Genetic Information as Instructional Content*

Ulrich E. Stegmann†‡

The concept of genetic information is controversial because it attributes semantic properties to what seem to be ordinary biochemical entities. I argue that nucleic acids contain information in a semantic sense, but only about a limited range of effects. In contrast to other recent proposals, however, I analyze genetic information not in terms of a naturalized account of biological functions, but instead in terms of the way in which molecules determine their products during processes known as template-directed syntheses. I argue that determining an outcome in a certain way is constitutive for being an instruction. On this account, the content of genetic information is identified with the template's properties, which determine the product in the way constitutive for instructions.

1. Introduction. Genetics and developmental biology treat DNA as containing information about aspects of the organism and its development. It is contentious exactly which aspects these are, but anything from protein structure (e.g., Crick 1958; Stent 1981) to complex phenotypic traits (e.g., Jacob [1970] 1974) has been suggested at some point. Whatever its precise content, information is thought to be stored in DNA and to be capable of being expressed in a right or wrong way. This idea is challenging: Aboutness, misinterpretation, and storage are semantic properties, but such properties are not posited by ordinary biochemistry. How, then, can semantic properties be integrated with biochemistry and molecular biology? If semantic properties are not normally attributed to molecules and

---

*Received April 2004; revised July 2005.

†Philosophy Department, King's College London, Strand, London WC2R 2LS, England, United Kingdom; email: ulrich.stegmann@kcl.ac.uk.

‡I am grateful to Martin Carrier, Ulrich Krohs, Eva Neumann-Held, David Papineau, and Richard Samuels for their comments on different versions of the manuscript, and to Wilfried Meyer-Viol for helping me to think about structure-preservation. I also wish to acknowledge the comments of two referees. Research was supported by the Deutsche Akademie der Naturforscher Leopoldina through funds from the German Ministry of Education and Research (BMBF-LPD 9901/8-83).

Philosophy of Science, 72 (July 2005) pp. 425–443. 0031-8248/2005/7203-0002$10.00

chemical processes, why should DNA and developmental mechanisms be exceptions? Is genetic information an empty or misleading metaphor as some have alleged (e.g., Sarkar 1996; Kay 2000; Keller 2000)?

To a large extent, the recent philosophical discussion has been framed in terms of the claim that genes contain information about *phenotypic* traits (Sterelny et al. 1996; Maynard Smith 2000; Griffiths 2001; Godfrey-Smith 2003; Moss 2003; Sarkar 2003; Wheeler 2003). This is probably the most far-reaching and controversial idea from a biological point of view. Surely, many biologists endorse it (e.g., Jacob [1970] 1974; Alberts et al. 2002; Davidson et al. 2002; Wolpert et al. 2002), but many others deny it (e.g., Stent 1981; Kauffman 1993; Noble 2002). One of the most prominent critics of this view has been Sydney Brenner (see de Chadarevian 1998). He and other skeptics emphasize that phenotypic traits are the most distant of genetic effects and those requiring the largest and most diverse set of additional causal factors to bring them about.

Here I put this claim about phenotypic traits to one side. For the more fundamental question seems to be whether DNA contains semantic information about *any* traits or effects. If it does, then it is likely to be information about, at least, the effects marked out by the least ambitious and least controversial claims regarding genetic information. The topic of this paper is, therefore, the legitimacy of these modest claims. These claims center around the idea that DNA carries information about some of its proximate effects, and they feature processes like DNA synthesis, RNA synthesis, and protein synthesis. Assertions like *DNA contains information about proteins* seem acceptable even to philosophers rejecting semantic information about phenotypic traits (Godfrey-Smith 2000b; Griffiths 2001; Moss 2003). However, with the exception of Godfrey-Smith's work (Godfrey-Smith 2000b, 2003), the adequacy of such limited claims has rarely been evaluated.

The debate about genetic information is closely linked to debates over genetic determinism and the concept of a gene.[1] The argument in this paper, though, is restricted to the question of whether *molecules* like DNA, which may or may not be (parts of) genes, contain semantic information. I will argue that they do: As templates for the synthesis of macromolecules, nucleic acids determine their products in a way that is constitutive for instructions in general. It is therefore legitimate to attribute instructional content to molecular templates. Thus, my proposal falls broadly into Godfrey-Smith's (1999) category of 'developmental role theories' of genetic information.

---

[1] For recent work on genetic determinism see, e.g., Kitcher 2001; Griffiths forthcoming. For recent literature about gene concepts see, e.g., Beurton et al. 2000; Neumann-Held 2001; Griffiths 2002; Moss 2003; Stotz and Griffiths forthcoming.

I start by outlining the most limited claims regarding genetic information. I analyze the way in which molecular templates determine the distinguishing features of their products. I then argue that everyday examples of instructions share this way of determining their outcomes, and further, that the controversial semantic features of DNA can be understood in terms of the properties of instructions. The last sections consider possible objections and two thought experiments from the recent literature. But first, a few words about what seem to be three major desiderata for any theory of genetic information.

2. Desiderata. It is common ground among philosophers of biology that satisfactory accounts of genetic information must capture its semantic nature (for an exception, see Sarkar 2003) and that accounts based on the mathematical theory of information fail to do so.[2] Therefore, explaining the following three semantic features is the first desideratum. First, the informational entities must be 'intentional' in a broad sense: they must be *about* some other entities. Second, the idea of misrepresentation must make sense (Sterelny and Griffiths 1999); for example, molecular outcomes must be capable of being *erroneous* in a stronger sense than just being an uncommon effect (Godfrey-Smith 1999). (A strong sense of error does not imply aboutness, for something can malfunction without being about anything.) Third, there must be an account of what it means for genetic information to be *stored* in DNA. One aspect of this idea is the persistence of a semantic content over time; another is the view that genes contain information that may not currently be used or expressed (Sterelny and Griffiths 1999). By contrast, it would not make sense to distinguish between cases in which smoke stores, rather than expresses, its (natural) information about fire.

The second desideratum for justifying the attribution of genetic information is a successful naturalization of its semantic features. It is difficult to see how semantic information could be legitimately attributed to any biological entity if it were not possible to account for it in nonsemantic terms. A glance at some recent theories of genetic information confirms that this desideratum is widely shared. According to Godfrey-Smith (2000b), DNA contains genetic information in virtue of its peculiar causal role in development. On teleosemantic accounts (Sterelny et al. 1996; Maynard Smith 2000), a stretch of DNA is about a certain developmental effect in virtue of this effect being the (etiological) biological function of this piece of DNA: It has genetic information about an effect just in case

---

[2] The shortcomings of this theory with respect to an account of genetic information have been discussed by a number of authors, e.g., Yockey 1992; Sarkar 1996; Godfrey-Smith 1999; Sterelny and Griffiths 1999; Maynard Smith 2000.

it has been selected to produce this effect in the past. And Jablonka (2002) argues that an allele is a source carrying information (about a state for a receiver), because the receiver's reaction depends on the source's structural properties, and because it tends to adapt the receiver to that state.[3]

The third desideratum proves less straightforward. Biologists typically think that genetic information is contained only in a rather small set of biological macromolecules, i.e., in nucleic acids and (perhaps) proteins. It seems plausible, therefore, that a satisfactory philosophical account should *restrict* the attribution of genetic information to these molecules. At first sight, this adequacy condition appears to underwrite a major complaint against teleosemantic accounts: If all that is necessary for containing information is to have a biological function, any trait selected for a particular function would contain information about its function (Godfrey-Smith 1999; Griffiths 2001).

This criticism is justified, but I think the desideratum should be relaxed. A theory of genetic information is not intended as a mere explication of the biologists' actual use of this term, but rather as a substantial enquiry into the presence of semantic features. It may turn out that other, non-genetic, developmental factors share the very properties in virtue of which DNA is attributed semantic information. Indeed, Griffiths' (2001) parity thesis states that this will be the case on all satisfactory accounts. From his perspective, it seems that accounts allowing for such a symmetric distribution of semantic information are desirable, and Jablonka's (2002) proposal appears to follow this approach. Regardless of whether one accepts these accounts, one ought to acknowledge that the limitation of semantic information to a few kinds of biomolecules may be the result of a theory of genetic information, but should not be its precondition. On the other hand, one need not conclude that the theory should apply to *every* biological trait. Rather, it seems sensible to refine the third desideratum: The biologists' paradigm cases of informational processes should be preserved and the theory's applicability may be modified and extended beyond these cases within reasonable limits. Attributing semantic information to legs because they were selected for walking is beyond reasonable limits. This refined standard retains, I believe, the force of the criticism against teleosemantic accounts.

---

[3] I take Jablonka (2002) to suggest that information sources contain information about the state or circumstance to which the receiver tends to adapt when receiving information. It seems that these states are not understood as the causal effects of alleles, genes, or pieces of DNA. For, as Jablonka indicates, it is difficult to see in what sense a receiver like a cell could be said to adapt to effects like proteins or certain degrees of pigmentation. However, they can hardly be thought of as environmental states either, because alleles are supposed to be about some of their effects, not about environmental conditions. But then, it is unclear what kind of states are meant.

3. Modest Claims about Genetic Information. Limited claims about genetic information center around the idea that DNA contains information only about some of its most proximate effects, but not about phenotypic traits. An early and influential version of this idea postulates that nucleic acids contain the information for determining the linear order of both amino acids in proteins and bases in nucleic acids (Crick 1958). Proteins and nucleic acids acquire their linear order when they are synthesized from their components by using an already existing macromolecule as a template (‘template-directed synthesis’. This process will be explained below).

Genetic information in Crick’s sense exhibits the three semantic features outlined above. First, genetic information involves a content, i.e., it is information *about* the linear order of the product’s components. This is apparent from Crick’s suggestion that the “sequence [of nucleic acid bases] is a (simple) code for the amino acid sequence of a particular protein” (Crick 1958, 152). Another phrase suggesting the presence of a content is “hereditary information required for sequentialization” (144), where ‘sequentialization’ denotes the ordering of amino acids in proteins during protein synthesis. Furthermore, the following definition shows that genetic information is not only about the linear order of proteins, but also about the base sequence of nucleic acids: “Information means here the *precise* determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein” (153).

The second semantic feature concerns the possibility of error. Crick (1958) remarked that “mistakes” (140) may occur during protein synthesis and that the order in product molecules can be “right” (144) or “correct” (155). This suggests that genetic information can be expressed correctly or incorrectly in a sense not adequately captured by uncommon effects.

With respect to Crick’s (1958) sense of genetic information, it makes intuitive sense to distinguish between information being stored and expressed. DNA can be said to express its information whenever it serves as a template for synthesizing another nucleic acid or a protein. When it does not, it may still do so later, and then it is natural to say that DNA stores its information.

Thus, Crick’s (1958) claims about genetic information are modest insofar as DNA is thought to contain information only about the synthesis of its products. Nevertheless, this is information in a semantic, rather than a purely correlational, sense. Moreover, there seems to be something in the process of template-directed synthesis itself that motivates the attribution of semantic information. This process does not appear to be described semantically only in a derivative sense (as would be the case if synthesis were described as the copying or transmission of a message that is not about the product itself).

Godfrey-Smith thinks that, if template and product belonged to the same molecular class, they would bear a "coding relation" to one another only in a special, stipulative sense of 'coding' or 'representation', and that it would be "very unusual" to use such terminology in those circumstances (Godfrey-Smith 2000b; cf. Godfrey-Smith 2000a). However, this interpretation is not supported by views such as Crick's (1958), according to which DNA replication (an example of template-directed synthesis) is an informational process in its own right, although both template and product are composed of deoxyribonucleotides. Crick even considered as informational a hypothetical process in which, unlike in replication, the four kinds of deoxyribonucleotides pair up "like-with-like" (Keyes 1999). (Note also that replication and transcription involve *direct* specification of incoming nucleotides and still count as information transfers.) Perhaps, though, Godfrey-Smith's substantial point is correct, i.e., that ascribing semantic information to such processes is *unwarranted* (Godfrey-Smith 2000b). I leave this issue open for the moment.

To this day, Crick's (1958) claims about genetic information are probably the most uncontroversial, and they were never idiosyncratic. When speculating that the base sequence of nucleic acids is a code for the amino acid sequence in proteins, he added that "[t]his hypothesis appears to be rather widely held" (152). As Crick said, he sought to state explicitly what he took to be a common understanding among his colleagues at the time.

Moreover, information about the order of products ("sequence information") was the kind of information featuring in his "central dogma of molecular biology" (Crick 1958, 1970). The 'dogma' is concerned with the directions in which genetic information is supposed to be transferred. According to Crick (1970), genetic information is transmitted from DNA to DNA, RNA, and proteins, as well as from RNA to DNA, RNA, and proteins. Although information transfers from *proteins* to DNA, RNA, and proteins are "possible transfers" (562), they do not actually occur in nature.[4] As Keyes (1999) emphasized, Watson et al. (1965) were more explicit in associating the three frequent information transfers with three molecular processes: replication or duplication (DNA → DNA), transcription (DNA → RNA), and translation (RNA → protein).

Although the 'dogma' has been repeatedly and severely criticized (reviewed in Thieffry and Sarkar 1998), it is still portrayed as a fundamental

---

[4] In the original version of the central dogma (Crick 1958), the occurring transfers are called "possible" transfers and the nonoccurring transfers "impossible": "[The central dogma] states that once 'information' has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible."

biological principle in current textbooks (e.g., Nelson and Cox 2000; Alberts et al. 2002). More importantly, these criticisms do not challenge the status of replication, transcription, and translation as paradigms of informational processes. For example, Temin and Mizutani's (1970) discovery of reverse transcription (DNA synthesis from a RNA template) *added* an information transfer to Watson's version of the 'dogma'.[5] Similarly, RNA-editing and RNA-splicing (modifications of RNA sequences prior to protein synthesis), do not undermine the (alleged) informational nature of transcription and translation. For although these processes limit the predictability of amino acid sequences from DNA sequences (Thieffry and Sarkar 1998), they occur only after a primary RNA transcript (hnRNA) has been produced. So DNA may still contain information for the primary transcript, and information may be transferred from mature (edited and spliced) mRNA to protein.

I will argue that something like the limited claims expressed by Crick are justified because of the way in which templates determine their outcomes. I therefore now consider template-directed synthesis in more detail.

4. Template-Directed Synthesis. Linear macromolecules like DNA, mRNA, and proteins consist of specific, and often nonrepetitive, sequences of basic components. They are instances of 'ordered sets' or '*n*-tuples' whose elements are the chemical components making up (parts of) the molecule. For example, the DNA-stretch *A-C-C-T-G* is an instance of the ordered set $\langle A, C, C, T, G\rangle$ where *A* is adenine, *T* is thymine, *C* is cytosine, and *G* is guanine. Ordered sets are defined by three properties: the number, the kind, and the linear order of its elements. Let us see how these properties are determined.

Template nucleotide bases are chemically active, i.e., they can attract and form hydrogen bonds with incoming nucleotides. Hydrogen bonds tend to form only between chemically and spatially complementary bases, i.e., Watson-Crick base pairing is highly specific in a given context. The chemical properties of a template adenine, for example, are such that only an incoming nucleotide 'carrying' thymine pairs with it and is then integrated into the growing product chain. Such a pairing is an instance of chemical regularities of the sort 'If X, then Y^* is integrated' (for brevity, 'is integrated' will be omitted in the following), where X and Y^* are placeholders for types of template and product components, respectively

---

[5] Similarly, epigenetic inheritance (Thieffry and Sarkar 1998) and, at some point, prion multiplication (Keyes 1999) have been interpreted as challenging the 'central dogma'. These examples are supposed to show that there are *additional* kinds of information transfers. See Keyes (1999) for the distinction between Watson's and Crick's versions of the 'central dogma'.

(the latter indexed by '*'), and 'integrate' means both the binding of a free product component Y^* to X and the linking of Y^* to the rest of product molecule. In replication, the pairing regularities between template and product bases satisfy the one-to-one function $\{\langle A, T^*\rangle, \langle T, A^*\rangle, \langle C, G^*\rangle, \langle G, C^*\rangle\}$, where the first position of each 2-tuple represents a template base, and the second position represents an attracted free base, i.e., a component of the nascent product molecule.

In translation, there is no direct interaction between mRNA codons and amino acids but rather there is a mediation through the specificities of molecules like tRNAs and aminoacyl-tRNA-synthetases. One kind of codon specifies one kind of amino acid, and one kind of amino acid can be specified by more than one kind of codon ('degeneracy'). As observed by Gatlin (1972), these specificities satisfy a many-to-one function. The set of specificities is commonly known as the 'genetic code'.

I take it then that the *identity* of a product residue is determined by two factors. One factor consists in the 'If X, then Y^*'-type reactions between template and product components. As long as the identity of a component at a specific template position is unspecified, any of the 'If X, then Y^*'-reactions may occur at this position. Consequently, the corresponding product component may be any Y^*. A second factor is therefore required to determine the identity of the product component. This is the presence of a particular component, say A rather than G, at a given template position. Given an A-token in the template and the regularity 'If A, then T^*', the presence of A restricts the number of reactions that may otherwise occur at this site to one of the type 'If A, then T^*'. In this sense, a particular template base specifies which kind of 'If X, then Y^*'-reaction will occur. Since the reaction leads to the integration of one particular product component (e.g., T^*), the template base also specifies the identity of the product component (given a set of pairing regularities). Furthermore, the *number* of product residues is determined by the number of template residues that serve as 'substrates' for reactions of the sort 'If X, then Y^*'.

The factor determining a product's *linear order* is the template's linear order.[6] This idea was first stated, with respect to protein synthesis, as the "sequence hypothesis" (Crick 1958): "the sequence of bases along the

---

[6] There are other prima facie plausible mechanisms, for example, determination by the set of chemical regularities and the identity of template residues. However, the product is not determined by these factors, because they allow for products with many different linear orders. For instance, the set of chemical regularities specified by the set of template tokens $\{C, C, A, G, T\}$ may yield 4^5 different linear orders in the product, e.g., A^*-G^*-G^*-T^*-C^* and A^*-T^*-C^*-G^*-G^*. Furthermore, a product's linear order is not specified by the interactions of product components among themselves either—a point emphasized by Polanyi (1968), Sanger, and Monod (see Judson 1993).

nucleic acid determines the sequence of amino acids of the protein being synthesized" (158). The product elongates through a series of reactions that result in one component being linked to the next. The order in which these reactions occur depends on which template components are next to each other. For example, given a section *A*-*C*-*G* in the template, one reaction will be of the sort 'If *A*, then *T**', the next will be 'If *C*, then *G**', and so on. The result is the product-section *T**-*G**-*C**.

I now want to draw attention to the seemingly trivial fact that the factors determining a product are present before the product is synthesized. That is, the nucleotides composing the template are present before the synthesis starts and they are arranged in a linear order. Consequently, once a section of nucleic acid has been chosen to be a template, it is fixed *prior to synthesis* which 'If *X*, then *Y**'-type reactions will occur and in which order. As Pontecorvo (1958) observed, templating is a process in which "a collection of building blocks (say, of the four main types of nucleotide), which could arrange themselves into any one of all the possible sequences, becomes arranged into only one sequence *because of the prexistence of such a unique sequence.*"

This fact distinguishes templating from other biochemical processes. Consider biochemical pathways in which different organic molecules are produced or transformed in a certain order in time. In the Krebs cycle, for example, oxalacetate (*O*) accepts an acetyl residue yielding citrate (*C*); subsequently, citrate is isomerised to isocitrate (*I*); then isocitrate is oxidized to 2-oxoglutarate (*G*), etc., until the last step yields oxalacetate again and the cycle starts anew. Like linear macromolecules, we can describe this pathway as an ordered set $\langle O, C, I, G, \ldots \rangle$ and then ask which factors determine the identity and the order of the elements of this 8-tuple.

The identity of a molecule is determined like a product component in template-directed synthesis. For example, the fact that *O* is followed by *C* (rather than by *G*) is determined both by the presence of *O* and the regularity 'If *O*, then *C*' (the regularity here concerns a transformation, not a pairing).

However, the order of elements in the Krebs cycle is determined differently. In the Krebs cycle, each reaction product becomes the substrate of the subsequent reaction. *C*, for instance, is present as the result of the reaction transforming *O* into *C*, and *C* then becomes the substrate of the subsequent reaction that transforms it into *I*. Hence, the order of the 8-tuple $\langle O, C, I, G, \ldots \rangle$ results from a series of 'If *A*, then *B*'-type reactions in which the presence of one molecule (*A*) determines the identity of the next (*B*), and so on. The order is not determined by the properties of a molecule present before the reactions occur. It is not even necessary that the set of substrates is present before the cycle starts, because the substrates are produced as the cycle unfolds.

5. Naturalizing Instructional Content. So far, I attempted to show, in nonsemantic terms, that templates determine their product in a distinctive way. From here, we can get to the semantic features by looking at the properties in virtue of which items—like cooking recipes, computer programs, and the like—are instructions.

Recipes and programs provide specifications of the kind and order of operations, which if carried out, produce an outcome. For example, a recipe for a cake consists of a list of ingredients and a number of specifications that determine the kind and order of actions, which if carried out, produce the cake; and the recipe is provided before it is acted upon. Similarly, a computer program consists of a list of interconnected commands that specify the kind and order of operations that a computer will perform if it runs this program. Programs usually contain specifications of conditional form, and therefore, they rely on inputs to specify which operations to execute. However, the range of possible operations is specified by the program.

It seems, then, that programs and recipes share a peculiar way to determine their outcomes. They specify the kind and order of operations that will result in a certain outcome. Importantly, they specify this before the operations are performed: With a certain program loaded or a particular recipe in place, it is determined which operations (among a set of alternatives defined by other programs or recipes) will occur. The idea that operations are specified before they are performed appears to be the basis for our practice to distinguish between merely specified operations and those that are, in addition, executed.

I suggest that this way of determining an outcome is *constitutive* for instructions like recipes and programs. It provides a way to analyze, in nonsemantic terms, what it is for something to be an instruction. *P instructs* the making of Q just in case P determines Q by specifying, in advance, the kind and order of operations that produce Q if the operations are carried out. The proposal is to naturalize instructional processes as a certain way to determine an effect. If this picture is correct, then the *content* of an instruction consists in the specifications of the kind and order of operations that will result in a certain outcome.

Templates, I suggest, determine their products in just the same way as do recipes and programs. For we saw that a nucleic acid serves as a template for the synthesis of a product just in case it determines the kind and order of product components in the following way: The nucleic acid is present before the start of synthesis and it determines, through the kind and linear order of its components, the kind and sequence of 'If X, then Y^*'-type reactions that will occur. The nucleic acid section reduces the number of possible pairing reactions to one at each of its sites. Thus, the section can be said to specify, before the start of synthesis, the kind and order of reactions that will result in the product if the reactions occur. Hence, templates determine their products in the same way in which recipes and programs determine their outcomes.

If this mode of determination is indeed constitutive for instructional processes, then it is justified to say that molecular templates contain instructional content for the synthesis of their products. On this view, the instructional content of a molecular template consists in those of its (nonsemantic) properties that determine a product in the characterized way. These properties are the number, kind, and order of its components, i.e., the *n*-tuple instantiated by the template. As I want to show now, the three alleged semantic features of informational molecules can be construed as features of instructions in general.

6. The Semantic Features.

6.1. Aboutness. Programs and recipes are said to be about the procedures or operations yielding a specific outcome (rather than about the outcome itself).[7] A recipe for an apple pie is about how to bake an apple pie; a program for calculating arithmetic means is about how to calculate arithmetic means. Another way to say this is that manmade instructions provide the instructional content for achieving a certain outcome. Similarly, the aboutness of molecular templates can be construed as having instructional content for the synthesis of the product. A template *P* is about the production of *Q* insofar as *P* provides the instructional content for synthesizing *Q*.

Rather than saying that manmade instructions like recipes and programs provide instructional content for executing particular tasks, we sometimes say that they contain the *information* about how to bake a pie or how to calculate arithmetic means. In these cases, the terms 'information' and 'instructional content' are synonymous.

Similarly, instead of saying that a template provides the instructional content for the synthesis of its product, we may say that it carries the information for it. These formulations express the same idea. But, when expressed in terms of information, we capture what is meant by *genetic information.* I suggest that the genetic information of molecular templates is their instructional content. Further, a template 'carries' or 'contains' this information in the sense that the template is an instance of a certain *n*-tuple.

6.2. Error. We believe that recipes and programs can be implemented

---

[7] This observation was pointed out to me by Martin Carrier and Ulrich Krohs (personal communication).

correctly or incorrectly. This type of error does not consist in a deviation from what usually happens, but rather in a deviation from the instructional content. It is, therefore, a strong kind of error. A recipe or program is carried out 'correctly' just in case the instantiated actions or processes are those that are specified as part of the instructional content, and they are carried out in the specified order. Conversely, the instructions are carried out 'incorrectly' just in case the processes and their order are not realized as specified.

Similarly, it makes sense to say that molecular templates are implemented, or expressed correctly or incorrectly. The template's information is being expressed correctly if the occurring biochemical reactions are instances of the kind of reactions specified by the template components. The result of such reactions would be the "right" or "correct" (Crick 1958) order of the product. However, the right (or wrong) order may also arise by other means. For example, molecular 'proofreading' mechanisms (e.g., Alberts et al. 2002) replace 'mismatched' nucleotides with 'correct' nucleotides turning a wrong order into the right one. That is, they replace tokens of one nucleotide type (any type other than the one that would result from the pairing reaction specified by the template) by tokens of another type (the type specified by the template). Further, new product molecules usually contain more matched than mismatched nucleotides. But, even if mismatched nucleotides were the rule, they would be mismatched in the sense that they were nucleotides of a type other than the one specified by the template.

6.3. Storage. Particular pieces of nucleic acids may or may not currently serve as templates for the synthesis of a product molecule. The ideas of information storage and expression can be explained in terms of this difference. If a piece of nucleic acid does not currently contribute to synthesize a product, we may say that its instructional content remains 'unused' or 'stored'. By contrast, whenever a nucleic acid does serve as template, it makes sense to say that the information of the template is 'expressed'.

7. DNA and Protein. If accepted, the proposed account justifies the idea that templates contain semantic information about the synthesis of their *immediate* products. This conclusion still falls short of Crick's (1958) claim that DNA contains information about the linear order of amino acids in proteins, because the immediate products of DNA templates are not proteins but (primary) RNA transcripts. I now want to put forward an argument to show that, under some conditions, DNA does contain information about proteins.

What would the argument need to show? On the present account, a particular mRNA's instructional content for synthesizing a protein is the *n*-tuple it instantiates. The *n*-tuple's linear order is part of this content, and it determines the linear order of amino acids in the protein. Now suppose that the RNA's order has been *preserved* (in some sense) from the DNA template, such that the mRNA would possess the same kind of order as the DNA. Then, the linear order of the DNA template would ultimately determine the protein's order in an instructional way. For the DNA's order would determine the linear order of both the mRNA and the protein. Therefore, what needs to be shown is that there is a sense in which the DNA's linear order is preserved in the mRNA's order.

Consider the section *C-T-G-A* of a DNA template. The bases in the template section stand in neighbor relations to each other (i.e., *C* is next to *T*, *T* is next to *G*, and *G* is next to *A*), and these relations constitute this template's linear order. For each kind of DNA base there is exactly one 'corresponding' RNA base, because the biochemical complementarities between DNA and RNA bases satisfy the one-to-one mapping $\{\langle A, U^*\rangle, \langle T, A^*\rangle, \langle C, G^*\rangle, \langle G, C^*\rangle\}$. Now suppose the section *C-T-G-A* is transcribed, and we compare the template's linear order with the order of its product. We can then ask whether the neighbor relations among the bases of the DNA template hold among the corresponding RNA bases of the transcript. For example, the neighbor relation '*C* is next to *T*' (in the DNA template *C-T-G-A*) holds in the RNA product just in case the base corresponding to *C* is next to the base corresponding to *T*, i.e., just in case the product base G^* is next to product base A^* (the transcript is $G^*\text{-}A^*\text{-}C^*\text{-}U^*$). If the relations hold in RNA, we say that the template's linear order (its structure) has been *preserved* in the RNA's linear order.[8] If, by contrast, G^* would be next to, say, U^* in the product, then the neighbor relations among DNA bases would not hold among RNA bases (the transcript might be $G^*\text{-}U^*\text{-}A^*\text{-}C^*$). As a matter of empirical fact, the transcript will be $G^*\text{-}A^*\text{-}C^*\text{-}U^*$. Therefore, transcription is structure-preserving (as is translation).[9]

---

[8] I suspect that this is what biologists mean when saying that (bacterial) DNA sequences 'match with', 'correspond to', or are 'colinear with' the amino acid sequences of proteins.

[9] The genetic code is a many-to-one mapping from codons to amino acids caused by the specificities between codons, tRNAs, amino acids, and aminoacyl-tRNA-synthetases. Suppose the section *G-A-C-U-G-U-C-C-C* of an mRNA is translated into a protein. Under the relevant subset of the genetic code,

$$\{\langle\langle G, A, C\rangle, Asp^*\rangle, \langle\langle U, G, U\rangle, Cys^*\rangle, \langle\langle C, C, C\rangle, Pro^*\rangle, \ldots\},$$

different protein sections could be produced if, contrary to fact, the linear order of the mRNA would not fix the order of amino acids (*Asp*: aspartate; *Cys*: cystein; *Pro*: proline). For example, we might find the sequence $Pro^*\text{-}Asp^*\text{-}Cys^*$, which does not preserve the template's linear order. There is only one amino acid order that preserves

Having shown that the DNA's linear order is preserved in the mRNA's order, I conclude that the linear order of the protein is ultimately determined by the DNA's order. That is, the DNA's linear order is the instructional content, and hence, the genetic information that the DNA template provides for protein synthesis. Of course, DNA provides the instructional content for protein synthesis only if the RNA transcript is not altered before translation. Since the primary transcripts of eukaryotes are usually modified by RNA-splicing and RNA-editing, it may only be in organisms like bacteria where DNA actually does contain the information about the order of amino acids.

8. Disanalogies. Conceptualizing gene function in terms of instructions is an approach dating back to the early 1950s and taken up repeatedly ever since (see Kay 2000). For example, François Jacob ([1970] 1974) interpreted the genetic material as instructions. Just as in the present account, a part of the justification for his claim was the fact that the template's linear order determines its product's arrangement. However, the idea that has not been worked out previously is, I suggest, that molecular templates share a certain *way* of determining outcomes with recipes and programs, and that it is this kind of determining an effect that makes them instructional. Nonetheless, given that previous instructional accounts of genetic information did not succeed in demonstrating semantic properties of nucleic acids, one might well wonder why the account defended here should fare any better. In particular, one might point to disanalogies between templating and anthropogenic instructions and argue that they forbid attributing content to molecules.

Recipes and programs do not just bring about a particular outcome; they are designed to do so. Perhaps what makes them be about an outcome is the fact that they were designed to produce it. If so, then either templates are about their outcomes because they are biologically designed to do so, or if they are not so designed, then they cannot be about their products. In the latter case, even if templates determined their products in ways structurally similar to recipes, they would not possess semantic content. (Weber 2005 directs a similar line of argument against Maynard Smith 2000.)

Recipes and programs are usually formulated with a purpose, and many natural templates have the biological function to contribute to causing a certain outcome. But I deny that recipes and programs are instructions *in virtue of* having a purpose, and hence, that templates are instructions in virtue of their biological functions. Suppose the steps of the recipe or

---

the neighbor relations among mRNA triplets, *Asp**-*Cys**-*Pro**, and this is the sequence actually produced by organisms.

the program had been arranged randomly in advance. Then the ‘cake’ may hardly be edible, and the program may not perform anything sensible at all. Nevertheless, these outcomes would have been determined by specifying all individual steps and their order of occurrence such that the steps produce the outcome if they are carried out. I take it that we would still regard the computer output as the result of executing a pre-specified series of operations, and the inedible lump as the result of carrying out some (nasty) sort of instruction. Therefore, the attribution of instructional content is not derived from the function of a template, even if it has one.

Another concern is that recipes and programs seem to carry meaning or instructional content because they are linguistic entities, at least at some level. They are written commands and they instruct what they instruct in virtue of being meaningful sentences. This point can be developed into two objections. First, one may argue that only linguistic entities can carry instructional content in any proper sense. A less strict position would allow for an analogous sense of content in DNA. If this line is taken, DNA would convey information to the extent that it is akin to a language (as some say, DNA is like a string of words). Moreover, its instructional content should then be analogous to the meaning of words. The second objection is that my account fails on the latter score, because it equates instructional content with a template’s *n*-tuple. And certainly, whatever a word’s meaning, it will never be the *n*-tuple given by its letters or something similar. ‘Dog’ does not mean something like $\langle d, o, g \rangle$. From this point of view, I simply seem to confuse semantics with syntax.

Many manmade instructions involve verbal descriptions and they have linguistic meaning. It is even plausible to assume that they convey their instructional content by linguistic means. But this does not show that their instructional content somehow arises in virtue of them being linguistic entities. Suppose I happen to find someone’s phone number together with a wallpaper index that assigns one digit to each wallpaper color (0-blue, 1-green, etc). I also unpack my bag of colored marbles and arrange them on a thread. I do so by matching each phone number digit to a colored marble according to the digit-color index. Surely, the phone number is the instructional content determining which kind of marble to put on the thread and in which order. Yet phone numbers do not mean anything in a linguistic sense, nor are they designed as instructions to produce chains of marbles.

This example, one may object, does not show that instructional processes can be individuated in purely physical terms. From a physical point of view, the phone number is just ink on paper. Similarly, nothing about the physical properties of the wallpaper index determines that it should be adopted as a rule. These items feature in an instructional process only after someone has decided to use them in this way.[10]

I agree that in this case, a purely physical description does not capture the instructional nature of the process. But this is unproblematic for the following reason: On the present account, an instructional process is meant to be a process in which an item bears a certain structural relation to an outcome. The relata may be linguistic, chemical, or other entities. But it is right to insist that, if they are chemical entities, then they must bear this structural relation just in virtue of their physicochemical properties. I believe that this is the case in template-directed synthesis. The template determines the features of the product because of the physicochemical properties of nucleic acid bases and their arrangement. The process is driven by the causal powers of the template and product components (plus those of enzymes and so on).[11] This marks an important difference with recipes and chains of marbles, where human action is required to produce the outcome. But I do not see that it undermines the idea that, in the essential structural aspect, both molecular templates and recipes have a common way of determining their outcomes.

9. Implications. The recent literature provides two interesting thought experiments about the conditions of attributing semantic information to molecules. Let us see how to evaluate them on the present account. Godfrey-Smith (2000b) argues that information requires, among other conditions, that template and product belong to two different chemical classes (such as to require a nontrivial rule of specification, Godfrey-Smith 2000a). In his thought experiment, proteins are replicated by synthesis from a protein template. A given template amino acid specifies, via a connector molecule, the same kind of amino acid in the growing protein. He suggests that this process would not involve a "coding relation," because there would be "no biological reason" for postulating such a relation (Godfrey-Smith 2000b), other than in the weak sense of information theory (2000a).

However, on the account defended here, there are biological reasons for claiming that templating is informational even if it does not involve two different molecular classes; for, attributing genetic information to a template is to claim that it determines its product in a certain way, i.e., through the step-by-step realization of operations specified in advance. And, templates so determine their products whether or not the compo-

---

[10] This objection was raised by one of the referees.

[11] Perhaps metaphors like the *self-executing* "code-script" (Schrödinger 1944) and the "unmoved *mover*" (Delbrück 1971, my emphasis) go some way in capturing this idea.

nents belong to the same molecular class. Moreover, the different-classes requirement has problematic consequences. Neither replication nor transcription involve different chemical classes, and hence, paradigm cases of informational processes, such as heredity, come out as noninformational.

Note also that my account is not tailored to rescue the primacy of DNA over other developmental factors. Template components may belong to any chemical class. As long as a template specifies its products' arrangement by determining the kind and the order of individual 'If *X*, then *Y**'-type reactions, the process is one of instruction. This explains why, in a thought experiment by Neumann-Held (forthcoming), a chain of enzymes codes for its product (assuming that '*x* codes for *y*' here means '*x* has information for *y*'). Neumann-Held (forthcoming) envisions a process in which a molecule is synthesized from several different components in a series of reactions catalyzed by different enzymes. The enzymes are linearly fixed in a membrane, and the order of the enzymes determines the order in which the components are joined together. Thus, each enzyme can be assigned to the component it helps to join. The conclusion seems to be that this process is not relevantly different from translation and that it is, therefore, as informational as translation. I agree that it is informational, and I suggest it is so because the chain of enzymes determines the product by determining the kind and the order of individual 'If *X*, then *Y**'-type reactions.

10. Conclusion. The instructional account of genetic information makes a reductive claim about what it is for an entity to instruct something: Instructing consists in an advance specification of the kind and order of steps yielding a certain outcome if the steps are carried out. It claims, further, that molecular templates determine their products in this way. For in a process like replication, one molecule specifies, prior to synthesis, the kind and order of chemical interactions that determine the kind and linear order of the product's components. If this is accepted, then it is legitimate to describe the template's properties, which so determine the product, as the instructional content (or information) for the synthesis of the product.

This account does not apply to other likely paradigm cases of informational processes in molecular genetics. The 'start signals' for transcription and translation provide one example. It seems that signaling the start of transcription is regarded as a transfer of genetic information that leads to, but is not based on, template-directed synthesis. However, the instructional account is not meant to cover all paradigm cases. The expression 'genetic information' seems to be used differently in different contexts, and it is a separate issue to decide whether it is legitimately used in all those contexts. My aim here was to show that there is at least one class of molecular processes exhibiting the semantic properties of aboutness, error, and information storage.

REFERENCES

Alberts, Bruce, et al. (2002), *Molecular Biology of the Cell*. New York: Garland Science.

Beurton, Peter, Raphael Falk, and Hans-Jörg Rheinberger, eds. (2000), *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press.

Crick, Francis H. (1958), "On Protein Synthesis", *Symposia of the Society for Experimental Biology* 12: 138–167.

——— (1970), "Central Dogma of Molecular Biology", *Nature* 227: 561–563.

Davidson, Eric, et al. (2002), "A Genomic Regulatory Network for Development", *Science* 295: 1669–1678.

de Chadarevian, Soraya (1998), "Of Worms and Programmes: *Caenorhabditis elegans* and the Study of Development", *Studies in the History and Philosophy of Biological and Biomedical Sciences* 29: 81–105.

Delbrück, Max (1971), "Aristotle-totle-totle", in Jacques Monod and E Borek (eds.)*, Of Microbes and Life*. New York: Columbia University Press, 50–55.

Gatlin, Lila L. (1972), *Information Theory and the Living System*. New York: Columbia University Press.

Godfrey-Smith, Peter (1999), "Genes and Codes: Lessons from the Philosophy of Mind?", in Valerie Gray Hardcastle (ed.), *Where Biology Meets Psychology: Philosophical Essays*. Cambridge, MA: MIT Press, 305–331.

——— (2000a), "Information, Arbitrariness, and Selection: Comments on Maynard Smith", *Philosophy of Science* 67: 202–207.

——— (2000b), "On the Theoretical Role of 'Genetic Coding'", *Philosophy of Science* 67: 26–44.

——— (2003), "Genes Do Not Encode Information for Phenotypic Traits", in Christopher Hitchcock (ed.)*, Contemporary Debates in Philosophy of Science*. London: Blackwell, 275–289.

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

——— (2002), "Molecular and Developmental Biology", in Peter Machamer and Michael Silberstein (eds.), *The Blackwell Guide to the Philosophy of Science*. Oxford: Blackwell, 252–271.

——— (forthcoming), "The Fearless Vampire Conservator: Philip Kitcher, Genetic Determinism, and the Informational Gene", in Eva M. Neumann-Held and Christoph Rehmann-Sutter (eds.), *Genes in Development*. Durham, NC: Duke University Press.

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

Jacob, François ([1970] 1974), *The Logic of Living Systems—a History of Heredity*. Reprint. Translated by B. E. Spillmann. Originally published as *La logique du vivant; une histoire de l'hérédité* (Paris: Éditions Gallimard). London: Allen Lane.

Judson, H. F. (1993), "Frederick Sanger, Erwin Chargaff, and the Metamorphosis of Specificity", *Gene* 135: 19–23.

Kauffman, Stuart A. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press.

Kay, Lily E. (2000), *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press.

Keller, Evelyn F. (2000), "Decoding the Genetic Program", in Peter Beurton, Raphael Falk, and Hans-Jörg Rheinberger (eds.), *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press, 159–177.

Keyes, Martha E. (1999), "The Prion Challenge to the 'Central Dogma' of Molecular Biology, 1965–1991. Part I: Prelude to Prions", *Studies in the History and Philosophy of Biology and Biomedical Sciences* 30: 1–19.

Kitcher, Philip S. (2001), "Battling the Undead: How (and How Not) to Resist Genetic

Determinism", in Rama S. Singh et al. (eds.), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*. Cambridge: Cambridge University Press, 396–414.

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

Moss, Lenny (2003), *What Genes Can't Do*. Cambridge, MA: MIT Press.

Nelson, David, and Michael Cox (2000), *Lehninger Principles of Biochemistry*. New York: Viking.

Neumann-Held, Eva (2001), "Let's Talk about Genes: The Process Molecular Gene Concept and Its Context", in Susan Oyama, Paul E. Griffiths, and Russell D. Gray (eds.), *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge, MA: MIT Press, 69–84.

——— (forthcoming), "Genes—Causes—Codes: Deciphering DNA's Ontological Privilege", in Eva Neumann-Held and Christoph Rehmann-Sutter (eds.), *Genes in Development*. Durham, NC: Duke University Press.

Noble, Dennis (2002), "Modeling the Heart—from Genes to Cells to the Whole Organ", *Science* 295: 1678–1682.

Polanyi, Michael (1968), "Life's Irreducible Structure", *Science* 160: 1308–1312.

Pontecorvo, Guido (1958), "Self-Reproduction and All That", *Symposium of the Society for Experimental Biology* 12: 1–5.

Sarkar, Sahotra (1996), "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology", in Sahotra Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht: Kluwer Academic Publishers, 187–231.

——— (2003), "Genes Encode Information for Phenotypic Traits", in Christopher Hitchcock (ed.), *Contemporary Debates in Philosophy of Science*. London: Blackwell, 259–274.

Schrödinger, Erwin (1944), *What Is Life?* Cambridge: Cambridge University Press.

Stent, Gunther S. (1981), "Strength and Weakness of the Genetic Approach to the Development of the Nervous System", *Annual Review of Neuroscience* 4: 163–194.

Sterelny, Kim, and Paul E. Griffiths (1999), *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press.

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

Stotz, Karola, and Paul E. Griffiths (forthcoming), "How Biologists Conceptualize Genes: An Empirical Study", *Studies in the History and Philosophy of Biological and Biomedical Sciences*.

Temin, H. M., and S. Mizutani (1970), "RNA-Dependent DNA Polymerase in Virions of Rous Sarcoma Virus", *Nature* 226: 1211–1213.

Thieffry, Denis, and Sahotra Sarkar (1998), "Forty Years under the Central Dogma", *Trends in Biochemical Sciences* 23: 312–316.

Watson, James D. (1965), *Molecular Biology of the Gene*. New York: W. A. Benjamin.

Weber, Marcel (2005), *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

Wheeler, Michael (2003), "Do Genes Code for Traits?", in A. Rojszczak, J. Cachro and G. Kurczewski (eds.), *Philosophical Dimensions of Logic and Science: Selected Contributed Papers from the 11th International Congress of Logic, Methodology, and Philosophy of Science*. Dordrecht: Kluwer, 151–64.

Wolpert, Lewis, et al. (2002), *Principles of Development*. Oxford: Oxford University Press.

Yockey, Hubert P. (1992), *Information Theory and Molecular Biology*. Cambridge: Cambridge University Press.