

Representation in the Genome and in Other Inheritance Systems

NICHOLAS SHEA

Faculty of Philosophy, University of Oxford, Oxford, OX1 4JJ, UK

(email: nicholas -dot- shea <a> philosophy -dot- ox -dot- ac -dot- uk)

Unedited final draft

Published version forthcoming in *Biology & Philosophy*

DOI : 10.1007/s10539-006-9046-6

Key words: genetic information, genetic representation, inheritance systems, ontogenetic and phylogenetic explanation, teleosemantics

Abstract. There is ongoing controversy as to whether the genome is representing system (Sterelny et al. 1996, Griffiths 2001). Although it is widely recognised that DNA carries information, both correlating with and coding for various outcomes, neither of these implies that the genome has semantic properties like correctness or satisfaction conditions (Godfrey-Smith 2002). Here a modified version of teleosemantics is applied to the genome to show that it does indeed have semantic properties – there is representation in the genome. The account differs in three respects from previous attempts to apply teleosemantics to genes. It emphasises the role of the consumer of representations (in addition to their mode of production). It rejects the standard assumption that genetic representation can be used to explain the course of an organism's development. And it identifies the explanatory role played by representational properties of the genome. A striking consequence of this account is that other inheritance systems could also be representational. Thus, a version of the parity thesis is accepted (Griffiths 2001). However, the criteria for being an inheritance system are demanding, so semantic properties are not ubiquitous.

Contents

- (1) Introduction: Is genetic information semantic?
- (2) The explanandum - adaptive complexity
- (3) Selectional vs. developmental properties of genes
- (4) Representation in the genome, and in other inheritance systems
- (5) Conclusion

(1) Introduction: Is genetic information semantic?

Understanding the genome has helped to demystify the miracle of life. As if the mere existence of living organisms were not enough, their staggering complexity is created anew in each generation starting from a single-celled embryo. Think of a peacock egg. Keep it in a warm place, next to a big pile of corn, and it will transform itself into the shimmering beauty of an adult male. How so? We now know a vital ingredient in the magic potion: DNA. Its stable double helix is thought to carry the information needed to sustain such adaptive complexity over the aeons. But what is this 'information' that DNA is said to carry? It's certainly unlike the information we read in the newspapers. In everyday currency, items of information are produced and interpreted with conscious understanding. According to one view, talk of genetic information is entirely different. It is just convenient shorthand for the special kind of causal process by which proteins are made from DNA. There is a striking rule of causal specificity between the four-base alphabet of the DNA molecule and the sequence of amino acids in the proteins produced from it. Who can resist calling this a code, when the rule is so like the spy's procedure for transforming one string of symbols into another? But the coded signals conveyed behind enemy lines are more than just strings of letters. They also carry messages. They have semantic properties: reporting facts and giving orders. Is the same true of DNA? Does it represent the world, making claims that could be false? Does it carry instructions, issuing directives that could go unheeded?

The rule of causal specificity between the linear order of bases on a strand of DNA and the primary structure of proteins produced from it can be called a code irrespective of whether it carries a message or instruction (Godfrey-Smith 2000b, 2002, Sarkar 2004, Stegmann 2005, Wheeler 2003, forthcoming). The existence of this coding relation does not imply that DNA has any semantic properties. It does not licence attribution to genes of correctness conditions or satisfaction conditions. Talk of coding is surely one reason why genes are called information carriers, and from it flows talk of transcription, translation, proofreading, etc. But this is not information in any semantic sense. Nor is it plausible that, in this sense, genes code for phenotypes, since the code only extends as far as the sequence of amino acids in the gene's first protein product. There is then a many-many relationship between expressed proteins and phenotypes. The coding relation also supports a notion of error. The teleofunction of the DNA triplet TGG is to give rise to the amino acid tryptophan – throughout evolutionary history the triplet TGG has contributed systematically to the reproduction of the organisms in which it is found only when it has given rise to tryptophan. Producing something other than the standard amino acid is an error, a malfunction. Wherever there are teleofunctions there is the possibility of error. But not all teleofunctions are representational. So the possibility of error is not diagnostic of representation.

DNA also carries correlational information (Dretske 1981). Genes correlate with the expression of their protein products. Some genes also correlate with phenotypic features. The best examples of the latter are genes for genetic diseases. But there are relatively few strong correlations between single genes and readily-identifiable phenotypes. Genetic diseases that depend upon a single locus mutation are a special case (and are studied extensively for that reason). Grice called such correlations 'natural information' (Grice 1957). It is information in the sense that its presence allows us to work things out. A significant portion of biomedical research is engaged in the task of inferring from the clues it provides. For example, geneticists can construct phylogenetic trees of the relatedness of people alive today by comparing non-coding sections of their DNA. This is why biology is being called an information science – because correlations which can be relied upon as evidence are a kind of information, and it requires large amounts of computing power (as well as ingenuity) to extract the inferences that DNA affords. However, correlational information is not representational content. Such correlations do not have conditions of correctness or satisfaction. So although the fact that DNA is rich in correlational information may explain a lot of informational talk in biology, it does not vindicate attributions to genes of representational content (so-called semantic information).

I reserve 'representational' for semantic properties like correctness and satisfaction conditions. The coding relation is not of the right kind to be representational. Nor are the current correlations between genes and their protein products (and other phenotypic properties). They are both part of a causal account of the processes of development which explains how DNA, non-genetic cellular machinery and environmental factors interact to give rise to the phenotype of a mature organism. Within such an account, there seems to be no room for a separate explanatory role for correctness conditions or satisfaction conditions. As with representational explanations in other domains, representational properties are not found in an explanation of the internal causal process. A representation's content – the message or instruction it carries – is a property of the representational token (the sentence, signal or pattern of neural firing), but non-semantic properties of the token threaten to monopolize the causal and explanatory responsibilities. With mental representation, the concern is that local properties of the brain, or syntactic properties of the processes it instantiates, do all the causal work (Field 1978). The same concern arises wherever representational properties are putatively located: what causal or explanatory work do they do? Despite their variety, many different theories of content can offer the same kind of answer. Local properties of an organism furnish a causal story about what goes on within the skin: about internal reactions to sensory input (described as surface stimulation) and the amazingly complex internal processes eventuate in bodily movements. But we also want to characterise what organisms do in their particular environments: the results of their behaviour and the distal states of affairs which lead to such behaviour. To predict and explain such relational properties of the organism we should expect to invoke relational properties of its internal states. Such relational

properties may capture real patterns in organism-environment interactions that are invisible if organism and environment are considered separately. Representational contents are relational properties suited to particular kinds of explanation.

Different kinds of representational contents are found in different systems. To identify representation in the genome I will use the 'infotel semantics' that applies to many simple representing systems, which I have developed elsewhere.¹ Infotel semantics combines a teleosemantic output condition with an informational input condition. It gets its purchase when a producer system and a consumer system interact via a range of intermediates, with the consumer system responding systematically to each different intermediate. If each resulting action of the consumer system has a specific evolutionary success condition (under which the response historically contributed to the survival and reproduction of the organism), then the intermediates will have content provided they are produced so as to covary with that success condition. The correlation need not be exceptionless (as required by Dretske 1981). It may only extend through some local region of space or time (Millikan 2000, app. B; Millikan 2004; Maclaurin 2002). And the probabilistic relation need only be sufficiently strong to have been worth acting on in the organism's evolutionary history (Godfrey-Smith 1991). Infotel semantics is applicable when there is an evolutionary success condition specific to each of the range of intermediates, and each intermediate carries the correlational information that its success condition obtains. Those conditions are then the contents of the various intermediate representations.

If genes do have representational contents then adverting to those contents should supply explanatory power which is unavailable when considering only correlation, coding and other non-semantic properties of genes. Section (2) locates an explanandum and identifies two types of explanation of it, in one of which representational properties of genes have an important explanatory role. These are ontogenetic and phylogenetic explanations, with genetic representation figuring only in the latter. Section (3) draws a corresponding distinction between developmental and selectional properties of genes. Section (4) uses these distinctions to support an argument that there is indeed representation in the genome.

(2) The explanandum – adaptive complexity

Life has an astounding ability to reform anew in each generation. Nothing else produces such structured, adaptive complexity. Leave a stone to interact with as rich and complex an environment as you like and it doesn't get much more complex. In fact, pretty much nothing does. Except living things. So striking are they in this respect that, until recently, people were convinced that they must possess a mystical life force, the *élan vital*. In the

modern understanding of life, if anything at all plays the role of *élan vital*, it is surely DNA. I will argue that DNA can indeed help explain adaptive complexity. In this section I differentiate two ways in which that explanandum may be addressed. To be more precise about the explanandum, the objects of explanation are organisms and their patterns of interaction with their environment. The explanatory question is:

(DAC) *How does adaptive complexity arise anew in each generation?*

I will argue that there are at least two ways of answering DAC ('development of adaptive complexity'), ontogenetic and phylogenetic, with representational properties of genes playing a role only in the latter type of explanation. The explanandum concerns the outcome of development – how did that peacock's adaptive complexity arise from that egg – and not just the existence of adaptive complexity in general. So 'ontogenetic' and 'phylogenetic' are being used in a special sense, as labels for two different ways of explaining the outcome of an organism's development. To preview the distinction, the ontogenetic explanation gives a stage-by-stage description of the causal processes leading from embryo to adult, whereas the phylogenetic explanation brackets these details and explains the outcome of development in terms of the instructions given to the developmental process. For a rough comparison, consider ordering a coffee. The former kind of explanation would chart the detailed process starting with pressure waves issuing from my lips, causing bodily movements of the coffee shop worker involving beans, milk and water, and issuing in a hot frothy-topped liquid in a paper cup in my hand. The latter approach would explain the same outcome in terms of my having *ordered a cappuccino*.

Preformationism, if true, would be an ontogenetic explanation of DAC. So would genetic determinism. But genetic determinism is false, and it seems unlikely that genes make a special causal contribution to the processes of individual development (Griffiths 2005). Where genes do play a special role is in a phylogenetic explanation of DAC. To see this, recall why DNA has been such a significant discovery. To start with, it is a very stable molecule with an incredibly numerous range of differentiable states. Each base-position can, independently, be filled by one of four types (A, C, G and T), giving rise to a combinatorial explosion of distinct possibilities. Each possible state has the potential to carry correlational information. The capacity of a system to carry different pieces of correlational information is often measured in bits. I will call this a system's information capacity. A system has a large information capacity when it can adopt a large range of different stable states which can be measured, acted on to produce behaviour, etc. The more readable states the system can adopt, the more bits of information are carried when it is in one of those states. Information capacity and correlational information are the two parts of Shannon's theory of communication (Shannon 1949). They are not equivalent.

¹ Shea, 'Consumers need information: supplementing teleosemantics with an input condition', in submission.

Correlational information considers the probabilistic relation between one possible state of one system and one possible state of another system. By contrast, information capacity is a property of a single system. Two systems come into view in Shannon's account of a communication channel. The quantity of information carried by a channel depends upon all the possible states of the communicating systems. There *is* a communication channel for the huge amount of order locked into the genome. Cellular processes of protein manufacture are sensitive to the different states of the genome, in accordance with the coding relation discussed above. Thus, the very many states of the genome correspond to large numbers of potential primary protein products – two systems are linked via an intermediate channel. However, neither the information capacity of a system nor the capacity of a communication channel depends on any message actually being conveyed. As Shannon recognised, his kind of information has nothing to do with representational content.²

Information capacity is ubiquitous, especially in the computer age. Storing the maximum amount of information in the minimum space is the goal of a whole branch of engineering. Artefacts like DVDs are designed to be very good at it. Each bump on the shiny underside of a DVD can be in one of two states. With lots of bumps there are wildly more possible states of a DVD than there are particles in the universe.³ Like genomes, DVDs give rise to amazingly complex outputs. Embedded in the same environment (my laptop), one produces a mountaineering documentary, another a Bollywood musical. As with genes, there is a complete causal explanation of how the bumps interact with the decoding machinery, reflecting a laser beam which is transduced and processed through various circuits until it issues in flickering lights on the screen and vibrations in membranes in the speakers. It would even be possible, in principle, to *predict* the results on the screen that would issue from a given disc, without actually trying it. But it's much easier to predict and explain the behaviour of a DVD disk if we adopt a different explanatory stance (cp. Dennett 1981). Then we predict that *that* disk will give rise to a documentary about the Slovenian ascent of Trango Tower (you can tell by the hair-raising pictures on the other, non-coding side; and by the title, 'The Slovenian Ascent of Trango Tower'.) This is to predict the output on the basis of what the film-maker meant the disk to do. This alternative explanation still relies on there being a huge variety of possible states of the DVD disc, to go with the huge range of things film-makers might want to show; but it

² 'The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.' (Shannon 1949, p. 3). In his gloss on Shannon's theory, Weaver took a similar view (although he did have more to say about the semantic aspects of communication), '*information* must not be confused with meaning. In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information.' (Weaver 1949, p. 99)

³ DVDs can carry over 10 gigabytes, which is 10^{11} (100 billion) bits of information. Each bit can be in one of two states, giving $2^{100\text{billion}}$ possible states. That is more than $10^{30\text{billion}}$ possible states. This compares to estimates of the number of particles in the universe ranging up to 10^{80} (10^{80} is the number of possible states of a system of 33 bytes).

brackets the actual process by which light inputs at a camera are coded onto a DVD and then decoded back into pictures on a screen. Predicting DVD behaviour from the latter stance relies only on the fact that the whole complex mechanism was designed to reproduce what the camera was aimed at.

So the complexity that results from playing a DVD can be explained in two quite different ways. One explanation charts the stage-by-stage causal processes. The other relies on teleofunctions. The two are complementary explanations of the same token outcome (the output on the computer screen). But they shouldn't be mixed. It is a familiar point that the same distinction applies to living organisms produced by natural selection, since they too have teleofunctions. Why does the heart pump blood? Because it has four chambers, arranged in serial pairs, with one-way valves between them, etc. Or: because that's its evolutionary function. DNA too has both current causal and teleofunctional properties. They feed, respectively, into ontogenetic and phylogenetic explanations of the phenotypes to which DNA gives rise. As with the DVD, both rely on there being a huge information capacity of the genome. But genes' teleofunctions do not figure in an ontogenetic explanation of the actual processes of individual development.

It has often been argued that any information about phenotypes carried by genes cannot form part of an explanation of the course of individual development.⁴ However, no one has noticed why. The reason is that the semantic properties of genes are a species of selectional property, as I will argue in section (4). The next section (3) distinguishes between selectional and developmental properties of genes and shows that only the latter should figure in ontogenetic explanations of DAC. So if we seek to explain the *course* of individual development – the chain of processes by which an embryo becomes an adult – we should not advert to the semantic information in the genome. However, when it comes to the other, phylogenetic, kind of explanation of DAC, the semantic content of the genome plays an important explanatory role. It furnishes one kind of explanation of DAC and so of the outcome, but not the course, of individual development.

(3) Selectional vs. developmental properties of genes

Biologists have worked with several different conceptions of the gene.⁵ Molecular genes are lengths of DNA. A single molecular gene is a stretch of DNA lying between a start codon and a stop codon – a section of DNA transcribed as a single unit. Only rarely is a molecular gene causally associated with a single phenotype. There is a many-many relation between genes and phenotypes: each gene contributes products that act to produce many different phenotypic features, and each phenotype relies causally on the action of many different genes. Development proceeds via a web of interactions between genes, cellular machinery

⁴ Eg, Lehrman 1970, p. 35, quoted with approval in Griffiths and Gray 2005, p. 420.

⁵ Throughout I use 'gene', unqualified, to mean molecular gene.

and the developmental environment, during which genes are switched on and off by other genes and by factors in the environment. A further complication is that a single gene can give rise to several different protein products, depending upon modulations to the translation machinery (which are only now beginning to be understood). All this makes charting the course of individual development a task of byzantine complexity.

There is another gene concept that predates the discovery of DNA and molecular genes. It is the classical gene. The classical gene is a difference-maker for a phenotypic difference; one which obeys the laws of Mendelian inheritance. Classical genes are identified instrumentally, without knowing what those difference makers are. Luckily for Mendel, the traits he studied in highly inbred sweet pea varieties could be associated with classical genes. In such cases there are often genetic differences at a single locus (or a few loci) which correlate with the phenotypic difference. The same is true of genetic diseases that display a Mendelian pattern of inheritance. A molecular gene can then be identified which predicts whether or not a person will have the disease. Moss has introduced the term gene-P (for 'predictive') for molecular genes whose presence correlates with a phenotypic difference, to be contrasted with genes-D (for 'development') (Moss 2001). A gene-P will also be a gene-D, but the converse is rare. So genes-D and genes-P are best seen not as different entities, but as different properties of a given stretch of DNA: as a gene-D a molecular gene is described in terms of its base sequence and expression into proteins; as a gene-P it is described in terms of phenotypic effects with which it correlates (if any).

Because of the complexities of development, not all molecular genes have gene-P properties. Genetic diseases are a rather special case. But the fact that organisms have evolved by natural selection on genes implies that, historically, there was a gene-P property for many evolved phenotypes. When a new molecular gene arises, it is acted on by natural selection if it results in a phenotypic difference which has an effect on fitness.⁶ Gene-P properties were the makers of the heritable differences which drove such episodes of natural selection. Thus, they give rise to teleofunctions. Each molecular gene has many effects (actual and possible). Amongst these, a phenotypic effect for which that gene was selected is one of its evolutionary functions (there may be several). I will call this a 'selectional' property of the molecular gene. Contrast 'developmental' properties of a molecular gene: its disposition to give rise to certain proteins and other downstream products. The two senses of genetic information discussed in section (1) – coding and correlation – are both developmental properties of genes, according to my definition. The coding relation is a relation of causal specificity between base sequences and amino acid sequences. The current correlations between a gene and its downstream products are also

⁶ New genes classically arise through mutation, but there may be other mechanisms, such as by the rearrangement of transposable elements of DNA (Moss 2003). When transposition results in stretches of non-coding DNA giving rise to phenotypic differences that are acted on by selection, it may not be possible to locate a gene-P which correlates with the new phenotypic difference. Instead, the relevant locus of heritable variation will be the difference maker in the DNA, whatever it is.

just causal role functions of the gene. By contrast, selectional properties are teleofunctions of the gene.⁷ A gene's developmental properties will figure in an ontogenetic explanation of DAC. But where the outcome of development is explained in terms of selectional properties of genes, that is what I have called a phylogenetic explanation of DAC. Selectional properties of genes are not suited to figure in an ontogenetic explanation of DAC.

Both classical genes and genes-P focus on phenotypes. They are uninterested in the tortuous process by which those phenotypes arise. That is also a merit of my selectional properties of genes. Natural selection acts on heritable phenotypic differences. It is indifferent to how they are produced. Abstracting away from the stage-by-stage processes of development, selectional properties of genes are custom-made to figure in phylogenetic explanations.⁸ This is especially important given that some developmental outcomes are canalised against environmental or genetic variation (Flatt 2005). Canalisation supplies multiple causal routes to the same phenotype. These pathways should be bracketed together for the purpose of giving phylogenetic explanations.

For F to be a selectional property of a gene G, there must have been a correlation between G and F at the time when G was selected. However, that correlation may be broken by subsequent evolution. A gene's protein products may become incorporated in multiple developmental pathways so that its action (at different times and sites in the developing organism) may be causally necessary for the appearance of very many phenotypes. It may also be re-deployed to make additional protein products, by means of different splicing by the decoding machinery. It's expression may even have become crucial to the viability of the organism. So a gene need not in general still be a gene-P for (ie, predictive of) its selectional properties. And each new episode of selection in which a gene is involved will furnish it with additional selectional properties. However, the scale of interdependence in development remains a matter of experimental investigation (Schlosser and Wagner 2004). To the extent that genes are quasi-independent (Lewontin 1978), a gene may still correlate causally with the products for which it was selected. Similarly, developmental modularity may preserve some of the gene-phenotype correlations which were acted on by natural selection.

Selectional properties of genes arise when genes are difference makers for heritable phenotypic effects (which are then selected for). Other developmental resources may have selectional properties in the same way. Chromatin marks are one plausible candidate.

⁷ For the distinction, see Godfrey-Smith 1993, Griffiths 1993. Griffiths calls them 'causal' and 'selected' functions, respectively ('Function, homology and character individuation, in submission).

⁸ In a sexually-reproducing species, each gene must be effective in the context of most of the other genes in the species' gene pool. Hence, the focus on selectional properties of *genes*. In asexual organisms (or mitochondrial DNA) it is better to consider the whole genome. Selection occurs between two genomes (which differ in terms of the relevant mutation). Thus, the relevant selectional properties will attach, in the first instance, to genotypes (typed by the DNA differences that were operative during episodes of natural selection); and the same for representational properties. For simplicity, I continue to talk only about the selectional properties of genes, but the same points apply to selectional properties of the genomes of asexual species.

These are modifications to the proteins in which DNA is packaged that affect which genes are expressed. Chromatin marks are known to be crucial for generating different cell lines during an individual's development so that, for example, a liver cell divides to form further liver cells ('somatic cell inheritance'), but it is not yet clear to what extent they give rise to differences between phenotypic traits which are heritable across generations of organisms ('trait inheritance') (Jablonka and Lamb 1995, Jablonka and Lamb 2005). The term 'inheritance' is potentially misleading. Only a subset of the resources which offspring inherit (in the everyday sense) from their parents are the basis for trait inheritance. A plant's parents pass down to it a handy location on the surface of the earth, where it can get the sunlight it needs. But this is not the basis of trait inheritance, since differences in this feature do not correlate with heritable differences in phenotype. Not having the feature (a location within the sun's reach) is fatal. Therefore, although an organism's relationship to a background resource may be necessary to the development and evolution of the organism, it does not give rise to selectional properties, since it is not a locus of heritable variation (at least, not when the background resource is necessary and relevant variation is fatal). Similar considerations apply to the cellular machinery that is passed on to the zygote. Structures like cell membranes and centrioles (used in cell division) cannot be synthesised from scratch. The zygote couldn't do without the samples that are passed on from its mother. But it is far from clear that differences in these structures give rise to phenotypic differences that are carried down the generations. One example is given by the chemical gradients within the zygote which are needed for the embryo to differentiate into different cell types according to its body plan (Weber 2005). As studied in *Drosophila*, the zygote gets these morphogen gradients directly from its mother. Furthermore, variations in the morphogen gradient passed on to the zygote are not necessarily fatal. They can lead to a viable phenotypic differences. But descent stops with the first offspring, because the morphogen gradient passed on by the F1 fly to the next generation is produced by gene transcription, and does not depend upon the nature of the gradient that it received as zygote from its mother. By contrast, changes of gradient caused by gene mutation are heritable. They are maternal effect mutants: the mutation first has its effect in the offspring. So morphogen gradients are not the basis for trait inheritance although, in an everyday sense, they are inherited by the offspring.

It is, of course, an open empirical question whether various non-genetic factors which are relevant to development are also the basis of heritable phenotypes. My purposes are equally well-served however these investigations turn out. The important point is that heritability is a substantial requirement.⁹ Very many developmental resources are, like genes, partially causally responsible for parent-offspring similarity;¹⁰ but they only have

⁹ Von Neumann 1966, p. 86 argues that it is a criterion for being a self-reproducing system, in the most general sense.

¹⁰ Mameli 2005 argues that not all *causes* of parent-offspring similarity are needed to *explain* such similarity.

selectional properties, in the way genes do, if they are the basis of heritable phenotypes.¹¹ Only this subset of developmental resources is needed to explain the evolution of an organism by natural selection. The set of selectional properties is also a proper subset of the set of teleofunctions. There is a significant debate about how teleofunctions should best be understood, which I won't attempt to resolve here, but on many views it is not required for having teleofunctions that an item should be a locus of heritable phenotypic variation. So selectional properties meet more stringent criteria than do teleofunctions in general. The heart has the teleofunction to pump blood, but the heart itself is not the reason why the trait of having circulating blood is heritable.

DNA has a further property that many other developmental resources lack. It has the evolutionary function of producing heritable phenotypes. Godfrey-Smith 1999 distinguishes between two kinds of function that DNA has: functions of particular molecular genes (option A) and functions of DNA in general (option B). The selectional properties of genes discussed above arise from teleofunctions under option A. But DNA as a whole also has teleofunctions under option B. It was selected to be a basis of heredity. Option A functions arise over phylogenetic time, during periods of selection. Option B functions are higher-order. They arise because of a series of episodes of selection. As a result, DNA, taken together with its associated developmental machinery, has this meta-function: to produce heritable phenotypes.¹² I will call the kind of developmental resources that have this meta-function inheritance systems.¹³ A developmental resource may be a basis of a heritable phenotype, and so have selectional properties (option A), without having the meta-function of producing heritable phenotypes (option B) and therefore being an inheritance system. For example, it may turn out that cell membrane structure is heritable, and that cell membrane structures are themselves the basis of that heritability; but it is less likely that membranes have the meta-function of producing heritable membrane types. If that is right, then cell membranes would have selectional properties without being an inheritance system. By contrast, if chromatin marking turns out to be a locus of heritable phenotypic variation at all (and not just of somatic cell inheritance and maternal effects), there probably would have been enough variation in the system's phylogenetic history for it also to have acquired the meta-function of producing heritable

¹¹ Griffiths 2001 treats as inheritance 'any biological mechanism which produces resemblances between parents and offspring' (pp. 399-400). Weber 2005 criticises this usage and emphasises that only hereditary replication is evolutionarily relevant (p. 239).

¹² Alberts et al. 2004, pp. 169-191. The meta-function – to produce heritable phenotypes – arises from a long history in which different stretches of DNA gave rise to different phenotypes. Because of the stability of such DNA-phenotype correlations, they could be acted on by natural selection. The cellular machinery of DNA maintenance and repair (Alberts et al. 2004, chs. 5 & 6) is further evidence that DNA has the meta-function to produce heritable phenotypes. Indeed, the postulated transition from the RNA world to DNA-based inheritance is explained by DNA's superior ability to perform this meta-function (Maynard Smith and Szathmáry 1995).

¹³ DNA is part of an inheritance system that is capable of giving rise to a fine-grained response to environmental conditions. It has been suggested that, to count as an inheritance system, a mechanism must be capable of generating fine-grained responses to environmental conditions (Griffiths 2001, p. 406). I'm uncertain, but I don't currently see a strong motivation for this additional restriction, at least for present purposes.

phenotypes, and thus for it to be an inheritance system. Learning by imitation in apes is another mechanism which probably satisfies the conditions on being an inheritance system.

To summarise, I have pointed to three nested categories of the properties which figure in phylogenetic explanations. At the bottom are teleofunctions, which are had by any structure whose effects contribute to its survival and reproduction (eg, the heart has the teleofunction of pumping blood). Selectional properties are a special sub-class of teleofunction, possessed by genes and other developmental factors which are the basis of heritable phenotypes. (Recall that not all causes of parent-offspring similarity are heritable, eg, morphogen gradients.) Finally, most restricted is the category of inheritance systems, which have the meta-function to produce heritable phenotypes. The machinery is now in place to tackle the question of genetic representation.

(4) Representation in the genome, and in other inheritance systems

The infotel semantics for representation in simple systems is applicable where a mechanism divides into three subsystems: a producer, a range of intermediates and a consumer. The output of the consumer system in response to each intermediate must have its own evolutionary success condition, specific to that intermediate, under which the type of output prompted contributed systematically to the survival and reproduction of the system. Representational contents arise when the producer subsystem manages things so that the intermediates correlate with those specific success conditions, the intermediates thereby being representations.

To apply this to the genome, consider a new gene H that arises by mutation. For example, suppose H (for hair) causes organisms to develop thick hair, being a new gene-P which correlates, at that stage in evolutionary history, with thick hair. If thick hair confers a selective advantage, the frequency of H in the population will increase (because the environment has become colder, say). In virtue of this episode of selection, gene H acquires a selectional property: the evolutionary function to produce thick hair. I will argue that gene H thereby also acquires representational properties. It has the following content: *the environment is cold* (conducive to thick hair), *grow thick hair*.¹⁴ To see that this is so, we need to identify, on the output side, an evolutionary success condition specific to gene H and, on the input side, a correlation between gene H and that condition's obtaining.

Evolutionary success conditions are conditions required for the gene to perform its evolutionary function in a systematic way. As well as the very general (that it should promote relative reproductive fitness), gene H has a more focused evolutionary success condition: that the environment is cold (ie, in a temperature range conducive to thick

¹⁴ Like many simple representations, it has both indicative and imperative contents. It is a pushmi-pullyu representation (Millikan 1996). For simplicity, I discuss the input and output conditions only for the

hair). Recall that infotel semantics requires each intermediate representation to have its own success condition, specific to it in the context of a range of other intermediates. Is that true of gene H? Yes, because DNA is part of an inheritance system which, as a whole, has the function of producing heritable phenotypes (Godfrey-Smith's option B above). This ensures that there are a variety of other genes with evolutionary success conditions to which we can compare the new gene H to see that, in our example, H does indeed have a specific evolutionary success condition: *that the environment is cold*. So gene H satisfies the output-based criterion on content.

It also satisfies the input-based criterion. Gene H carries the correlational information that the environment is cold. This is harder to see, because when we consider only the lifetime of an individual, we find no relevant correlation. The mutation just arose by chance. But now consider the species' gene pool after there has been selection for gene H. From the presence of gene H in that gene pool we could infer that the environment had been cold (or otherwise conducive to gene H's phenotypic effects). That is, presence of gene H in the population correlates with the environment having been cold. From the perspective of evolutionary time, H correlates with cold. Because of natural selection, over this timescale the producer system gave rise to more instances of gene H because it was cold and would have stopped producing gene H had the environment not been cold. Gene H thereby meets our criteria for being a simple representation.

The genes' representational properties arise because we can identify, in the mechanism of genetic inheritance, three component systems: producers, consumers and intermediate representations. The DNA in the zygote is produced by the mechanisms of reproduction and acts as intermediate. It is consumed by a developmental system – cellular machinery and environmental factors – to give rise to a new organism, with all its phenotypic traits. To discern that there is a range of potential intermediates between producer and consumer we have to shift to the perspective of evolutionary time. Over the phylogenetic history of a species there has indeed been a range of different intermediates passed between generations, leading a variety of phenotypes to be output by the consumer mechanism. Godfrey-Smith 2006 doubts whether it is legitimate to lump together all the diverse mechanisms of development and treat them as a single consumer system, especially since the DNA of the zygote is reproduced in multiple copies that are active in individual cells during development. However, evolution too treats the developmental system as a whole, acting on its outcomes irrespective of how they are produced. An analogy might help diffuse the objection. Consider a composer who writes a score for a chamber group. She listens to the piece performed and then repeatedly tweaks the score and listens again until she is happy with the final output. The chamber group is a single consumer system of her score. This is so despite the fact that, in practice, one of the group makes photocopies of the score so that each instrument plays from its own copy. Similarly, it is the zygotic

indicative content.

DNA which has semantic properties, acting as a representational intermediate between generations. Those representations are consumed by the whole developmental system to issue in the phenotypes on which natural selection acts.

Notice how the timescale has shifted. In standard examples of simple representing systems, like the honeybees' nectar dance, a range of intermediates is produced during an individual lifetime (order 0). That system has an evolutionary history (order 1) in virtue of which the intermediate states have content. But with DNA we need to move to evolutionary time just to discern the range of intermediates. We also need to move to evolutionary time to discern the correlational information needed to satisfy the input condition. When the range of messages themselves are sent over evolutionary time (order 1), their content depends on *that* mechanism having an evolutionary function (order 2). To put it another way, to be a consumer system a mechanism must have the function of responding to a range of intermediates with a variety of different outputs. The genetic inheritance system does indeed have that function, but it is a function of DNA and its consumers as a whole (Godfrey-Smith's option B above), not of particular genes (option A). It is the function discussed above: to produce heritable phenotypes. And it is a meta-function (order 2) which arises as a result of a series of episodes of selection of heritable phenotypes (order 1). In short, the developmental system which reacts to zygotic DNA by outputting an organism's phenotype is a genuine consumer system, precisely because it is part of an inheritance system, according to the criteria discussed in the last section. Recall that not every cause of a heritable phenotype is part of a *bona fide* inheritance system.

Both Godfrey-Smith and Papineau argue that genes cannot have indicative contents.¹⁵ Papineau in particular is worried that, although genes initiate causal sequences with biologically purposeful outcomes, there is 'nothing in this akin to the gearing of these causal sequences to variable environmental circumstances'. But neither Papineau nor Godfrey-Smith considers the relevant timescale (ie, evolutionary time) over which selected genes do correlate with environmental conditions. To display the indicative content of a gene, suppose the environment changes so as not to be conducive to its selected phenotypic effects. For example, consider our gene H for thick hair expressed during a temporary warm spell. Things then go wrong: the developmental system faithfully produces an organism with thick hair, but it does badly. Where does the blame lie? The developmental system is still carrying out the imperative for which the gene was selected. But the gene is now misleading its consumer – it is telling the developmental system that *the environment is cold* when it's not. So we can explain the failure in contentful terms: the content carried by gene H is now false. Things can go wrong at the consumer end too. Suppose something interferes with the development of an organism with gene H so that it fails to grow thick hair. The organisms start to get chilly, and reproduce less well as a result. This time failure is to be explained by reference to the imperative content: the

¹⁵ Godfrey-Smith 1999, pp. 310-311; Papineau 2003, pp. 121-122.

consumer system is no longer following the instruction to *produce thick hair*. So false content (or failure of a satisfaction condition) explains failure.¹⁶

Similarly, correctness explains success. We can explain an organism's having a phenotypic trait by the existence of a gene which tells that the environment is conducive to that trait and instructs the developmental system to produce it. The individual cases ramify into an explanation of DAC for the phenotype of the whole organism. The organism developed that complex phenotype because the zygote contained a large number of genes instructing the developmental system to produce a large number of traits. This representational explanation of adaptive complexity is a phylogenetic one, which abstracts away from the detailed processes of development, focusing only on gene-phenotype and gene-environment correlations. As such, it is perfectly compatible with, and complementary to, an ontogenetic explanation of DAC. Genes will also play a role in the ontogenetic explanation, but only as one of a number of factors that interact in highly complex ways to eventuate in mature organisms. However, only developmental properties of genes will figure in the ontogenetic explanation. By contrast, genes' representational contents are fixed by their selectional properties, and so can figure only in phylogenetic explanations. That is why the semantic information about phenotypes carried by genes can have no role to play in an ontogenetic explanation of DAC, as noted at the end of section (2) above. It also shows why a negative conclusion about semantic information in genes should not be drawn from their failure to play a role in an ontogenetic explanation of DAC. The representational content of a gene is amongst its selectional properties and so can engage only with phylogenetic, not ontogenetic, explanations.

Since contents arise from selection, some genes which have clear developmental properties will not be representations. A gene-P for a genetic disease will not be representational (unless, like sickle cell anaemia, it was selected for some other reason). Nor will a new mutation; which is to be expected – these are both outcomes that can be explained ontogenetically, but not phylogenetically. Hitchhiker genes also fail the test (Wheeler 2003). They arise where a gene-P for a phenotypic outcome which is selected for is genetically linked to a gene-P for some other phenotypic outcome which is selectively neutral, so that the two genes tend to be inherited together. The case is explained by combining the developmental properties of the neutral gene with the selectional properties of the selected gene. So the neutral gene will not represent; nor should we expect it to.

This account has not been designed only to apply to genes. Accordingly, I accept a version of Griffiths' parity thesis (2001): that the semantic information carried by genes may also be found in other non-genetic developmental resources. However, I do not accept that every causal factor in development will carry semantic information, if genes do.

¹⁶ Thus, genetic representation neither implies nor requires genetic determinism. Although the imperative contents of genes may sound determinative (Oyama 1985, Griffiths 2005), it must be remembered that genetic instructions may easily go unsatisfied, genetic indicatives may well be false of the current environment, and none carries any normative force. In this respect, it is quite like the representational contents in (tabloid) newspapers.

Genes only represent, according to my account, because they play a very particular role in inheritance. They are intermediaries in a mechanism which has the meta-function to produce heritable phenotypes. Otherwise the representational framework would not apply. So there is gene-type representation only in genuine inheritance systems. As we saw in section (3), that is a substantial restriction. Chromatin marks may turn out to be an inheritance system, but morphogen gradients almost certainly won't. Imitation learning in primates will probably count. But for animals other than humans, whether non-genetic inheritance systems have any real importance is a wide-open empirical issue.

My account of genetic representation differs from previous attempts to apply teleosemantics to genes in being constrained by DAC, the explanandum to which attributions of genetic representational content are addressed (rather than just trying to capture uses of the term 'information' in biology), in distinguishing between ontogenetic and phylogenetic explanations of DAC, and in identifying representation in the genome as a selectional property which can only figure in phylogenetic explanations of DAC. Jablonka 2002 has only indicative contents. Sterelny et al. 1996 has no role for consumers. Maynard Smith 2000 and Sterelny 2000 appeal to the arbitrariness of the causal specificity between DNA triplet and amino acid. Wherever there are several steps between the mechanism of inheritance and the phenotypic expression on which selection acts the specificity between representational intermediate and its content is likely to be arbitrary. However, I see no reason why arbitrariness should be criterial for representation, especially in simple representing systems.

If genetic representation were founded in there being a genetic code (eg, Wheeler 2003, forthcoming), then a system of inheritance by protein samples would not count as representational (Godfrey-Smith 2000a, 2002). By contrast, my account would locate genetic information in inherited protein samples, if that were how inheritance worked.¹⁷ That is a virtue since, as we have seen, it allows that there is representation in any genuine inheritance mechanism, whether or not it operates via the kind of code found in DNA. Thus, the view is neutral on the contingency of the DNA-protein code. It is also neutral on the importance of other mechanisms of inheritance. However, it is empirically committed to DNA being what I have called a genuine inheritance system, that is, to its having the meta-function of producing heritable phenotypes. Once that condition is met though, the view is compatible with a range of findings about the copying fidelity of inheritance systems, and of their ability to generate variation or resist outlaw replicators. Nor is it committed to a particular level of developmental modularity or canalisation.¹⁸

¹⁷ There are good reasons why inheritance does not work by protein samples. The combinatorial power of DNA bases is a good way to perform DNA's meta-function of producing heritable phenotypes.

¹⁸ Sterelny 2004 is a useful discussion of the importance of these various properties.

(5) Conclusion

Living things have the characteristic ability to regenerate dazzling adaptive complexity in each generation, starting from a fertilized egg. This outcome can be explained in two ways. Ontogenetically, we can chart the intricate course from genes, cellular machinery and environmental factors to the mature phenotype. Developmental properties of genes figure in this explanation, but as one causal factor amongst many. Three kinds of non-semantic property which might be called ‘information’ have a role to play in explaining ontogeny. Firstly, correlations between genes and their downstream products: expressed proteins and, in some cases, phenotypic traits. Secondly, the coding relation – the causal specificity between DNA bases and the primary structure of the proteins which result from their expression – figures in explanations of development. Finally, ontogenetic explanations rely on the fact that DNA is a stable molecule with a very large number of differentiable states, ie, on its large information capacity.

There is a complementary phylogenetic explanation of the zygote’s capacity to give rise to the adaptive phenotype of a complex organism. The zygote contains DNA, with its high information capacity, which is a representational intermediate consumed by the mechanisms of development. Over evolutionary time the mechanism of differential survival and reproduction acts so as to send a range of genetic messages to the consumer system about which phenotypes to produce, and about the conduciveness of the environment to those phenotypes. Since the representations themselves change only over evolutionary time (order 1), the identity of the consumer system can only be discerned from the perspective of higher-order evolutionary time (order 2) over which the developmental system acquires the meta-function – to produce heritable phenotypes – which makes it a representation consumer. That is, genes are an intermediate in an inheritance system. Gene-based selection ensures that genes carry correlational information about the conduciveness of the environment to the phenotypic traits with which they correlate during selection. Therefore, the conditions are satisfied for application of infotel semantics, a framework for representation in simple systems. We can then see that genes do indeed have correctness conditions and satisfaction conditions, albeit of a rather low-level variety. Other genuine inheritance systems will have representational contents in the same way, but few developmental factors have the right kind of meta-function to be part of an inheritance system, and thus to be representational. Unlike the ontogenetic account, the representational explanation of the development of adaptive complexity is phylogenetic. It both adverts to and explains relational properties of the organism, complementing the ontogenetic explanation by abstracting away from the details of the process of development in order to capture real patterns in the organism’s interactions with its environment. So it is in phylogenetic explanations that we should advert to representation in the genome and in other inheritance systems. Three kinds of properties which may be called ‘information’ can figure in ontogenetic explanations: correlational information,

coding and information capacity. Semantic information cannot.

Acknowledgements

Many thanks to the following for comments on this paper and earlier versions of this material: Peter Godfrey-Smith, Susan Hurley, Matteo Mameli, James Maclaurin, Ulrich Stegmann and Kim Sterelny; and audiences at the Universities of Oxford, Reading and at the Intelligent Autonomous Systems Laboratory at the University of Western England. The author gratefully acknowledges the support of the British Academy which funded this work through a postdoctoral research fellowship.

References

- Alberts, B., D. Bray, et al. 2004. *Essential Cell Biology*. New York & Abingdon, Garland Science.
- Dennett, D. C. 1981. True Believers: The Intentional Strategy and Why It Works. In A. F. Heath (ed.) *Scientific Explanation*. Oxford, O.U.P.
- Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, M.A., MIT Press.
- Field, H. 1978. Mental Representation. *Erkenntnis* 13, pp. 9-61.
- Flatt, T. 2005. The evolutionary genetics of canalization. *The Quarterly Review of Biology* 80(3), pp. 287-316.
- Godfrey-Smith, P. 1991. Signal, decision, action. *Journal of Philosophy* 88, pp. 709-722.
- Godfrey-Smith, P. 1993. Functions: consensus without unity. *Pacific Philosophical Quarterly* 74, pp. 196-208.
- Godfrey-Smith, P. 1999. Genes and codes: lessons from the philosophy of mind. In V. Hardcastle (ed.) *Where Biology Meets Psychology: Philosophical Essays*. London / Cambridge, MA, MIT Press.
- Godfrey-Smith, P. 2000a. Information, arbitrariness and selection. *Philosophy of Science* 67, pp. 202-207.
- Godfrey-Smith, P. 2000b. On the theoretical role of "genetic coding". *Philosophy of Science* 67, pp. 26-44.
- Godfrey-Smith, P. 2002. On genetic information and genetic coding. In P. Gardenfors, J. Wolenski and K. Kajania-Placek (eds.) *In the Scope of Logic, Methodology, and the Philosophy of Science*. Dordrecht, Kluwer. Vol. II, pp. 387-400.
- Godfrey-Smith, P. 2006. Information in biology. In D. Hull and M. Ruse (eds.) *The Cambridge Companion to the Philosophy of Biology*. Cambridge, CUP.
- Grice, P. 1957. Meaning. *Philosophical Review* 66, pp. 377-388.
- Griffiths, P. E. 1993. Functional analysis and proper functions. *British Journal for the Philosophy of Science* 44, pp. 409-422.

- Griffiths, P. E. 2001. Genetic information: a metaphor in search of a theory. *Philosophy of Science* 68, pp. 394-412.
- Griffiths, P. E. 2005. The fearless vampire conservator: Philip Kitcher, genetic determinism and the informational gene. In E. M. Neumann-Held and C. Rehmann-Sutter (eds.) *Genes in Development: Re-reading the molecular paradigm*. Durham, NC, Duke University Press.
- Griffiths, P. E. and R. D. Gray 2005. Discussion: three ways to misunderstand developmental systems theory. *Biology and Philosophy* 20, pp. 417-425.
- Jablonka, E. 2002. Information: its interpretation, its inheritance, and its sharing. *Philosophy of Science* 69, pp. 578-605.
- Jablonka, E. and M. J. Lamb 1995. *Epigenetic Inheritance and Evolution: the lamarkian dimension*. Oxford / New York, OUP.
- Jablonka, E. and M. J. Lamb 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge, MA, MIT Press.
- Lehrman, D. S. 1970. Semantic and conceptual issues in the nature-nurture problem. In D. S. Lehrman (ed.) *Development & Evolution of Behaviour*. San Francisco, W. H. Freeman & Co., pp. 17-52.
- Lewontin, R. 1978. Adaptation. *Scientific American* 239, pp. 156-169.
- MacLaurin, J. 2002. The resurrection of innateness. *Monist*, pp. 105-130.
- Mameli, M. 2005. The inheritance of features. *Biology and Philosophy* 20, pp. 365-399.
- Maynard Smith, J. 2000. The concept of information in biology. *Philosophy of Science* 67, pp. 177-194.
- Maynard Smith, J. and E. Szathmáry 1995. *The Major Transitions in Evolution*. Oxford, Freeman.
- Millikan, R. G. 1996. Pushmi-pullyu Representations. In J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 9. Atascadero, CA, Ridgeview Publishing, pp. 185-200.
- Millikan, R. G. 2000. *On Clear and Confused Ideas*. Cambridge, Cambridge University Press.
- Millikan, R. G. 2004. *Varieties of Meaning*. London / Cambridge MA, MIT Press.
- Moss, L. 2001. Deconstructing the gene and reconstructing molecular developmental systems. In S. Oyama, P. E. Griffiths and R. D. Gray (eds.) *Cycles of Contingency: Developmental Systems and Evolution*. London / Cambridge, MA, MIT Press.
- Moss, L. 2003. *What Genes Can't Do*. London & Cambridge, MA, MIT Press.
- Oyama, S. 1985. *The Ontogeny of Information: developmental systems and evolution*. Cambridge, Cambridge University Press.
- Papineau, D. 2003. Is representation rife? *Ratio* 16, pp. 107-123.
- Sarkar, S. 2004. Genes encode information for phenotypic traits. In C. Hitchcock (ed.) *Contemporary Debates in Philosophy of Science*. Oxford, Blackwell, pp. 259-274.
- Schlosser, G. and G. P. Wagner, Eds 2004. *Modularity in Development and Evolution*, London / Chicago, University of Chicago Press.
- Shannon, C. E. 1949. The mathematical theory of communication. In C. E. Shannon and W.

- Weaver (eds.) *The Mathematical Theory of Communication*. Urbana, University of Illinois Press.
- Stegmann, U. E. 2005. Genetic information as instructional content. *Philosophy of Science* 72(3), pp. 425-443.
- Sterelny, K. 2000. The "genetic program" program: a commentary on Maynard Smith on information in biology. *Philosophy of Science* 67, pp. 195-201.
- Sterelny, K. 2004. Symbiosis, evolvability, and modularity. In G. Schlosser and G. P. Wagner (eds.) *Modularity in Development and Evolution*. London / Chicago, University of Chicago Press.
- Sterelny, K., K. C. Smith and M. Dickson 1996. The extended replicator. *Biology and Philosophy* 11, pp. 377-403.
- Von Neumann, J. 1966. *Theory of Self-Reproducing Automata*. London / Urbana, University of Illinois Press.
- Weaver, W. 1949. Recent contributions to the mathematical theory of communication. In C. E. Shannon and W. Weaver (eds.) *The Mathematical Theory of Communication*.
- Weber, M. 2005. *Philosophy of Experimental Biology*. Cambridge / New York, Cambridge University Press.
- Wheeler, M. 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. Synthese Library vol. 320, pp. 151-164.
- Wheeler, M. forthcoming. Traits, genes and coding. In M. Matthen and C. Stephens (eds.) *Handbook of the Philosophy of Biology*. London / Amsterdam, Elsevier.