DOI:
10.1007/s10539-010-9232-4

Document Version
Early version, also known as pre-print

What's Transmitted? Inherited Information.

Nicholas Shea

1 Introduction

In response to worries that uses of the concept of information in biology are metaphorical or insubstantial, Bergstrom & Rosvall (B&R) have identified a sense in which DNA transmits information down the generations. Their 'transmission view of information' is founded on a claim about DNA's teleofunction.

Sterelny et al. (1996) and Maynard Smith (2000) were the first to argue that genetic information should be understood in terms of teleofunctions, leading to a sense in which genes carry semantic information, with conditions of correctness or satisfaction. In Shea (2007a) I argued that a modified teleosemantic account could be sustained in the face of various objections to these views. The account has two key elements. First, semantic content comes into view only when we see DNA as carrying messages between generations, rather than sending messages during the course of individual development. Second, we need to be clear that there is a consumer of these messages. I argued that there is a real consumer: the process of development takes zygotic DNA as input and produces phenotypic traits as output – bracketing the intricacies of the actual processes of individual development. The message carried by DNA down the generations can only be discerned by considering the way that genes are designed to be consumed.

Bergstrom and Rosvall's (B&R) focus on transmission properties of DNA is a significant advance. They agree with the first point from Shea (2007a) about transmission down the generations, but not the second, because they see their transmission view of information as a rival to semantic accounts. This commentary argues that it is complementary. The idea that DNA is transmitting information down the generations only makes sense if it is carrying a message, that is to say if it has semantic content. Section 2 below argues that B&R's transmission view needs to be supplemented with a sense in which DNA carries semantic information. But the account in Shea (2007a) also needs the insights B&R bring, because it is based on the claim that DNA has the function of carrying information down the generations. By showing that DNA is adapted to playing the role of a Shannon-type communications channel, B&R offer a strong argument that DNA really does have that biological function. Furthermore, by taking the perspective of a communications engineer, B&R offer novel insights into the way evolution has designed the DNA-based inheritance system to operate. Section 3 goes on to address the objection that there is no real consumer of zygotic DNA of the sort that would underpin teleosemantic contents.

2 Semantic Contents

Critics of genetic information claim that genes and other developmental factors are on a par in carrying information (Griffiths & Gray 1994, Griffiths 2001, p. 398). B&R's answer is that only elements on which natural selection acts constitute an information channel – only those which are the basis of heritable variation [p. 5]. Such elements will thereby have

teleofunctions. In fact, when they come to define transmission information, B&R make a significant further restriction. Only a factor X which has a very particular teleofunction, 'to reduce, by virtue of its sequence properties, uncertainty on the part of an agent who observes X', carries transmission information.[1]

I agree that some such metafunction is needed to vindicate the existence of substantive genetic information. But, even if we read 'agent' in a very thin sense, B&R's formulation invites the question: about *what* is the agent's uncertainty reduced? B&R rightly argue that we don't need to know what DNA represents in order to be able to see that it transmits information.[2] We can infer that from evidence that DNA is adapted to functioning as a communications channel. But DNA could only have the metafunction B&R specify if there *is* a message that it carries, which presupposes that there is a sense in which DNA genuinely does carry semantic information.

B&R do seem to accept that there is a message in the genome, and that it concerns phenotypic features of the organism:

> we know that genes are transmitted from parent to offspring in order to provide the offspring with information about how to make a living (e.g. metabolize sugars, create cell walls, etc.) in the world. [p. 4]

So there is a functional message about 'how to make a living'. There is a real puzzle about how natural selection could have adapted DNA to this transmission role. Does it require controversial lineage-based selection? Or can it be accounted for by the long-run fitness benefits of improved transmission fidelity, with the short-run fitness costs of improvements in fidelity being zero or very small (as I have suggested: Shea 2007a, p. 323, Shea 2009, p. 2432)? Despite these uncertainties, the evidence is strong that DNA has been adapted to a transmission role, so that it does have the evolutionary function of transmitting heritable phenotypes down the generations. B&R point to six lines of evidence in support:-

(i) it stores and transmits an arbitrary sequence
(ii) a long sequence is stored in a small space
(iii) the sequence is indefinitely extensible
(iv) the sequence is inert and structurally stable
(v) the sequence is very easy to replicate

---

[1] It seems that B&R intend this to be a necessary condition as well.

[2] Following de Ruyter van Steveninck et al. (1997), B&R observe that calculating how much information is being carried does not depend upon knowing what the message is. However, it does require that you know what the semantically-significant coding elements are (the syntax, not the semantics, roughly). For example, de Ruyter van Steveninck et al. assumed that rate coding, rather than phase coding, say, was the bearer of relevant messages in the fly's neurobiology.

(vi) the mapping from codon triplets to amino acids is optimised by reference to facts about how it is transmitted, by comparison to randomly-generated mappings

The mapping is optimised with respect to transmission errors in two ways. Firstly, simple translational errors generate the same amino acid, or an amino acid which is similar in an important chemical property, its affinity for water (and in its polar requirement, which is closely related) – although similarity in other chemical properties is not optimised (molecular volume and isoelectric point). Secondly, the mapping is optimised to the errors that are more common, given observed biases in rates of point mutation and in mistranslation.

Since features (i) to (vi) are not restricted to multicellular organisms, they cannot be explained away as adaptations for somatic cell inheritance. The conjunction of these features is good evidence that DNA has been adapted to serve as a channel for communicating information down the generations. They increase fidelity, so that useful adaptations are preserved. Feature (vi) also aids evolvability by smoothing the mutational landscape, making it more likely that common mutations lead to phenotypes that are nearby in the space of phenotypic possibility. I will argue that features (i) and (vi) do not just concern the transmission of the sequence itself, but also depend on the sequence carrying semantic information.

At first pass, arbitrariness (i) and optimisation (vi) seem to be in tension: optimisation implies that the code is anything but arbitrary. Rather, it is a one-in-a-million solution to the communication problem. Indeed, arbitrariness itself is a puzzling concept, since in any multi-stage causal process, an intermediate *could* be replaced by an alternative, provided the there are compensating changes in the way it is produced and acted upon. To take one example, the relation between the direction of the honeybee's nectar dance and the direction of nectar appears to be systematic, not arbitrary, but the system could work just as well with a different relation, provided the behaviour of producers and consumers were adjusted in corresponding ways.

The literature on costly signalling indicates a better way to think about arbitrariness. It is not a categorical property, but a matter of degree, depending upon the costs associated with moving to a different signal. Some female frogs use the pitch of male frog sounds as a proxy for size, hence fitness. That signal is relatively non-arbitrary because there would be a very high cost for a small frog to produce large vocal apparatus. Correlatively, the arbitrariness of the genetic code pointed to at (i) lies in the fact that there is little difference in cost between different ways of setting up the mapping between codon triplets and amino acids. As a result, the code is arbitrary with respect to its message about phenotypic properties. If the message were just the sequence itself, then the code would not be arbitrary.

The tension between arbitrariness and optimisation is resolved when we see that arbitrariness consists in the fact that costs to the developing individual are largely independent of how the triplet-amino acid map is set up. That is what allows the mapping which is chosen to be optimised in the light of common errors due to point mutations. It is from the point of view of the problem of keeping that mapping stable over evolutionary time that we see that

different mappings have different costs, because of the distribution of errors introduced as the message is transmitted down the generations.[3]

Notice that the non-arbitrariness in (vi) is dependent on the content of the message being sent. The smoothness of the mutational landscape is specified with respect to chemical properties of amino acids for which DNA codes. Chemical similarity will tend to lead to proteins with similar physiological, hence phenotypic effects. That is optimization with respect to a presumed message. So both these lines of evidence that DNA is adapted to an informational function depend upon assumptions about what DNA is coding for, and so on the assumption that it *has* semantic content.

Finally, B&R's argument about the directionality of transmission relies on an assumption about what is coded. They argue that the data processing inequality reveals the directional flow of Shannon information. When there are correlations between a random variable X measured by a sender and another random variable Y, measured by a receiver, then the mutual information between X and Y is symmetric: each carries the same amount of information about the other ($I(X ; Y) = I(Y ; X)$). Asymmetry arises, according to B&R, when the message Y concerns some feature of the external world W with which the variable observed by the sender itself correlates. For then the data processing inequality ensures $I(W ; X) \geq I(W ; Y)$. That asymmetry only arises if the content of the message is some further property of the world. If the function of the DNA sequence were just to reduce the uncertainty on the part of an agent about sequence properties themselves, then we would be back to the symmetry of mutual information. B&R's prescient observation about the directionality of the flow of genetic information is surely right, but it relies on DNA carrying semantic information about something other than its own sequence properties.

3 Consumers

One prominent argument against genes carrying semantic contents accepts that teleosemantics would deliver such contents if DNA had a genuine representation consumer. The objection is that in reality there is no such consumer (Godfrey-Smith 2007). B&R emphasise that their transmission information flows down generations, orthogonally to the processes going on in individual development, and caution against focusing on 'how information goes from an encoded form in the genotype to its expression in the phenotype' [p. 4]. The absence of any real consumer would undermine the teleosemantic approach and thus support B&R's attempt to give a semantics-free account of the information transmitted by DNA down the generations. This section answers the objection about consumers by taking a difficult case: single-celled bacteria. I argue that even there, in the absence of a clear distinction between zygote and organism, there really are consumers of the message carried by DNA.

---

[3] That is an idealisation, since the code is also optimised by reference to errors that occur in individual development (e.g. frame-shift errors, ribosome "traffic congestion"). Such optimisation is only possible because, in every other way, the two synonymous codons impose very similar developmental costs.

Both B&R and Shea (2007a) rely on a division between the genetic inheritance channel and the developing organism. It is harder to spot the channel in the case of bacteria, since the germ-line DNA continues to play an active role in the metabolic activity of the organism throughout its lifetime. If the organism (bacterium) is the consumer of genetic information, the message itself (DNA) looks to be a substantial *part* of the consumer. A second problem is to make a distinction between *producer* and consumer of the genetic message, since in bacterial cell division it may be unclear which of the two resulting cells is the mother and which is the daughter.

The first problem is to distinguish the consumer from the message. Granted, eukaryotes have much more complex developmental programmes than prokaryotes. So it is more obvious in eukaryotes that the DNA is being read so as to guide development, rather than just playing a stable, but essentially physiological / metabolic role in the activity of the cell. But in bacteria there is still a fundamental asymmetry. Changes in DNA will have downstream effects far into the future, if they are viable at all. Viable changes in other cellular factors have much more short term effects. Further, DNA plausibly has the teleofunction of playing this role. The evidence about DNA's informational metafunction highlighted by B&R is equally applicable to bacteria. So the DNA is the message and the rest of the organism is the consumer. The message sent by DNA continues to be consumed during the lifetime of the organism, but that does not undermine a distinction between message and consumer. If other factors, like chromatin marking, also have the function of passing on heritable phenotypes, then there will be other channels of inheritance. But that is a demanding test, so there is no threat that most of the bacterium will end up counting as message. There is a clear distinction between message and consumer, and much of the organism constitutes the consumer.

A further peculiarity in bacteria is that two copies of the DNA are made before the cell has divided; indeed, those copies will have begun the duplication process again before the cell walls have been sealed around the first division (Lau et al. 2003). Here, B&R's observation that data storage and data transmission are mathematically equivalent comes in handy. Both initial duplication of the DNA, and stable preservation of those duplicates in the cell, are part of the process of transmitting genetic information to offspring cells. If we need to identify a point when consumption begins anew, it is the moment when the dividing cell wall closes to make two cells. That is the point when the metabolic activity of the two cells is separated so as to create two organisms. There will be fuzzy boundaries here too of course, but the difference between one organism and two organisms is an uncontroversial distinction that is relied on for lots of purposes in biology. The account of inherited information has no need to postulate extra biological properties that have not previously been recognised as significant. When the cell has divided, a new consumer comes into existence, and the message carried by the genome begins to be read by that new organism.

A final problem is to say, in bacteria, which of the two resulting cells is the producer and which is the consumer. It may be possible to distinguish between mother and daughter in bacterial cell division (Stewart et al. 2005), in which case the mother is the producer of the message, which gets passed to the daughter cell and begins to be consumed when the daughter cell is formed. Even if there is no asymmetry, we could just treat one cell as mother and the

other as daughter. But it might be better, given the symmetry of the situation, to see two new cells being formed out of the original, so that the parent cell sends two DNA messages, one to each of the new cells which are formed. If so, there would be two new consumers, in addition to the parent producer of the message.

It may not matter which way we go. Indeed, although teleosemantics relies on there being genuine consumers whose reality is independent of semantic facts, it is not clear that it needs there to be well-defined producers. Infotel semantics adds to teleosemantics the additional requirement that the message should carry correlational information (Shea 2007b), but it does not require there to be a mechanism which has the function of producing that correlation. In the case of genetic information, it is the process of natural selection that generates correlational information in the genome. So there is no need to view individual organisms as producers. The problem of differentiating producer from consumer when a bacterium divides just does not arise. The whole process is well-characterised in terms of natural selection giving rise to information over phylogenetic time, read in ontogenetic time by a series of the organisms, which are the consumers of the message carried by DNA. That is exactly the framework relied on by B&R and illustrated in their figure 4 – but it does not dispose of the notion of consumers. The idea that individual organisms merely reacting to DNA are consumers remains critical to the claim that DNA carries semantic information.

4 Conclusion

B&R's transmission view of information depends on there being a sense in which semantic information is carried by the genome, because they rely on DNA's carrying a content or message. So the inherited information of Shea (2007a) is complementary to B&R's transmission view of information, not a rival. Correlatively, the several strands of B&R's ingenious paper serve significantly to strengthen the case for the existence of substantive genetic information, underpinned by DNA's teleofunctions. An important worry about teleosemantic accounts of semantic information in the genome is their reliance on the reality of information consumers. However, even in the difficult case of bacteria the reality of DNA consumers has a biological basis which is independent of debates about genetic information.

Acknowledgements

Many thanks to Carl Bergstrom for patient discussion of his view, and to Peter Godfrey-Smith for an extended dialogue on these issues over several years, including comments on this commentary. The research reported here was supported by the OUP John Fell Research Fund, the James Martin 21st Century School, the Oxford Centre for Neuroethics and Faculty of Philosophy, University of Oxford.

References

de Ryter van Steveninck et al. (1997), *Science* 275, 1805

Godfrey-Smith, P. (2007). Innateness and genetic information. In P. Carruthers & S. Laurence & S. Stich (Eds.), *The Innate Mind: Foundations and the Future* (pp. 55-68). Oxford / New York: O.U.P.

Lau, I. F. (2003). Spatial and temporal organization of replicating *Escherichia coli* chromosomes, *Molecular Microbiology*, 49, 731-743.

Maynard Smith, J. (2000). The concept of information in biology. *Philosophy of Science*, 67, 177-194.

Shea, N. (2007a). Representation in the genome, and in other inheritance systems. *Biology and Philosophy, 22*, 313-331.

Shea, N. (2007b). Consumers Need Information: supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research, 75*(2), 404-435.

Sterelny, K., Smith, K. C., & Dickson, M. (1996). The extended replicator. *Biology and Philosophy*, 11, 377-403.

Stewart EJ, Madden R, Gregory P, Taddei F (2005) *PLoS Biol* 3:295–300.