – by the postulation of an analogy between a particular everyday form of symbol use, and the domain that the theorist is trying to understand. This analogy can be very partial, while still exerting influence on how the phenomena are described and understood.

How does the analogy work? A central aspect of everyday symbol use is that one object is used to "stand for" another. More precisely, a person guides behavior directed on one object or domain by attending to the state of another. This is the schematic core of everyday symbol use, and it shows up abstractly in many philosophical analysis of semantic phenomena (eg., Millikan 1984), as well as in models of signaling games and the evolution of meaning (Skyrms 1996). This basic pattern is also installed in the basic

picture that Shannon used in his theory of information: we have a source, and a signal whose state can be consulted to learn something about the source.

A central feature of this "causal schematism" is the distinction between some mechanism that reads or consumes the signal, and the signal itself. In genetic case, the idea that semantic description is guided by this model is quite helpful. First, we see that the basic cell-level machinery of transcription and translation is, in fact, a fairly good instance of the schematic structure in question. The ribosomal/tRNA machinery is, in effect, a reader or consumer of nucleic acid sequence, with the function of creating protein products that will have a variety of uses elsewhere in the cell. We also see that this realization of the causal schematism *only* applies at the cell level, at the level at which the transcription and translation apparatus shows up as a definite part of the machinery. So we see why it is true – if it is true, as I think it is – that the use of informational or semantic language in explaining how protein molecules are made is legitimate and well-motivated, while the use of this language when talking about the role of genes in producing whole-organism phenotypes is not. Once we think in terms of the influence of analogy and a causal schematism here, we can also note a connection to the discussion of the comparative chemical "passivity" of DNA discussed earlier in this section. Paradigmatic cases of messages in everyday life are rather physically passive, too, having their significant effects only via their interpretation by a reader or consumer.

This second category of factors motivating informational description of genes involves a kind of model-based theorizing, in a sense that was developed for the analysis of very different parts of science (Giere 1988, Godfrey-Smith forthcoming, Weisberg forthcoming). The term "model" gives us some definite purchase here after all.

My third category involves a role for the informational framework that is not part of a causal hypothesis, a posited mechanism, or anything of that kind. Instead, it involves a commitment to a way of demarcating and categorizing an entire domain. Via the informational framework, a commitment is made to the importance of one set of properties and the unimportance of another. One set of properties of biological molecules is foregrounded, by introduction of a language that can naturally accommodate them, while another set of properties is backgrounded. What is being foregrounded is *sequence* properties, as objects of study, as opposed to all the other chemical properties of genes

and associated biological molecules. The suggestion is that rather than attributing some particular causal powers to DNA sequence, the informational framework often functions to make sequences in general the primary focus of study. What results is a form of abstraction akin to that seen in statistical mechanics; there is a focus on a distinctive level of description and a particular set of statistical features of interactions between particles, abstracting away from lots of other properties (Griesemer 2005). The informational framework also brings with it a set of conceptual tools that are suited for the analysis of sequence properties, as opposed to other chemical properties. However, it should be added here that there are conspicuous uses of informational language in biology in contexts where sequence properties are not treated as central, for example in the description of hormonal signaling. In these cases I would emphasize the second of the three factors discussed in this section, the role of a causal schematism derived from public symbol use.

Suppose the actual patterns of use of informational language in genetics are in fact guided by factors like these, in a context-sensitive mixture. The use of the informational framework is guided by some real features of genetic mechanisms, but also by application of an schematic causal model that guides many or most uses of semantic language. It reflects and reinforces a general disciplinary focus on sequence properties as opposed to others. This would steer us away from the idea that there is some definite but hidden set of properties being posited by such language, that might or might not be real. I will leave somewhat open how this set of ideas relates to the more standard lines of thought outlined in the previous section. In some ways, the two can complement each other. In other ways, there is probably some tension.

<center>*   *   *</center>

References

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

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R. (1986). *The Blind Watchmaker*. New York: Norton.

Dembski, W. (1997). "Intelligent Design as a Theory of Information." *Perspectives on Science and Christian Faith* 3: 180-190.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge MA: MIT Press.

Francis, R. (2003). *Why Men Won't Ask for Directions: The Seductions of Sociobiology*. Princeton: Princeton University Press.

Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: Chicago University Press.

Godfrey-Smith, P. (1999). "Genes and Codes: Lessons from the Philosophy of Mind?" in V. Hardcastle (ed.), *Where Biology Meets Psychology: Philosophical Essays*. Cambridge MA: MIT Press.

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

Godfrey-Smith, P. (2001). "Information and the Argument from Design." In R. Pennock (ed.), *Intelligent Design Creationism and its Critics: Philosophical, Theological and Scientific Perspectives*. Cambridge MA: MIT Press, 2001, pp. 575-596.

Godfrey-Smith, P. (forthcoming). "The Strategy of Model-Based Science." To appear in *Biology and Philosophy*.

Griesemer, James. (2005), "The Informational Gene and the Substantial Body: On the Generalization of Evolutionary Theory by Abstraction", in M. Jones and N. Cartwright (eds.), *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences*. Amsterdam/New York: Rodopi.

Griffiths, P. E (2001). "Genetic Information: A Metaphor in Search of a Theory," *Philosophy of Science* 68: 394-412.

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

Jablonka E. (2002) "Information: Its Interpretation, Its Inheritance and Its Sharing." *Philosophy of Science* 69: 578-605.

Jacob, F. and J. Monod (1961). "Genetic Regulatory Mechanisms in the Synthesis of Proteins," *Journal of Molecular Biology* 3: 318-356.

Kay, L. (2000). *Who Wrote the Book of Life? A History of the Genetic Code.* Palo Alto: Stanford University Press.

Kitcher, Philip. S. (2001), "Battling the Undead: How (and How Not) to Resist Genetic Determinism", in R. Singh, C. Krimbas, D. Paul and J. Beatty (eds.), *Thinking About Evolution: Historical, Philosophical and Political Perspectives*. Cambridge: Cambridge University Press.

Knight, R., S. Freeland and L. Landweber (1999). "Selection, History and Chemistry: The Three Faces of the Genetic Code." *Trends in Biochemical Sciences* 24: 241-247.

Marcus, G. (2004). *The Birth Of The Mind: How A Tiny Number of Genes Creates the Complexities of Human Thought*. New York: Basic Books.

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

Mayr, E. (1961). "Cause and Effect in Biology." *Science* 134: 1501-6.

Millikan, R. (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", in D. Hull, M. Forbes and K. Okruhlik (eds.), *PSA 1992, Volume 1*. East Lansing: Philosophy of Science Association, pp. 335-348.

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

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

Shannon, C. (1948). "A Mathematical Theory of Communication," *Bell Systems Technical Journal* 27: 279-423, 623-656.

Shea, N. (forthcoming). "Representation in the Genome, and Other Inheritance Systems."

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

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Smith, B. (2002) "The Foundations of Computing." In M. Scheutz (ed.), *Computationalism: New Directions*. Cambridge MA: MIT Press

Stegmann, U. (2004). "The Arbitrariness of the Genetic Code." *Biology and Philosophy* 19: 205-222.

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

Weisberg, M. (forthcoming). "Who is a Modeler?" To appear in *British Journal for the Philosophy of Science*.

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

Wright, L. (1973). "Functions." *Philosophical Review* 82: 139-168.