

Appeared in: *British Journal for the Philosophy of Science* 60 (2009), 1–17.

PENULTIMATE DRAFT

DNA, Inference, and Information

Ulrich E. Stegmann

ABSTRACT

This paper assesses Sarkar's ([2003]) deflationary account of genetic information. On Sarkar's account, genes carry information about proteins because protein synthesis exemplifies what Sarkar calls a 'formal information system'. Furthermore, genes are informationally privileged over non-genetic factors of development because only genes enter into arbitrary relations to their products (in virtue of the alleged arbitrariness of the genetic code). I argue that the deflationary theory does not capture four essential features of the ordinary concept of genetic information: intentionality, exclusiveness, asymmetry, and causal relevance. It is therefore further removed from what is customarily meant by genetic information than Sarkar admits. Moreover, I argue that it is questionable whether the account succeeds in demonstrating that information is theoretically useful in molecular genetics.

- 1 *Introduction*
- 2 *Sarkar's information system*
- 3 *The pre-theoretic features of genetic information*
 - 3.1 *Intentionality*
 - 3.2 *Exclusiveness*
 - 3.3 *Asymmetry*
 - 3.4 *Causal relevance*
- 4 *Theoretical usefulness*
- 5 *Conclusion*

1 Introduction

Genetic information is generally regarded as being ‘intentional’ and exclusive (e.g., Godfrey-Smith [1999]; Sterelny and Griffiths [1999]; Griffiths [2001]). Intentionality in this context is the idea that the information carried by genes is of a sort similar to that carried by things like street maps or beliefs. Maps and beliefs can represent something as being a certain way it is not. Such ‘semantic’ or ‘intentional information’ is often contrasted with natural information, which cannot be false. By exclusiveness I mean the view that the information carried by genes is only carried by genes, or nearly so. Information makes genes special and separates them from non-genetic factors of development. Intentionality and exclusiveness are among the essential ingredients of the customary notion of genetic information. They should be reconstructed by any account purporting to defend this notion as scientifically respectable.

Yet, the two ingredients have proven difficult to reconcile. Teleosemantic theories provide accounts of intentionality based on evolutionary function (Sterelny et al. [1996]; Maynard Smith [2000]). A trait can have a function while failing to satisfy it. By assimilating information to function, teleosemantic theories hope to make sense of the idea that a gene can carry information while failing to properly express it. Though prominent, teleosemantic theories do not deliver exclusiveness: they scatter semantic information generously across both genetic and non-genetic factors of development (Godfrey-Smith [1999]; Griffiths [2001]).¹ One response to this difficulty is to reject the notion of genetic information altogether (e.g., Weber [2005]). Another is to continue the quest for reconciling intentionality with exclusiveness, e.g., by modifying the teleosemantic approach (Shea [2007]) or by developing alternative accounts of intentionality, such as those provided by the developmental role theory (Godfrey-Smith [2000]) and the instructional theory of genetic information (Stegmann [2005]). The third response is to abandon one of these components in order to make genetic information a defensible concept. For instance, by surrendering exclusiveness in favour of intentionality, development becomes a process guided by semantic information that resides in both genes and a great variety of non-genetic factors. This option is in line with the parity thesis (Griffiths [2001]). According to the parity thesis, both genetic and non-genetic factors carry information on any legitimate theory of that notion. So, assuming a defensible notion of semantic information were available, both genetic and non-genetic factors would have it, and this fact would rule out exclusiveness.

The second option is to surrender intentionality in favour of exclusiveness. This move is defended by Sahotra Sarkar ([2003]), who rejects any demand that theories of genetic information need to involve semantic information in order to be adequate². He calls his account ‘deflationary’

for giving up on intentionality. Roughly, the deflationary account aims to capture our general sense of information by defining a ‘formal information system’. This system is then found to be (sometimes) instantiated by the mechanisms of protein synthesis. Hence, genes are said to carry information in a perfectly legitimate sense. Furthermore, Sarkar accepts that many non-genetic factors of development frequently carry the same sort of information, but he maintains that genes are nevertheless informationally privileged because the genetic code is arbitrary, whereas the non-genetic factors do not enter into arbitrary relations. And, according to Sarkar, arbitrariness together with our general notion of information yields ‘semiotic’ information, which is unique to genes.

This paper assesses Sarkar’s defence of genetic information. I first expound the deflationary account, arguing that both ‘specific’ and ‘semiotic’ information are versions of what is known as epistemic information (Fodor [1990]): roughly, *A* carries information for *B* if we can learn from *A* something about *B*. The following sections then explore the extent to which the deflationary account captures the ordinary notion of genetic information. While Sarkar may not intend his account to do justice to that notion, he does assert that it happens to capture a good portion of it. Yet I aim to show that the deflationary account vastly departs from the customary notion, and to a greater extent than Sarkar maintains. The final section assesses the theory by its explicitly stated goal, which is to show that the concept of information is useful in molecular genetics (Sarkar [2003, p. 261]). On this separate score I believe the theory does better, although its usefulness still turns out to be rather modest.

2 Sarkar’s information system

Sarkar’s ‘formal information system’ is defined as consisting of two sets, *A* and *B*, and a relation connecting *A*-elements with *B*-elements. The relation also connects the sets as a consequence of connecting their elements. The relation is called ‘information relation’ and denoted by ‘*i*’. It will become apparent below why Sarkar regards this relation as informational. For the moment it should be stressed that no constraints are placed on the obtaining of this relation between an *A*- and a *B*-element other than that they relate to one another. For example, they do not need to cause one another.

Formal information systems are characterised by two additional features. First, every *B*-element is related to at least one *A*-element (but not necessarily vice versa). Second, each set may contain ‘equivalence classes’, which consist of ‘informationally equivalent’ *A*- or *B*-elements, respectively. Sarkar does not define informational equivalence. Instead he gives as an example of an equivalence class the set of DNA triplets that specify one kind of amino acid (p. 267).³ So if set

A consists of DNA triplets and set B of amino acids, then the informationally equivalent DNA triplets are those that relate to one and the same amino acid. This example suggests that A -elements are informationally equivalent if they bear an information relation to one and the same B -element (cf. footnote 5).

Two conditions guarantee increasing degrees of specificity.

(1) 'Differential specificity: suppose that a and a' belong to different equivalence classes of A . Then, if $\iota(a, b)$ and $\iota(a', b')$ hold, then b and b' must be different elements of B .'

(2) 'Reverse differential specificity: if $\iota(a, b)$ and $\iota(a', b')$ hold, and b and b' are different elements of B , then a and a' belong to different equivalence classes in A .' (p. 267)

According to Sarkar, both conditions hold with respect to protein synthesis in prokaryotes; but only the first holds in eukaryotes (p. 270). In order to understand these claims it is useful to assess the extent to which these conditions are satisfied in certain set theoretic relations. For it then becomes apparent that prokaryotes and eukaryotes differ because they instantiate different set theoretic relations and because only some of these relations satisfy both the specificity conditions.

Four classes of set theoretic relations are important here: one-to-one, many-to-one, one-to-many, and many-to-many relations. I take it that all four relations satisfy differential specificity. This is because differential specificity appears to be a consequence of considering non-equivalent A -elements (although Sarkar does not put it this way). Pick any two A -elements, a and a' , and the requirement that they belong to different equivalence classes ensures that they relate to distinct B -elements. After all, for a and a' to be non-equivalent *is* for them to relate to distinct B -elements, rather than to the same one. This consequence holds in all four classes. Reverse differential specificity, however, holds only in one-to-one, many-to-one, and one-to-many relations⁴. In many-to-many relations the B -elements may be related to informationally equivalent A -elements, in which case reverse differential specificity does not obtain. Suppose a relates to two B -elements, b and b' , and there is a second A -element, a' , which relates to b' . If we now pick the two distinct elements b and b' , then the corresponding A -elements, a and a' , will belong to the same equivalence class because they both specify b' .⁵

Now, the molecular genetic mechanisms of protein synthesis instantiate some of the four set theoretic relations. The DNA triplets in prokaryotes specify amino acids in a many-to-one mode, because several codons may specify one amino acid (degeneracy of the genetic code). Since the relation is many-to-one, both conditions (1) and (2) obtain in prokaryotic protein synthesis. By

contrast, eukaryotes only satisfy condition (1). Their codons and amino acids not only relate many-to-one as in prokaryotes (due to the degeneracy of the code), but the relation is also one-to-many. For example, RNA editing may alter an RNA codon transcribed from a DNA triplet such that one kind of DNA triplet may come to specify several kinds of amino acids. As a consequence, eukaryotic codons relate to amino acids in a many-to-many mode, which precludes reverse differential specificity but satisfies differential specificity.

Sarkar introduces two further conditions, which together are intended to capture the idea that *A*-elements are arbitrarily related to *B*-elements. The first condition is ‘medium independence’ (‘A1’, p. 268), according to which there is ‘no preferred representation’ of an information system. The absence of a preferred representation seems to consist in all implementations being ‘epistemologically on a par with each other’ (p. 268). I understand this condition as follows (for an alternative reading see section 3.2). If there are two physically different yet isomorphic implementations of an information system, then we can learn from the *A*-elements of implementation 1 as much about *B* as we can learn from the *A*-elements of implementation 2. This interpretation fits Sarkar’s example for medium independence, i.e., a physical piece of DNA that is isomorphic to a string of symbols on paper. For we can learn the same things from either implementation, i.e., we can derive the same amino acid sequence. The second condition for arbitrariness is ‘template assignment freedom’ (‘A2’, p. 268). This condition requires that various alternative assignments between *A*- and *B*-elements were evolutionarily possible. Again, this condition is arguably satisfied by the genetic code, because the current assignment of codons to amino acids is one of several evolutionarily possible outcomes.⁶

The point of these various conditions is, of course, that on Sarkar’s view they justify describing the relation between sets *A* and *B* in informational terms. The idea is that *A* carries some degree of information for *B* depending on which of the conditions is satisfied. By definition, if (1) obtains, *A* carries ‘specific information’ for *B*; if both (1) and (2) obtain, then ‘*A* alone carries specific information for *B*’ (p. 267); and if (1), (A1) and (A2) obtain, then *A* contains ‘semiotic information’ for *B* (p. 270) or, in other words, *A* ‘encodes’ *B* (p. 269). As it turns out, genetic and non-genetic factors differ with respect to which conditions they satisfy. The genetic code satisfies the arbitrariness conditions in both prokaryotes and eukaryotes. It also satisfies conditions (1) and (2) in prokaryotes and (1) in eukaryotes. Hence the overall conclusion that ‘DNA encodes proteins; for prokaryotes, DNA alone encodes proteins’ (p. 269). While not excluding the possibility that non-genetic factors of development could in principle satisfy all conditions, Sarkar asserts that, as a matter of fact, they at most carry some degree of specific, but not semiotic, information: ‘So far,

there is no evidence that any of them satisfy (A1) and (A2)' (p. 270). Genes are therefore privileged over all other developmental factors in being the only carriers of semiotic information. The exclusiveness of genetic information is thus restored.

Before discussing various difficulties, one step in Sarkar's argument still needs to be addressed. The argument moves from certain conditions to information, but what licenses this move? All Sarkar says about why his formal information system is an *information* system is this: 'The justification for these conditions [1 and 2] is that they capture what is customarily meant by information in any context: for instance, they capture the sense in which the present positions and momenta of the planets carry information about their future positions' (p. 267). We are not being told more about this customary sense. But presumably the sense in which the present positions and momenta of the planets carry information about their future positions is that knowledge of the present positions allows us to infer, together with accepted theories, the future positions. On this notion of information, *A* carries information for *B* if one can learn about *B* from *A*. Sarkar's notion of information therefore appears to be a species of natural information, i.e., an 'epistemic access theory' of indication (Fodor [1990])⁷. Presumably therefore, Sarkar regards *t* as an informational relation because, given the above conditions, it licenses inferences and, hence, grounds an epistemic sense of information.

It is tempting to think that the notions of semiotic and specific information stand for different sorts of information, not least because semiotic information is explicitly distinguished from semantic information and also from the information of communication theory (pp. 262 and 264, respectively). It may seem as if, on the deflationary account, the privilege of genes is a privilege to an exclusive sort of information. But this is not what is shown. The (intended) difference between genetic and non-genetic factors is that genes are arbitrarily related to some developmental outcomes, whereas non-genetic factors are not. Yet this fact makes no difference to the kind of information carried by genetic and non-genetic factors, respectively. The information is epistemic in either case. Semiotic information is just the epistemic information of arbitrary systems. There is no reason why the epistemic information within the class of arbitrary systems should be somehow different from the epistemic information carried by non-arbitrary systems.

Let us now see whether the deflationary theory captures the notion of genetic information as customarily understood.

3 The pre-theoretic features of genetic information

3.1 Intentionality

The most obvious worry with the deflationary account is the very idea of abandoning intentionality. For, since the customary notion of genetic information appears to be semantic, any adequate theory of this notion needs accounting for intentionality. Sarkar briefly considers this worry, and I take the following quote to be the essence of his remarks:

‘There is no reason to suppose that any concept of *biological* information must be “semantic” in the sense that philosophers use that term. Biological interactions, at this level, are [...] not about meaning, intentionality, and the like; any demand that such notions be explicated in an account of biological information is no more than a signifier for a philosophical agenda inherited from manifestly nonbiological contexts, in particular from the philosophy of language and mind. It only raises spurious problems for the philosophy of biology’. (p. 262)

This remark may appear to beg the question. If biological interactions indeed lacked semantic properties then theoretical accounts could safely ignore them. But whether they do has not been decided one way or another. This issue is a crucial part of what is at stake in the debate.

However, the move to abandon intentionality should not be dismissed too quickly. Perhaps the biologists’ usual understanding does not involve intentionality after all⁸. If it is not a feature of the customary notion, theories aiming to explicate this notion will not need to capture it.⁹ But Sarkar offers no argument to the effect that intentionality is not such a feature. Although biologists like Richard Lewontin ([2001]) endorse a deflationary understanding¹⁰, it is not in line with the way in which genetic information is generally used. For instance, one of the dominant pre-theoretic notions of genetic information, the one featuring in the ‘central dogma’, is ‘hereditary information required for sequentialization’ (Crick [1958, p. 144]), i.e., information required for arranging amino acids (and nucleic acids) into specific sequences during polymerisation. If taken literally, information about how to arrange amino acids can be implemented correctly or incorrectly, just as information about how to bake a cake. And since even the most frequent way of implementing a baking recipe may be incorrect, correctness on the literal reading is not a matter of producing the most frequent result. This in turn motivates the view that customary notions of genetic information involve a strong form of error, which is evaluative and non-statistical.

3.2 Exclusiveness

In contrast to intentionality, Sarkar takes his account as preserving and accounting for exclusiveness, because no non-genetic factor of development satisfies the arbitrariness conditions. Semiotic information is therefore carried only by genes (p. 270). In this section I argue that the

deflationary account fails in so privileging genes. As the arbitrariness conditions are formulated, they are satisfied by a range of non-genetic factors for development and even by biological entities unrelated to development.

First, consider the role of auxin in early angiosperm embryogenesis. Auxin is a ‘plant hormone’ that establishes the apical-basal axis of the angiosperm embryo and is thus a non-genetic and non-environmental factor of development. Angiosperm zygotes divide into the apical cell, which develops into the proembryo, and the basal cell, which develops into the suspensor. Auxin accumulates in the apical cell and its presence is crucial for specifying the apical cell’s fate, i.e. for its becoming the founder cell of the proembryo (Friml et al. [2003]). At this early stage of development, auxin therefore relates one-to-one to cell fate, i.e., auxin relates to becoming the proembryo, whereas its absence relates to becoming the suspensor. Since both differential and reverse differential specificity hold for one-to-one relations, auxin concentration alone carries specific information for cell fate. This conclusion is unproblematic for Sarkar, because he explicitly accepts that non-genetic factors may carry *specific* information for development (p. 270).

However, auxin concentrations also satisfy the arbitrariness conditions. First, the physical auxin-cell fate system is epistemically on a par with its isomorphic representation on a piece of paper insofar as we can infer the same conclusions concerning cell fate from either. The system therefore exhibits medium independence. It also exhibits template assignment freedom.¹¹ For although auxin induces the transcription of auxin-dependent genes in angiosperms today (Jenik and Barton [2005]), it seems evolutionarily possible that auxin could act as a transcriptional repressor (a protein might evolve specificity to auxin, the binding of which might prevent formation of the transcription complex). If so, the auxin-cell fate assignment would be reversed (everything else being equal): high auxin concentrations would repress transcription and therefore specify suspensor-fate, whereas low concentrations would induce transcription and specify proembryo-fate. Hence, auxin concentration alone carries *semiotic* information for cell fate.

The second example concerns environmental factors for development. Many aquatic plants respond flexibly to varying environmental conditions by developing differently shaped leaves (plastic heterophylly). One well-studied species is *Proserpinaca palustris*, which inhabits shallow, seasonally flooded depressions of wetland floors (Wells and Pigliucci [2000]). During winter, the submerged shoots produce filamentous leaves. In spring, vertical shoots rise through the water column and produce increasingly entire (lanceolate) leaves. Submergence is but one of several factors determining whether a given plant individual will develop filamentous or entire leaves. Short daylengths result in filamentous leaves irrespective of whether the shoot apex was submerged

or not. Long daylengths and air exposure generate entire leaves, as do long daylengths in combination with submergence and high temperatures. The relation between environmental factors and leaf shape is therefore many-to-one and the factors alone carry specific information for leaf shape.

The environment-leaf system also exhibits arbitrariness. From physical instances of environmental factors we can learn as much about leaf shape as we can learn from an isomorphic system on a piece of paper (medium independence). Furthermore, given adequate selection pressures it seems possible that the environmental variables could generate the reverse leaf shapes in *P. palustris* (e.g., short daylengths specifying entire instead of filamentous leaves), thereby satisfying template assignment freedom. For instance, a reversal of the environment-leaf assignment should be expected under sufficiently strong selection for low relative surface area in submerged, rather than aerial, leaves (and vice versa). There is no obvious constraint on aquatic plants that their submerged leaves must be filamentous and their aerial leaves entire. Since the environment-leaf system exhibits an arbitrary many-to-one relation, we should conclude on Sarkar's account that the environment alone carries semiotic information for leaf shape.

Finally, semiotic information is found in biological systems unrelated to development. Many animal taxa are vertically stratified in tropical rainforests. Birds in Costa Rican lowland forests provide one example. Whereas ground doves and wrens inhabit the forest floor, and hummingbirds and flycatchers the understory, the canopy top is inhabited by yet other bird taxa, notably toucans and parrots (Bourlière [1983]). Butterflies, too, can be vertically stratified. In a Bornean rainforest, some species inhabit the canopy, whereas others live in the understory (Schulze et al. [2001]). Thus, forest layers (or height intervals) relate one-many to their animal inhabitants, thereby carrying specific information for them. In addition, assignments of forest heights to animals appear to be arbitrary in Sarkar's sense. First, the physical system is epistemically on a par with an isomorphic instantiation of that system on a piece of paper; in either case we can learn from forest height about the inhabitants. Second, the forest layer-animal system exhibits template assignment freedom. True, it is unlikely that, say, highly specialised fruit-feeders like toucans would evolve into ground-dwelling ant-eaters, or wrens into large-winged birds that roam the tree tops. But even such dramatic reversals can hardly be excluded as evolutionarily impossible. So again there is semiotic information. The forest layer-animal system manifests an arbitrary one-to-many relation and we are therefore committed on the deflationary account to saying that forest layers encode their animal inhabitants.

Let me consider two responses to these examples. First, one might object to interpreting the condition of medium independence in terms of epistemic parity; representational symmetry might seem the better interpretation. Take Sarkar's example of a system that does *not* satisfy medium independence: the present, physical states of the planets (p. 268). He says that an isomorphic paper implementation of the planets' states represents the physical planets in a way in which the physical planets do not represent the paper implementation. There is a representational *asymmetry* between the two implementations of this information system. Now, the example nonetheless satisfies epistemic parity; we can learn from either implementation equally well about the planet's future positions. Since epistemic parity occurs in an example that lacks medium independence, it cannot be equivalent to medium independence. So, medium independence is better interpreted as representational symmetry between two isomorphic implementations of an information system.

It is correct that this example fits ill with the epistemic parity interpretation of medium independence. However, if medium independence is taken to mean representational symmetry, we encounter a consequence that appears to be worse: the genetic code turns out to be non-arbitrary on the deflationary account, because it is not representationally symmetric. A paper implementation of the physical relations between codons and amino acids represents those relations in a way in which the physical relations do not represent the paper implementation. It is not clear why a paper implementation should be representational with regards to planets but somehow less so with regards to the genetic code. In order to avoid this difficulty, I interpreted medium independence as epistemic parity.

Here is a second response to these examples. Why not introduce a more demanding notion of arbitrariness and of template assignment freedom in particular? For instance, one might suggest that assignment freedom is a matter of randomness, rather than of evolutionary contingencies. The thought may be this. According to Crick's ([1968]) frozen accident hypothesis of code evolution, the assignments during the incipient stages of code evolution were entirely accidental. The basis of assignment freedom is therefore the randomness of these early assignments, not the possibility that different selection regimes would have generated different assignments, or that a given selection regime could have led to various alternative assignments (cf. footnote 6). Since my examples exhibit assignment freedom only in the sense of evolutionary contingency, they are not genuine counter-examples if assignment freedom is construed as randomness. The difficulty with this move is that the frozen accident hypothesis is poorly supported and probably false (Knight et al. [1999]). So if assignment freedom were to be construed as randomness, not even the genetic code would exhibit assignment freedom and, hence, arbitrariness. While there are other notions of arbitrariness

(Stegmann [2004]) that may be more amenable to Sarkar's purposes, I shall not pursue this issue further here.

It seems fair to conclude that the conditions of arbitrariness, as proposed by Sarkar, endow many non-genetic factors with semiotic information for development. Contrary to his assertion, there is ample evidence that non-genetic factors satisfy his arbitrariness conditions.

3.3 Asymmetry

There is an additional worry about the deflationary account, concerning the alleged direction of genetic information. On the pre-theoretic notion, DNA carries information for proteins (and perhaps for other developmental outcomes), but *not* vice versa. It seems to me that the specificity and arbitrariness conditions do not deliver this asymmetry.

Let set *A* consist of prokaryotic amino acids and set *B* of the corresponding DNA triplets. Nothing about the formal information system blocks this exchange of elements between *A* and *B*. The amino acids are then related one-to-many to their DNA triplets and the system therefore exhibits differential specificity: pick any two amino acids belonging to different equivalence classes and the corresponding triplets will be distinct. Proteins therefore contain 'specific information' for DNA. What is more, proteins *encode* their specific information for DNA. This is because the prokaryotic genetic code satisfies both conditions for arbitrariness. First, the evolutionary possibility of reassigning DNA triplets to different amino acids (which Sarkar accepts) implies the reverse, i.e. that amino acids could be reassigned to different triplets (template assignment freedom). Second, we can learn from physical amino acid sequences as much as we can learn from isomorphic symbols on paper (medium independence). Hence, prokaryotic proteins encode specific information for DNA just as DNA encodes specific information for proteins.

Three responses might be offered. First, one might deny that asymmetry fails. If genetic information is identified with the epistemic information that is carried by genes, as opposed to the epistemic information carried by proteins, then genetic information is asymmetric after all. For the epistemic information gained from nucleic acids has its source in nucleic acids and is about, say, proteins, whereas the information in the opposite direction (from proteins to nucleic acids) is, by definition, not genetic information. This response is beside the point, however, because the worry is that proteins carry the *same kind* of information about DNA that DNA carries about proteins, i.e., semiotic information. Hence, DNA is not informationally privileged over proteins in the sense that it carries information of a sort which proteins do not.

Second, one might place restrictions on the information relation such that the desired asymmetry is restored. For example, *A*-elements may be required to be the causes of *B*-elements or to causally contribute to their obtaining. Since in protein synthesis DNA causally contributes to producing polypeptides, but not the other way around, DNA carries information about polypeptides, but not vice versa. But this suggestion is unsatisfactory. Apart from being ad hoc, informational relations do not in general depend on the carrier causing whatever the information is about. As testified by the famous trails of quails (Dretske [1988]), it is often the effects that carry information about their causes. Also, we can in fact gain some knowledge about prokaryotic DNA templates from the polypeptides they determine. Using the genetic code we can infer for any amino acid either a specific codon or a set of codons. As far as epistemic information is concerned, we do not want to exclude proteins from having it. A dilemma ensues: the information relation either remains unrestricted, in which case genetic information is symmetric, or it is restricted, in which case proteins are denied the sort of information they uncontroversially carry.

Finally, one may object that when the *A*-elements are amino acids, then there are more *B*-elements per *A*-element as compared to the case when the *A*-elements are DNA triplets. That is, the one-to-many component is stronger in the first instance (this is because it is due to degeneracy rather than RNA-editing, and the former generates relatively more *B*-elements per *A*-element). Consequently, proteins carry less semiotic information for DNA than DNA carries for proteins, and this may be seen as a reason to drop information talk with respect to proteins altogether. As Sarkar observes, heterogeneity can compromise the utility of informational interpretations (p. 270). However, heterogeneity and utility are matters of degree. Even if proteins carry very little information, it still is semiotic information.

Ultimately, the deflationary account cannot block the flow of information from proteins back to DNA. It is likely that this is an unwanted result by the theory's own lights. As Sarkar observes repeatedly, on his account DNA carries semiotic information for proteins. And while he does not explicitly say that his account excludes the reverse, this appears to be the understanding.

3.4 Causal relevance

Finally, it is difficult to see how epistemic information plays the alleged causal role within organisms. On the customary understanding, genetic information is causally effective; it helps to bring about certain developmental outcomes. For instance, it can guide protein synthesis by determining the linear arrangement of amino acids. But given that our inferences are not even components of biological systems, how could they play the relevant causal role?

One might be tempted to restore causal relevance for the deflationary account as follows. For *A* to carry information for *B* is for *A* to stand in a relation to *B* that allows us to draw from *A* inferences about *B*. What is constitutive of epistemic information is not our ability to make inferences, but rather a certain relation between *A* and *B*. Whether this relation obtains is an empirical matter and independent of drawing inferences. Furthermore, these relations are exemplified by causally relevant entities in the case of protein synthesis. The many-to-one relation between prokaryotic codons and amino acids is instantiated by a set of causal processes, which include stereochemical interactions, and so on. So, since the informational relations are realised by causal processes, the former may appear to be causally relevant.

But in the case of prokaryotic codons, the many-to-one relation is still distinct from the causal relation between codons and amino acids. On Sarkar's account, the information relation is not identified with a causal relation, though it may be exemplified by items that also exemplify causal relations. There is therefore no sense in which codons causally contribute to producing amino acid sequences in virtue of the fact that they instantiate an informational relation. Consider, by way of contrast, a version of the teleosemantic theory (Sterelny et al. [1996]). It has the potential to make information causally relevant because carrying information for, say, peptide sequences is identified with having the function to cause peptide sequences. To the extent it is reasonable to say that codons specify peptide sequences in virtue of the codons' function to cause certain amino acids to be added to the peptide chain, it is reasonable to say that they specify amino acid sequences in virtue of carrying information for them.

4 Theoretical usefulness

The lesson from the previous sections is that Sarkar's proposal fails to capture four essential components of the pre-theoretic notion of genetic information, i.e., intentionality, exclusiveness, asymmetry, and causal relevance. On this score, then, it is inadequate. But this entails nothing with respect to the proposal's value as a substantive theory of genetic information. By a substantive theory¹² I mean one which asserts that genes carry something appropriately called 'information', and which provides an account of this information. I take it that Sarkar wants to provide a substantive theory in this sense, because he writes: 'Either informational talk should be abandoned altogether or an attempt must be made to provide a formal explication of "information" that shows that it can be used consistently in this context and, moreover, is useful.' And he goes on to say that he provides 'a sketch of one such attempted explication' (p. 261). Such a theory need not capture the customary sense because this sense may overlap little with any defensible kind of information.

Perhaps, the pre-theoretic idea that information is intentional is untenable; indeed, it may turn out that none of the four features can be salvaged from the customary notion.

Let us see then whether the deflationary account fares better as a substantive theory. Does it show that attributing information to genes is useful? Sarkar does not qualify usefulness or explain explicitly how his account is useful. But we can get a grip on this issue by construing usefulness as theoretical usefulness and, in particular, as predictive and explanatory power. This construal is reasonable for a theory and it is the one Sarkar adopted in earlier papers ([1996], [2000]).

First consider predictive power. The many-to-one relation between prokaryotic DNA and amino acids enables us to predict amino acid sequences from DNA sequences. In eukaryotes, the complexities of post-transcriptional and post-translational modifications result in the many-to-many specifying properties of the DNA template, and this strongly restricts the extent to which predictions are feasible. So the informational relation, ι , provides some predictive power, although it is effectively restricted to prokaryotes.

A similar picture emerges with respect to explanatory power. The sequence of prokaryotic amino acids is causally determined, and explained, by the DNA template sequence in conjunction with the many-to-one specifying properties of prokaryotic codons. Since the many-to-one relation is an instance of Sarkar's informational relation, ι , this relation is explanatory for prokaryotic protein synthesis. In eukaryotes, however, this relation is many-to-many. This very fact limits the extent to which ι causally determines (together with a template) a given amino acid sequence. In eukaryotes, the informational relation therefore only plays a limited role in explaining amino acid sequences.

Thus, the informational relation ι is to some extent predictive and explanatory. But it is questionable whether this translates into predictive and explanatory power for the concept of information. One reason is that the modest extent to which the relation is predictive and explanatory may be regarded as being insufficient for licensing information talk. This was Sarkar's earlier position ([1996]): he argued that the customary notion of genetic information as (DNA) sequence is of limited predictive and explanatory usefulness, for the reasons outlined above, and should therefore be abandoned¹³. Yet his new account is explanatory and predictive to the same extent, and for the same reasons.

More importantly, on an epistemic access theory of information, our ability to draw inferences from A about B is essential for what it is for A to carry information about B . Without anyone having this ability, A would not be informational. Carrying information does not reduce to whatever enables or justifies our inferences, i.e., ι according to the deflationary account. The explanatory power of ι and that of A 's carrying information about B may therefore come apart;

epistemic information may do no explanatory work at all even though ι , one of its components, does. Indeed, this seems to be the case with respect to explaining amino acid sequences. The explanation for why such-and-such amino acid sequence arose is not that we were able to infer it from its template sequence (epistemic information). It is rather that there was such-and-such a template sequence with many-to-one specifying properties. In other words, ι 's explanatory value does not render explanatory the attribution of (epistemic) information.

5 Conclusion

The deflationary account promises to show that the concept of information has a legitimate theoretical role to play in molecular genetics. I doubt that the account redeems its promise. It turns genetic information into a species of epistemic information and it secures only a limited degree of usefulness for the notion of information, if any at all. Moreover, the notion of genetic information we end up with is far removed from ordinary usage. In deflating genetic information we lose sight of what appear to be four essential features of the pre-theoretic notion: intentionality, exclusiveness, asymmetry, and causal relevance. This in itself would be unproblematic if indeed they could not be salvaged. But just which features can be salvaged from the ordinary notion of genetic information has not been settled.

Funding

British Society for the Philosophy of Science (doctoral fellowship); British Academy (postdoctoral fellowship).

Acknowledgements

My thanks to David Papineau, Sahotra Sarkar and two anonymous referees for comments on earlier versions of this paper.

*Department of History and Philosophy of Science
University of Cambridge
Free School Lane
Cambridge CB2 3RH
U.K.
us227@cam.ac.uk*

References

- Bourlière, F. [1983]: 'Animal species diversity in tropical forests', in F. B. Golley (*ed.*), 1983, *Tropical rain forest ecosystems, structure and function*, Amsterdam: Elsevier, pp. 77-92.
- Crick, F. H. [1958]: 'On Protein Synthesis', *Symposia of the Society for Experimental Biology*, **12**, pp. 138-67.
- Crick, F. H. [1968]: 'The Origin of the Genetic Code', *Journal of Molecular Biology*, **38**, pp. 367-79.
- Dretske, F. [1988]: *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.
- Fodor, J. [1990]: 'Semantics, Wisconsin Style', in J. Fodor (*ed.*), 1990, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, pp. 31-49.
- Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., Offringa, R. and Jürgens, G. [2003]: 'Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*', *Nature*, **426**, pp. 147-53.
- Godfrey-Smith, P. [1999]: 'Genes and Codes: Lessons from the Philosophy of Mind?' in V. G. Hardcastle (*ed.*), 1999, *Where Biology Meets Psychology: Philosophical Essays*, Cambridge, MA: MIT Press, pp. 305-31.
- Godfrey-Smith, P. [2000]: 'On the Theoretical Role of "Genetic Coding"', *Philosophy of Science*, **67**, pp. 26-44.
- Griffiths, P. E. [2001]: 'Genetic Information: A Metaphor in Search of a Theory', *Philosophy of Science*, **68**, pp. 394-412.
- Jenik, P. D. and Barton, M. K. [2005]: 'Surge and destroy: the role of auxin in plant embryogenesis', *Development*, **132** (16), pp. 3577-85.

- Knight, R. D., Freeland, S. J. and Landweber, L. F. [1999]: 'Selection, history and chemistry: the three faces of the genetic code', *Trends in Biochemical Sciences*, **24**, pp. 241-7.
- Lewontin, R. C. [2001]: 'In the beginning was the word', *Science*, **291**, pp. 1263-4.
- Maynard Smith, J. [2000]: 'The Concept of Information in Biology', *Philosophy of Science*, **67**, pp. 177-94.
- Sarkar, S. [1996]: 'Biological Information: A Skeptical Look at some Central Dogmas of Molecular Biology', in S. Sarkar (ed.), 1996, *The Philosophy and History of Molecular Biology: New Perspectives*, Dordrecht: Kluwer, pp. 187-231.
- Sarkar, S. [2000]: 'Information in Genetics and Developmental Biology: Comments on Maynard Smith', *Philosophy of Science*, **67**, pp. 208-13.
- Sarkar, S. [2003]: 'Genes Encode Information for Phenotypic Traits', in C. Hitchcock (ed.), 2003, *Contemporary Debates in Philosophy of Science*, London: Blackwell, pp. 259-72.
- Sarkar, S. [2005]: *Molecular Models of Life: Philosophical Papers on Molecular Biology*, Cambridge, MA: MIT Press.
- Schulze, C. H., Linsenmair, K. E. and Fiedler, K. [2001]: 'Understorey versus canopy: patterns of vertical stratification and diversity among Lepidoptera in a Bornean rain forest', *Plant Ecology*, **153**, pp. 133-52.
- Shea, N. [2007]: 'Representation in the Genome and in Other Inheritance Systems', *Biology and Philosophy*, **22**, pp. 313-31.
- Stegmann, U. E. [2004]: 'The arbitrariness of the genetic code', *Biology and Philosophy*, **19**, pp. 205-22.
- Stegmann, U. E. [2005]: 'Genetic information as instructional content', *Philosophy of Science*, **72**, pp. 425-43.

Sterelny, K. and Griffiths, P. E. [1999]: *Sex and Death: An Introduction to Philosophy of Biology*, Chicago: University of Chicago Press.

Sterelny, K., Smith, K. and Dickison, M. [1996]: 'The Extended Replicator', *Biology and Philosophy*, **11**, pp. 377-403.

Weber, M. [2005]: *Philosophy of Experimental Biology*, Cambridge: Cambridge University Press.

Wells, C. L. and Pigliucci, M. [2000]: 'Adaptive phenotypic plasticity: the case of heterophylly in aquatic plants', *Perspectives in Plant Ecology, Evolution and Systematics*, **3** (1), pp. 1-18.

Wheeler, M. [2003]: 'Do genes code for traits?' in A. Rojszczak, J. Cachro and G. Kurczewski (eds), 2003, *Philosophical Dimensions of Logic and Science: Selected Contributed Papers from the 11th International Congress of Logic, Methodology, and Philosophy of Science*, Dordrecht: Kluwer, pp. 151-64.

Footnotes

¹ As Wheeler ([2003]) pointed out, teleosemantic theories also suffer from the opposite difficulty in being too restrictive; they exclude genes that have not been selected for, such as hitchhiking genes.

² The account is foreshadowed in Sarkar ([2000]). Furthermore, Sarkar ([2005]) supplements Sarkar ([2003]) with a section on the methodological challenges for demonstrating that genetic codes other than the standard code were evolutionarily possible.

³ Throughout, page numbers refer to Sarkar ([2003]) unless noted otherwise.

⁴ Whether reverse differential specificity holds in one-to-many relations depends on whether a and a' ought to be distinct. If a and a' must be distinct, then any two distinct B -elements will relate to informationally non-equivalent A -elements and, therefore, reverse differential specificity holds. On the other hand, if they need not be distinct, then the two B -elements may be related to the same A -element. Consequently, a and a' would not belong to different equivalence classes and reverse differential specificity would therefore not apply.

⁵ One referee suggested that this example would work only if a' also related to b (in addition to b'), such that both a and a' related to exactly the same B -elements, on the grounds that the latter is required for a and a' to belong to the same equivalence class. However, the sort of equivalence Sarkar needs in order to exclude reverse differential specificity in eukaryotes is not available on this strong version of equivalence, though it is on the weak one adopted here (if two A -elements relate to at least one common B -element, then they belong to the same equivalence class). Consider again the original example of a many-to-many relation (a relates to both b and b' , but a' relates only to b'). On the strong version, a and a' are non-equivalent and the corresponding b and b' are distinct, which satisfies differential specificity. Furthermore, the distinct b and b' pick out the non-equivalent a and a' , thus satisfying reverse differential specificity. The strong version of equivalence therefore restores reverse differential specificity even in some many-to-many cases. Of course, the weak version has the counter-intuitive consequence that two A -elements sharing one B -element are informationally equivalent even if they relate to otherwise vastly different corresponding sets of B -elements.

⁶ Sarkar ([2003]) argues that the evolutionary possibility of alternative assignments is supported by the frozen accident hypothesis (Crick [1968]). Sarkar ([2005, p.275]) elaborates on this point as follows. There may be several local optima for energy efficient codon-amino acid assignments, one

of which is occupied by the standard genetic code. The one occupied was the optimum accessed first in evolution. But this fact is accidental; a different optimum might have been accessed first. Note that this sense of accidental, the possibility of alternative ‘solutions’ to a given selection regime, is not the sense of accidental used in Crick’s hypothesis. For Crick, the genetic code is accidental because the initial stages of code evolution, which lead to the first assignments, were not driven by selection at all (cf. section 3.2).

⁷ In Fodor’s words, ‘*R* represents *S* if you can find out about *S* from *R*’ (p. 34). This is the epistemic, as opposed to the causal, theory of indication. Fodor uses the term ‘representation’ instead of ‘indication’.

⁸ This is indeed Sarkar’s claim (pers. comm.).

⁹ It should be noted that a semantic theory of genetic information could be appropriate even if the pre-theoretic notion was not semantic. A semantic theory might be needed to account for the biological facts or it might be shown to be explanatory. But then it would not be an explication of the pre-theoretic notion.

¹⁰ Lewontin ([2001]) writes in his review of Lily Kay’s 2000 monograph *Who Wrote the Book of Life?*: ‘[To] say that DNA contains determinative information about amino acid sequences is simply to say that a knowledge of the DNA sequence is sufficient to provide knowledge of the amino acid sequence but not vice versa’; he suggests, moreover, that any other understanding of information would be metaphoric and misleading.

¹¹ Of course, auxin concentrations are not templates for cell fate. But Sarkar’s condition of template assignment freedom only requires that two sets of entities, which need not include templates, can be evolutionarily reassigned to one another.

¹² Not to be confused with Sarkar’s ([2000]) ‘substantive’ role of theories or concepts, which they have when they ‘explicitly occur[...]’ in a scientific entity.

¹³ Sarkar emphasises that in his 1996 he did not advocate an outright, but rather a conditional, rejection of genetic information: it is illegitimate if we cannot provide a technical explication different from the usual one in terms of coding (pers. comm.; cf. 2003, p. 261).