Introduction

The notion of information has developed in a number of different ways (as discussed in this volume), and many of them have been applied to biology, both usefully and gratuitously, and even misleadingly. These multiple notions of information have not surprisingly led to apparently contradictory claims by authors who have really been talking past each other, although there are also substantive issues at stake. The aim of this chapter is to review some of the ways that notions of information have been used in biology, to disentangle them, and to evaluate their implications and aptness, as well as to point out some of the more widespread confusions.

In particular, I will compare the use of information as a technology of measurement, which does not imply that there is anything present that might be called ‘information’, with a stronger usage of information in biology that attributes information to biological systems in a non-instrumental way. This distinction between instrumental and substantive uses of information in biological studies often turns on the notion of information used, so it is important in each case to be clear what is at stake. Where there is a choice, I will focus on the substantive use of information in biology. Roughly, substantive use of information uses information in an explanatory way in addition to any representational instruments.¹ I will not discuss what falls under the general heading of

¹ Sarkar (2000) makes a similar distinction between heuristic and substantive uses of information, but as an avowed instrumentalist he does not see a clear distinction. In particular, one would assume he sees no special role for explanation in the way that Chomsky (1959), for example, distinguishes between descriptive and explanatory adequacy. I believe that some of Sarkar's obtuseness about the role of information in biological systems is a result of blindness to the distinction, resulting in a failure to consider things relevant to the higher standards required for explanation. Maynard Smith (2000a, 2000b) attributes
bioinformatics in this chapter, since this is more about how we represent biological information than the information itself.

It will be impossible to cover all the varied uses of information concepts by biologists, so I will look primarily at cases that seem to be historically significant or else philosophically pivotal (the two often correspond).² The central case I will look at is heredity. The association of information with heredity goes back at least to Weissmann (1904), and was adopted by such disparate biologists as Francis Crick (1958) and Konrad Lorenz (1973). It is difficult to find well known theoretical biologists who object to the use of information concepts in relation to genetics, and if anything the use of information concepts in biology has increased over the last few decades. Dawkins (1986: 112) declared: “If you want to understand life, don’t think about vibrant, throbbing gels and oozes, think about information technology.” Increasingly, the “throbbing gels and oozes” can themselves be understood as made up of molecular machines that process information (Holzmüller 1984, Schneider 1991a, 1991b, 2000, 2006, Darden 2006). In order to give a strong grounding in accepted theoretical biology, I will take my lead from the role of information assigned by Maynard Smith and Szathmáry in The Major Transitions in Evolution (1995). They argue that the increase in complexity observed (however roughly) in some lineages results from relatively few major transitions in how genetic information is passed down from generation to generation. As we shall see, things are possibly and probably more complicated than this relatively simple hypothesis, but following it critically will raise some important philosophical issues. Importantly, however, Maynard Smith and Szathmáry use information explanatorily and their views and usages are fairly authoritative; therefore, presumably they pick out authoritative substantive uses of information in biology.

² In 1987 I did a search on the last three years of Biological Abstracts (on CD, at the University of Indiana Biology Library). Based on the abstracts, I tried to judge whether the use of information in the paper was required, or was more or less gratuitous. I found as few as seven abstracts that seemed to me to use the information concept in some essential way. The situation was worse for the entropy concept, which had only two clearly non-gratuitous mentions out of over 200 papers that used it. I would expect that things have not improved, so there is understandable suspicion about the use of these related concepts.
There are a variety of mathematical technologies that can be used for information measurement, but they fall into three general classes (Kolmogorov 1965): statistical (e.g., Shannon and Weaver 1949), combinatorial (a variation on Shannon methods not used directly by Shannon himself), and algorithmic (Chaitin 1987). The last has inspired two technologies for information measurement that have been applied to DNA and other biological objects: minimum description length – MDL (Rissanen 1989) and minimum message length – MML (Wallace and Freeman 1987). It is worth noting that the statistical methods are best used on ensembles, whereas the combinatorial and algorithmic methods work best on individuals. This suggests that the latter methods are more appropriate for dealing with information in biological organisms, even though the statistical approach is used so widely that it is often taken to be information theory (for more on this, see Winnie 2000). Despite this, each of the general classes of methods can be used on any particular subject matter with clever adaptation. Thus there is nothing in the mathematical methods themselves that distinguishes the use of information technology in studying the properties of a system from the substantive attribution of information to a system. In particular, the instrumental usefulness of information technologies does not in itself imply the existence of substantive information within the system being studied, at least not without more being said. The instrumental usefulness of information may, for example, reflect epistemic considerations such as how we decide to organize our data. Furthermore, mathematical methods are limited by their nature to the syntactic (formalizable) aspects of information, and are not suited in themselves to dealing with issues of communication, representation, semiosis and meaning, all of which are have an importance in biology that they do not have in, say, physics.3

In order to deal with these issues, and evaluate whether or not information in some substantive role has a place in biology, it is useful to give a classification of the ways in which information has been thought to play a substantive role in the sciences. Ignoring many fine distinctions, the basic ways can be placed into an increasingly nested hierarchy: “it from bit”, negentropy, hierarchical negentropy, functional information, and

---

3 This is also true of formal methods in general, including the Barwise-Seligman idea of an information channel in distributed systems (Barwise and Seligman 1997). Their approach invokes “regularities”, which cannot be understood purely formally (accidental regularities do not carry information), but even this informal part of their approach does not imply anything more than non-accidental relations, which can be found in the most basic physics.
meaningful information (see Figure 1). Each inherits the logical and ontological commitments of the containing views, but adds further restrictions.

The most liberal and inclusive view is the “It from bit” view. It has originated independently from so many people that it is pointless to attribute an origin, though it probably goes back to Leibniz’ view that the world has a logical structure of perceptions based in the ability to discriminate. The term is due to John Wheeler, and the view has recently been powerfully if controversially championed by Stephen Wolfram (2002). On this view, any causally (dynamically) grounded distinction makes a difference, thereby ensuring that it is information (MacKay 1969, see also Bateson 1973). On this view information is objective, and there is nothing else.

The negentropy view of information is a restriction on the It from bit view. Only those Its that are capable of doing work (either directing and using energy or sorting things) count as information. The rest is disorder. This view is due to Schrödinger (1944), though the groundwork was done by Szillard, and the implications were generalized and clarified by Brillouin (1962), though the significance is still controversial (Earman and Norton 1998, 1999). The motivation for this view is that work is required for control, and the information in microstates beyond that in macrostates is hidden from view in macroscopic interactions (Collier 1990b). Negentropy measures the capacity for control (in bits this is the number of binary discriminations that a system can possibly make).

The next view is a restriction of the negentropic approach to particular levels of a physical hierarchy, so that information is relativized to a cohesive level of an object, such as an organism or a species. The view is due to Brooks and Wiley (1988), Collier (1986, 2003) and Smith (1998). The idea is that not all negentropy is expressed at a given level, and the “Its” available are level relative. This information is a measure of the constraints on the objects within the level; because of their connection to biological and cognitive form, Collier (1990a, Collier and Hooker 1999) calls this expressed information enformation to distinguish it from other forms of negentropy (for example, statistical

---

\[^4\] Stan Salthe (1985, 1993) also uses similar technology very widely, but to different effect, as does Robert Ulanowicz (1986, 1997) in ecology. I will not discuss these uses here, as it would take me much further afield from the issues of heredity and function that are my main focus. Wicken (1987) and James Kay and Eric Schneider (1994) specifically avoid this information technology, at least in name, though they have been in close communication with the authors that do explicitly use this levels based information technology. They also use the technology very widely, including ecology.
information due to nonequilibrium conditions is sometimes called *intropy*).\(^5\) Expressed information is relative to a certain level in a hierarchy (Collier 2003), an idea that will be clarified below. Expressed information at higher levels is able to control information at lower levels only to a certain degree (sometimes called downward causation), but can control information at its own and higher levels more completely (Collier 1990b, 2003). This asymmetry is biologically important.

Restricting further, we have functional information, which is the expressed information that is functional. This sort of information is easily seen as information from outside the system. It has both syntax and semantics (reference), but does not require that the information is information for the system itself.\(^6\) Functional organization is biologically important, of course, but at least one common account of biological functionality tends to suppress the informational aspect. Whether or not we can call information that arises through functionality meaningful has been a subject of some debate. The nature of meaning is the great object of desire for information theory. I will address this issue throughout this chapter, but especially in the final two sections before the conclusion.

Within the scope of meaningful information is intentional information, or cognitive content. At the next level of restriction is social information, though some authors hold that cognitive content depends on language, which is a social activity. I will not discuss these levels further here, which is not to say that either they are unimportant, or are in some sense reducible to the information forms that I do discuss. These forms of information are better discussed in the context of cognition, though the biological roots of cognition are interesting, and connect to various forms of biological information.

**Information as a tool**

The use of information theory as a tool in biology is fairly widespread. Biological systems are both complicated and complexly organized, so information theory can be

---

\(^5\) Collier (1990a) borrowed this term from engineering usage. Lyla Gatlin (1972) called this information stored information, but this name is somewhat misleading, as it does not reflect the dynamical and often active nature that *expressed information* allows, nor its hierarchical nature.

\(^6\) Maturana and Varela (1980) apply this distinction by calling functional organization information externally, but not internally. I think this distinction is not sustainable (Collier 2004a, 2004b).
used to calculate or estimate the information content in biological structures from macromolecules to whole organisms, as well as in and between populations. This is not controversial, nor is this sort of application of information theory peculiar to biology. Similarly, communications theory can be used to analyze various biological channels, such as sensory processes, molecular communication, neural communication, intraspecies and interspecies communication and ecological networks in terms of their capacity, connectivity, order, and organization. Algorithmic information theory and its variants, MDL and MML are also useful for the last three, and Charles Bennett’s notion of logical depth may provide a measure of organization. Another useful tool, at least potentially, is the notion of information channel developed by Barwise and Seligman (1997) for the logic of distributed systems. Biological information channels, whether artifactual or inherent, are nothing if not distributed. Again, there is nothing particularly biological about these applications, and many of them are known from systems theory, electronics and computer science. Some of the applications, however, present interesting issues for the philosophy of biology, especially concerning whether the instrumental use of information is sufficient to explain the use of the idea of information by biologists.

DNA is probably the biological entity most closely associated with information. Maynard Smith and Szathmáry consider only hereditary transmission as the basis of information involved in increasing biological complexity, though they do mention major phenotypic changes that opened up new habitats, sensory inputs, and physiological mechanisms for adaptive (and not) radiation without direct change in hereditary channels. They are therefore committed (Maynard Smith and Szathmáry 1995, p. 8) to a gene-centered approach as outlined in Williams (1966) and explicit in Dawkins (1976). In light of recent work on developmental systems theory (Griffiths and Grey 1994, Oyama 2000, Jablonka and Lamb 1995) and niche construction (Odling-Smee et al 1996), both of which point to non-genetic channels of heredity, it now seems unlikely that the gene-centric view can be sustained.7

7 Griffiths (2001), Godfrey Smith (2000) and Sterelny (2000) criticize the gene centred approach offered by Maynard Smith, but seem to carry this over to substantive claims about information in biology, though they vary in the strength of this particular criticism. Griffiths argues for a “parity principle” that seems to lead him to the conclusion that information must be treated the same for all modes of heredity. I see no basis for this claim. On the other hand, Maynard Smith’s (2000b) response to alternative modes of heredity that these are ultimately grounded in the genes seems to be either wrong or beside the point. All biology is
One has to be careful here, however. Sterelny (2000) raises the issue that the regulation of genes depends on the cytoplasm, which is passed down maternally. He suggests that this presents a nongenetic mode of heredity that may contain semantic information. Now it may be true that semantic information is passed on in the cytoplasm, but at least for the regulatory part of the cytoplasm this is best understood not as information, but as part of the information channel that carries genetic information reliably. The cytoplasm is continuous from mother to offspring, and does not reproduce like the genes do. It plays a role in heredity, but not that of carrying information – it provides a continuity of channel for the genetic information, so that its expression is reliable. I will have more to say about channels later.

Despite the likely existence of other channels of inheritance, the most widely assumed view remains that DNA is the predominant if not only focus of biological information transmission. Genetic information is of undeniable importance, so it is worth looking at in more detail. I will look at some alternative channels of inheritance and their significance later. The workings of genes alone have implications for biological information channels, and how best to understand them, and most of the main issues can be understood within the scope of this focus.

DNA is often said to code for proteins, regulation, and various phenotypic properties from chemical networks in the body to social phenomena. None of this is straightforward, not even the mapping of DNA onto proteins. Since many biochemicals (such as, to take an instance, opium) are not proteins, there is no gene that codes for them. This is even more obvious for chemical networks and especially for behavior. Fortunately it is possible to get to the heart of the nature of genetic information without going into details of these complexities, but the technique required for doing so (information channels) perhaps opens up the field for biological information too far. More will be said about this soon. For the time being, focus will be on the channels of genetic heredity in terms of the Weissmannian view of separation of developmental and

ultimately grounded in chemistry and physics, but that does not mean that there aren’t special biological principles.

8 For discussion of these issues and a general empirically grounded critique of the centrality of the gene see (Keller 2000).
germ channels, rendered as the “central dogma” of molecular biology. This view is close enough to being true at the molecular level that it is useful.

With the above restrictions, the major transitions in evolution, according to Maynard Smith and Szathmáry are:

1. Replicating molecules → Populations of molecules in compartments
2. Independent replicators → Chromosomes (linked replicators)
3. RNA as gene and enzyme → DNA + protein (genetic code)
4. Prokaryotes → Eukaryotes
5. Asexual clones → Sexual populations
6. Protists → Animals, plants, fungi (cell differentiation)
7. Solitary individuals → Colonies (non-reproductive castes)
8. Primate societies → Human societies

Table 1: The Major Transitions (after Maynard Smith and Szathmáry 1995)

Several things to note about these transitions are 1) they occur in only some lineages, so they are not inevitable and do not confer an absolute fitness advantage, but are contingent and confer a fitness advantage at best only relative to others in their lineage 2) entities that were capable of independent replication before the transition can replicate only as part of a larger whole after the transition, 3) selfish mutations leading to conflict between levels of inclusion are possible in the larger whole – they happen – but there are so many chances for suppressor mutations in the whole that the influence of selfish mutations is rather small, 4) the transitions typically involve differentiation and specialization that increase efficiency, and 5) the notion of coding appears only in transition 3, but information concepts are applicable from the start.

Maynard Smith and Szathmáry note that 1-4 are common to all of the transitions, suggesting that there are some general principles at work. Since the early transitions are not part of biological evolution per se, but occur in molecular or chemical evolution, these general principles are not strictly biological. However, they note that point (5) marks what is perhaps the most significant transition – a division of labor that requires coding and translation. The major part of their book, not surprisingly, deals with this transition. I will argue below that transition 3 significantly enhances the role for
substantive information by separating through significant dynamical decoupling the roles of the energy and information budgets in prebiotic and living systems, and opens the door for semantic information in biological systems.

Arguably, to be alive requires this sort of separation of function and the requisite dynamical decoupling between metabolism and replication (Brooks and Wiley 1988, Brooks et al 1989, Maynard Smith and Szathmáry 1995, and many others), but nothing incontestable appears to rest on this definition of ‘living’, since the functional and dynamical separation are a matter of degree. In any case, the definition allows us to distinguish between chemical evolution, in which replication and metabolic processes are not distinct, and biological evolution, in which they are. A useful distinction made by Maynard Smith and Szathmáry is between limited replication and indefinite replication (1995, pp. 41-42). The former allows only a limited variety of forms dependent on the structure of the replicators, acting as templates. This sort of replication is highly subject to interference from parasitic replicators (ones that replicate at the expense of system organization, but nonetheless use system resources) that limits the size of reliably replicating structures. Limited replication cannot, therefore, support open-ended evolution. Sustained evolution is possible with template reproduction involving complementary base pairing, as with DNA. Is this sort of structure necessary for sustained evolution? If it is, then the dawn of coding is equivalent to the dawn of indefinite replication as well as the distinct decoupling of information transmission and metabolism. This would support the definition of living system given at the start of this paragraph. Unfortunately, as is often the case in biology, the answer is a bit fuzzy: the transition from RNA genes and enzymes to DNA code specializing in information transmission with protein enzymes taking care of the catalysis required for metabolism is not sharp, and the history of the transition is still muddy and incomplete.

The details of transition 3 as they are currently known are given by Maynard Smith and Szathmáry (1995, chapters 5, 6 and 7). As far as the evolution of the code itself goes, Maynard Smith and Szathmáry describe it using Orgel’s phrase: “like coming into focus of an obscure picture.” The evolution of the code was gradual, and it came to be more reliable and efficient through selection. Likewise, the replacement of RNA enzymes (ribozymes) by protein enzymes was gradual, and probably evolved in parallel
with the evolution of the code. Similar gradual evolution seems to apply to other aspects of transition 3. With the development of protocells (which I will not discuss), longer strands of RNA could be selected at the protocellular level. The fitness advantage comes through the linking of related genes, increasing reliability, but at the expense of some efficiency of replication, since linked genes take longer to reproduce. Other factors were no doubt important. In principle it is possible to gain increased stability through double stranded RNA, but the chemical stability of DNA, produced through the reduction of RNA, gave it a selective advantage (probably appearing first before the evolution of translation and protein enzymes). Its appearance is the final component of transition 3. The result is stable, stored reliably transmitted information that is reliably translatable into proteins. It should be noted that transition 3 has continued in minor detail through the selection of more efficient and reliable components. The transitions are not completely sequential, and the processes making up the transitions are gradual and evolve in parallel.

The general character of the transitions is that they involve greater complexity in how genetic information is translated and transmitted. This increased complexity increases the reliability, speed and/or efficiency of transmission and/or translation, and also opens up new regions of adaptive space that can be occupied. Each of these changes is produced accidentally and is then retained by relative fitness advantages, or so the story goes. The common aspects of the major transitions are shared between biotic and prebiotic evolution, so there is nothing particularly biological about them. Transition 3, however, appears to be a boundary (however fuzzy) between the prebiotic and biotic. If there is something special about biological information, this is where to look.

It from bit, intropic and level intropic substantive views of processes are all found outside of biology, so the relevant level for specifically biological information is the functional level. Instrumentally, the intropic and level intropic views are useful for describing the formation of self-organized structures, their replication and heredity in prebiotic evolution, as Maynard Smith and Szathmáry do in their book. This sort of process occurs in physics and chemistry as well, and continues into biological processes that support functionality. Maturana and Varela (1980) relegate the formation and support of autopoietic structures such as cells to such processes, with functional processes (operational processes) internal to the autopoietic system. They do not apply the notion of
Information to the internal perspective of autopoietic systems (at least for cells, perhaps not for higher level autopoietic systems), but argue that the concept of information is useful only from an external perspective. Thus, arguably (and it seems to be their considered opinion), information is useful only instrumentally for describing cells, even though they have a robust notion of functionality based in organizational or operational role. Despite their instrumental use of information, many of their followers talk of information internally, using the convenient connection from systems theory between the concept of operationality (function and control) and the concept of information somewhat unquestioningly. Maturana and Varela define autopoiesis in terms of operational closure, indicating that there is no information flow into or out of an autopoietic system, which I believe makes the information concept redundant in discussing the internal operations of the system. This, however, violates the general idea of openness central to most systems theory, and many of their followers have dropped the closure requirement, typically without comment, and allow autopoietic systems to be open to information. This undermines the justification for treating information as useful only for an external description of autopoietic systems, though of course it is still possible to treat information theoretic accounts as only of instrumental value. I have given principled reasons elsewhere (Collier 2002) for thinking that, contrary to Maturana, even cells are open to information, and that the self-organization process itself requires that the resulting system be open (Collier 2004a). Functionality depends on some sort of organizational closure, but it need not be complete and is not complete in biological cases (Collier 1999b, 2000, 2004b). These are based in well established principles of open systems theory (Ingarden et al 1997), so I will not go into more detail here.

The decoupling of energy and information budgets, which is a matter of degree, and increases through the major transitions, permits self-organization within the information system itself (Collier 1986, 2003, Brooks and Wiley 1988, Brooks et al 1989, Layzer 1990). The degree to which this occurs is presently unclear (it is an empirical matter), but it is a potential source of new organization and information within the information system itself, including within adaptive space itself (Layzer 1980, Collier 1998). This permits “minor transitions”, allowing gradual increases in the size of information space. These transitions, like the major transitions, are chance events, but are
favoured probabilistically. On the level intropic account they increase both information and entropy (disorder of the lower level) simultaneously, a phenomenon well known in physics (Landsberg 1984, Layzer 1990). Furthermore, as previously mentioned, there are other channels of heredity through the environment by way of niche construction (Odling-Smee et al 1996) and developmental systems more generally (Griffiths and Grey 1994, Oyama 2000). The interaction of multiple channels of information not only allows the influence of those other channels, but also sets up conditions favourable for further self-organization. Thus Maynard Smith and Szathmáry’s focus on DNA is questionable, as there may well be other informationgenic (or morphogenic, to use a less awkward word) processes in biology other than DNA selection based on the functionality of its phenotypic expression alone.

Setting this issue aside for the moment, I will focus on the complexities of the instrumental use of information in DNA in order to get clearer what is implied by its use in order to clear the ground for the discussion of the substantive use of information in biology more generally. It is worth noting, however, that the self-organization of information systems and developmental and environmental channels for the inheritance of information can all be understood with the resources of the intropic and level intropic accounts of information, without invocation of a substantive use of functional information tied to genetic information, even though it presupposes the partial decoupling of information and energy (but see below on storage). To understand functional information we need to look more closely at how gene expression and phenotypic selection works.

The route from information stored and transmitted from DNA to the phenotype of an organism is much more complex than the replication of genetic information in reproduction. Replication is fairly well understood, but gene expression, especially in multicellular eukaryotes, is very complex and not very direct. Fortunately, it is possible to avoid the complexities here. James MacLaurin (1998, see also Collier 1999a) has observed that substantive information has the nice property that if you wiggle something at one end of an information channel the result at the other end will reproduce the aspects of the information transmitted. This means that the complexities of gene expression, such
as those discussed by Evelyn Fox Keller (2000), can be ignored in the discussion here, no matter what their scientific interest otherwise.

Marcello Barbieri (2001) describes the “bridge between genes and organism” in two parts. The first part is transcription of DNA into primary transcripts. The second part, with eight steps, is epigenetic. The steps are shown in Table 2.

1. Splicing → Messengers
2. Translation → Polypeptides
3. Folding → Proteins
4. Protein Assembly → Organelles
5. Organelle Assembly → Cells
6. Cell Assembly → Tissues
7. Tissue Assembly → Organs
8. Organ Assembly → Organism

Table 2: Epigenesis (after Barbieri 2001)

These steps correspond to common phenotypic levels. There are various minor steps as well, such as the formation of control networks, membranes and their various inclusions, etc., as well as the back control of earlier processes. For example, opium is not composed of proteins, and its components must be produced under control of the appropriate tissue kind by networks composed of proteins and other substances. Similar observations could be made about behavior and other complex phenotypic states that are not the directly composed of proteins. The minor steps in general do not fit as well into a levels account. For example, networks involving hormones can extend across the whole organism, but their production and action is always local. I am ignoring these sorts of networks in this chapter, since they would need too much space to discuss, and would take me away from the focus on genetic information. Nonetheless, there are interesting issues concerning the extent to which these can be reduced to genetic information, the extent to which they can be analyzed as communication channels, and the way(s) in which they are committed to substantive information accounts. But these interesting issues must be set aside here in favor of the focus on genetic information.\footnote{Maynard Smith (2000a) suggests that enzymes should not be understood informationally, but that hormones should. This is in line with the usual notion that hormones signal, but enzymes merely facilitate}
changes, then we have genetic information expressed phenotypically. Knockout

Barwise and Seligman (1997) offer an account of information flow in distributed

systems that is widely applicable in biology as well as non-biological systems. The basic

idea is of an *infomorphism* based on regular relations between two sets of classifications

relating types to tokens. Networks of infomorphisms can be constructed to form

information channels that have the intuitively expected properties. However, regularities

are not sufficient for the purpose of information channels (accidental regularities do not

count for the transmission of information) so the Barwise and Seligman account is

already substantive to some degree, despite appearances. The use of types in the

formalization suggests that there must be an abstract aspect to infomorphisms, and thus

information channels, that is not substantive, but Barwise and Seligman prove that types

and tokens are logical duals in the formalization, so there is nothing to stop taking types

as substantive. For example, a type can be a filter that sorts out tokens of a particular

kind. Realizations of information channels must embody the types, which will be

something like a filter. Genetic information is expressed at various levels up to the

organism through such channels, and is combined and filtered to produce phenotypic

properties both complex and simple, subtle and coarse. The distributed network of

channels permits complex interactions to form the information in traits, and there can be

(and are) other sources of information that are not genetic (nor environmental, if self-

organization occurs within the organism). That such channels exist is beyond doubt; the

contributions to phenotypic form from genetic sources and the processes involved are less

well known. The existence of such channels from genes to phenotype will depend on

conditions within the organic environment (high acidity, for example, would make

expression impossible, blocking the first step – generally the contribution of the internal

environment will be more significant). However, despite this non-genetic dependency,

chemical reactions. Nonetheless there have been some attempts to treat enzyme action biosemiotically. The

idea is that they contain a message ‘carry out chemical reaction X here’. Given that enzymes are functional,

and are not merely chemicals that happen to be there, this idea is not as preposterous as it might at first seem.
the information expressed phenotypically through the cross-level channels is still genetic information. This is sufficient for enough genetic determination of the phenotype to be evolutionarily significant. This is quite aside from the issue of the extent to which the internal environment is itself genetically determined.

The issue of genetic determinism is fraught with confusion. As far as we are concerned with whether heredity or environment (nature or nurture) cause certain traits, we want to know their relative contribution to those traits. However, for evolutionary biology, the important issue is typically not the relative contribution, but whether there is any genetic contribution at all to differences in traits (Wilson 1988, pg. 8). The idea is that over evolutionary time environmental variation will be statistically insignificant (it will come out in the wash of time), and genetic differences will be all that matters to selection processes. Thus, if something is the product of selection, and thus more adaptive than its alternatives, it will be an adaptation (West-Eberhard 1992), and presumably will be functional on many standard accounts of biological function (Wright 1973, Millikan 1989, Neander 1991). If so, this would allow us to bridge the gap between genetic information expression and function, taking us to the next substantive level of information: functional information. I will discuss this further below, but a few more technical details are required first. The important point here is that the connection between genetic information and function need not deal with the complications of gene expression despite numerous recent criticisms of the idea of genetic determination in general, nor need it account for all aspects of phenotypic traits. Even in cases in which there are plausible emergent forms, such as has been claimed for the early development sea urchins by systems biologists who have studied in full the first 16 stages (Davidson et al 2002), genetic change can lead to developmental change (different attractors become more likely). Again we see the advantage of the information approach in that it can explain even across non-reducible levels.\(^\text{10}\)

So far I have kept mostly to the technical application of information theoretic methods to biology. However, in explaining their application, I have shown that at least substantive notions of information are required to make sense of the applications. The

\(^{10}\text{An analysis of the conditions required for this in terms of a formal account of levels and information across levels was given by Collier (2003). The basic ideas, let alone the details require much too much space to be recounted here.}
Information in Biological Systems – John Collier

substantive notions required so far, though, are not peculiarly biological in any way, until perhaps we come to the issue of adaptation. It is time to look in more detail at the peculiarly biological.

**Information storage and processing**

Transition 3 (Table 1), as discussed in the previous section, is the most significant, marking the transition from prebiotic to biotic evolution. At the very least, before transition 3 the processes are questionably biological, while after transition 3 they are clearly biological. Maynard Smith and Szathmáry describe the transition as being to a genetic code from RNA as both gene and enzyme, but I have described this in the previous section as a transition to decoupled information and energy budgets, with specialization for heredity and metabolism. The reason I described it this way in the previous section is that I did not want to presume there, like they do, that the notion of code, which strongly implies a robust sense of information, is required to make sense of the dynamical decoupling. I will set the issue of codes aside until the next section. In this section I will deal only with the implications of the decoupling for the explanatory role of information. Transition 3 also implies the decoupling of genotype and phenotype. Replication after the transition requires storage and transmission of information that is supported by, but largely unaffected by metabolic processes in any way specific to the information stored and transmitted. This information, then, is a candidate for specifically biological information. What is specifically biological about its nature, and why would we call it information?

Perhaps the most obvious reason to call the hereditary processes after transition 3 informational processes is that they involve storage and transmission. Gatlin (1972), for example, places great emphasis on these aspects of genetic information, and gives it no independent characterization, at least not explicitly. However attractive this idea might be, it can’t be right, because energy is also stored and transmitted in organisms. Almost exclusively, the vehicle for energy transmission and storage in organisms is ATP, but we

---

11 Godfrey-Smith (2000) argues that coding is not necessary, since proteins could be used to replicate themselves. True, but this would not allow open ended-evolution unless there were some sort of protein code allowing a finite number of proteins to map an open-ended range of possible products. However, Godfrey-Smith’s suggestion is consistent with decoupling without coding.
Information in Biological Systems – John Collier

are not inclined to call ATP information bearing. The reason, I think, is obvious: ATP is not discriminating; information is. A requirement for information to be discriminating is that its embodiment is complex. This is a direct consequence of information theory: the amount of information capacity of a channel can be no greater than the complexity of its embodiment.

While it is theoretically possible for information to be transferred without any net transfer of energy (Bennett 1973), this can occur only in fully conservative systems, so information transfer will typically also involve energy transfer. Why, then, would we want to refer to information rather than energy in certain biological processes, and especially in the focal case of this article, genetic information transmission? The answer has to do with guidance and control, at least, and possibly function and meaning (semantics) as well. I will deal with guidance and control in this section, and function and meaning in later sections.

Shannon (1949) observed that the notions of information and constraint are interchangeable. The Barwise and Seligman (1997) formalization of the idea of an information channel places the constraints in non-accidental regularities characterized as infomorphisms. These are grounded in classifications that have an embodiment in relations among tokens. However, many purely physical systems can be characterized in the same way12, so what, if anything, is peculiar to biology? The best answer available is that biological information channels typically show organized complexity (Collier and Hooker 1999). They are complex themselves, carry complex information, and are interconnected in complex ways that show considerable logical depth (Bennett 1985, see also a similar concept dubbed sophistication by Atlan 1972), indicating organization. In the physical sciences, boundary conditions are typically distinguished from laws governing dynamical transitions, which are regarded as peculiar to typical circumstances. In biology, however, the boundary conditions themselves have considerable regularity, and embody the special laws of the discipline (if any – if not, their closest analogue), or at least the foundation for those laws, such as Mendelian genetics and Fisher’s population genetics. This is another aspect of the dynamical decoupling of energy and information in

---

12 See (Collier 1999a) for a general characterization of causal connection that is a case of Barwise and Seligman channels with the classes restricted to dynamical kinds.
biological systems: the information system is free to form its own regularities, more or less free from any special restrictions of the boundary conditions on the energy budget. In evolutionary time, this has led to the production of more complex informational regularities of the sort described by Maynard Smith and Szathmáry as major transitions, as well as the minor transitions of Brooks and Wiley (1988). If we were to focus only on the energy budget, most of this organized complexity would be missed. For this reason it is at least inconvenient to reduce biological processes to energy governed ones; whether reduction also misrepresents biological processes requires more investigation.

Inasmuch as the regularities in boundary conditions and their interactions guide changes in the energy of a system, it is natural to refer to them controlling the system. In particular, the genes place boundary conditions on traits, and it is natural to say that they have some control of the traits that are produced epigenetically. Although the traits also have information about the genes, the relation is asymmetrical, since only some of the genetic information is expressed in the traits, and they are at best signs of the genetic information, and the genetic information is not an expression of the traits. The reason for this is that infomorphisms and the logic of distributed systems are not like standard logic: in general one cannot deduce from the knowledge that there is a channel from A to B that changes in the state of B will lead to corresponding changes in the state of A. For example, changing the dials in the control of a nuclear reactor that indicate it is out of control to indicate that it is in control will not put the reactor back into control. At best it will break the channel. However, changing the conditions in the reactor by using the reverse control channel has some hope that the dials will correctly indicate the reactor is back in control. In this sense the genes control the traits, but not vice versa.

There are two reasons to reject the idea that genes control traits: 1) control might be regarded as requiring function, but function is not required for explaining gene expression, and 2) control might be regarded as requiring intention, which adds to function some sort of meaningful purpose, but genes are not intentional. This brings us back to the issue of whether the functional kind of substantive information in Figure 1 needs to be invoked in biology, and also raises the issue of intentionality, meaning and semantics. I will address these issues later. Before that I want to address the issue of
coding, a notion Maynard Smith and Szathmáry use in their characterization of transition 3, but which I deliberately set aside in this section.

**Codes**

In the previous section it has been established that the decoupling of metabolism and information required by Transition 3 implies information channels, especially channels from DNA to phenotypic traits. These channels are grounded in classifications grounded in processes that show a regular and organized structure. The regularities are both essential and sufficient for the existence of such channels. Why would Maynard Smith and Szathmáry also require that DNA be a code? It has two major implications for inheritance and variability, required for evolution. Before the discrete character of genes was understood, objections were raised to Darwin’s theory of evolution by selection that sexual reproduction would lead to a mixing of genomes, and a tendency to converge towards some intermediate state, which is not what we observe, and not what we need for the origin of species (divergence). The discrete character of genes resolved this problem. It also permits recombination in the reproductive process, and recombination is known to be more effective in creating variable phenotypes than mutations alone. Of course both of these were the result of innovations later than Transition 3 itself, which applies to nonsexual bacteria. Since evolution by selection is not anticipative, these advantages could not have underlain Transition 3, however useful they were in later transitions.

One clear advantage of a code is the reduction of ambiguity in the regularities underlying the information channels involved in gene expression. This leads to an increase in fidelity of reproduction even in nonsexual organisms. The discrete character of changes in the genome is also advantageous even to even to nonsexual reproducers in that it introduces a degree of modularity into genotypic and resulting phenotypic changes. This modularity presumably makes it easier to change some parts of an organism without changing others. This conservatism is more likely to lead to mutated organisms that are viable. Thus a genetic code has an immediate advantage for even nonsexual organisms, as well as opening up the possibility of later major transitions. It is difficult to see how a more holistic form of heredity could be equally successful, but it is also difficult to rule out the possibility a priori. It is safe to say, however, that if such a mechanism had
evolved and become dominant, the evolution of sexual reproduction would not have occurred.

What is a code? Barbieri (2001) argues that codes 1) are rules of correspondence between two independent worlds, 2) give meanings to informational structures and 3) are collective rules that do not depend on individual features of their support structures. The independence in this case is grounded in dynamic decoupling in which at least one of the “worlds” is informational in the sense of the last section. It is not necessary that both “worlds” be informational in this sense, since it is possible for information to be expressed in a non-informational structure or process. Strictly speaking, requirement (3) does not imply discreteness, but discreteness at least greatly enhances the possibility of both collectivity and independence of support structures, as well as the efficiency of the code in the sense of using the same parts in different combinations to express different information. Collier (1986) introduced the notion of a physical information system that requires some degree of discreteness, but the advantages, all other things considered, are increased with the degree of discreteness. Such systems satisfy the storage and transmission requirements of the previous section, as well as Barbieri’s requirements (1) and (3). The most controversial requirement is the second one, which will be discussed later.

Is the code concept required in biology? Although it is possible to regard hereditary and expression processes entirely in terms of energetic transformations, as in the first section, taking a non-substantive or very weak substantive view of information, genetic information behaves so as to satisfy the requirements of a physical information system, so something more is going on than just transformations of energy. To miss this point is to miss the special character of biological information, not to mention belying the way in which accomplished biologists like Maynard Smith and Szathmáry talk about the systems they study. Although reductionists might argue that such talk is unnecessary, their position is based on a metaphysical view that need not hold, and probably does not hold for gene expression (recall Davidson et al 2002), and possibly for the other forms of biological codes that Barbieri discusses. The reductionist position is thus metaphysically dubious, factually inadequate, and flies in the face of the way experts talk.
Barbieri (2001) points out that the bridge between genes and proteins has one genetic step and at least four levels of epigenetic processes. The first is widely regarded by biologists as a codified assembly, but the epigenetic processes are typically regarded as catalyzed assembly. The difference is between the sort of processes found prior to Transition 3 and those found after the transition. Barbieri argues, however, that the assumption that epigenetic processes are of the older catalyzed form, with no clear distinction between metabolic informational processes, has not been proven. Much of his book takes up an argument that many epigenetic processes are also processes of codified assembly. He provides rich evidence from molecular biology that splicing at step 2, translation at step 3, signal transduction to form organelles, cells and tissues are also codified. Given the advantages of codification described in this section, perhaps this should not be very surprising. It is less easy to accept, however, that the codes require anything more than a syntax (the rules), but Barbieri also argues for a semantics or meaning (condition 2).

**Information and meaning**

The idea that meaning or semantics is required for fully understanding biological information has been attractive, but also highly controversial. I confess that in my own work so far I have had no need of the idea, but I have taken what I think is a very cautious approach to meaning, and perhaps I have not yet encountered the sorts of problems that require the hypothesis of meaning for their solution. Barbieri trumps his book as a revolutionary manifesto for what he calls “semantic biology”. Whether or not he is successful in this, he shows fairly convincingly that the code concept (in its restricted, non-semantic physical information system form) applies far more broadly than is generally accepted.

One thing that is generally agreed is that meaning requires function. A common further requirement is intentionality. Matthen and Levy (xx) have argued for intentionality in the immune system, but to the best of my knowledge there are no other arguments for intentionality within biological systems except for the mind. On the other hand, *teleosemantics* (Millikan 1987, Mathen xx) argues for continuity between selection processes and semantic representation, which suggests at least the possibility of non-
mental intentionality, but the idea has not been developed beyond the immune system. Barbieri (2001) seems to take it for granted that a code implies a semantics, but this must be non-intentional, if intentionality is peculiar to the mind and perhaps a few specialized systems like the immune system. I will discuss this possibility in the remainder of this chapter, with a discussion first of function, representation and biosemiotics. This discussion will be necessarily cursory, since a complete discussion would be worthy of a book (or several), and almost all of the main aspects are highly controversial.

The standard account of function used in biology is the etiological account (Wright 1973, Millikan 1989, Neander 1991). According to this account, a trait is functional if it is selected for, meaning that it is an adaptation. On this account, a selected trait will contain information produced by certain genes that are selected along with the trait. Thus, it is sometimes said, the selected genes have information about the environment, as well as about the traits that they express information in. The problem with this account is that meaning is an asymmetrical relationship on most accounts (like control, incidentally). I have already argued that this presents no special problem for the information in traits, but for environmental features the problem is not so easily solved. On the etiological account, it is certainly arguable that since the genes are (indirectly) selected, they are under the control of the environment. Thus the genes don’t have the right relation to the environment to have information about it in the same way that they might have information about traits. At best, the genes could be signs of environmental features, not meaningful representations. It all depends on which way the channel goes. One can’t have it that both that the genes mean the traits and that they mean the environmental features that selected them. To put it another way, genes are sometimes described as blueprints for the organism, but if this is fair then the etiological account implies that the environment is a blueprint for the genes. Something has gone wrong. 13 Perhaps it is the idea of meaning here, but perhaps it is the etiological view of function.

13 The symmetry problem has been pointed out in one form or another by Sarkar (2000), Sterelny (2000) and Winnie (2000). Winnie gives a solution very close to the one I propose. Sterelny and Godfrey-Smith (2000) are also concerned with symmetries between genes and their contexts. I have already argued that the contexts serve the role of channels. Of course there might be regulatory genes that serve a regulation function and code for channel construction. These issues are complex, and need to be untangled, but they do not seem to me to present any special difficulties if we have a suitable account of function that breaks the symmetries.
The etiological account is just seems wrong in several very obvious ways. Jeff Foss (1994) has noted that we can typically assign function without knowing anything about etiology (though etiologists will argue that this is often fallible). Alternative accounts of function focus on organizational role (Maturana and Varela 1980, Rosen 1991, Cummins 1983), with selection and resulting adaptation being explained in terms of differences in functionality rather than defining function. On this account the function of genes is heredity and the guidance of ontogeny. Selection acts on the results as a sort of filter, creating a channel guiding the gene pool to greater fitness (see, especially, Winnie 2000 for a helpful account). The representational role, if any, remains always in the genes. Perhaps this is a reason to reject the etiological view in favor of the organizational role view of function, but it depends on the cogency of the genes representing. It should be noted that, *mutatis mutandis*, similar problems can be raised for teleosemantics.

Representation typically requires some sort of system of rules that does not depend on their underlying substratum. Physical information systems (or Barbieri codes minus the meaning) have these. What more is required? The usual answer would be *interpretation*. Without interpretation, a representation is useless, non-functional. This suggests that we should look for some sort of interpretation in biological systems if we wish to find meaning. On standard accounts of meaning, the interpretation is given by the semantics, which is an abstract relation between symbols and their reference. This will not do for biological representation, however, since the relation has to be embodied in concrete biological processes. In order to correct this deficit, Bend-Olaf Küppers (1990) suggested that we include, along with the syntax (rules) and semantics (reference) of the genes a pragmatics. His view at that time was that the pragmatics was given by selection, but we have seen the problems with *that* view. Küppers has told me since that he has abandoned his earlier view, but the move towards pragmatics is a good one. How do we get a satisfactory biological pragmatics (if we can)?

**Biosemiotics**

Biosemiotics is an attempt to apply semiotics to biological systems in order to account for communication from the molecular level through the behavioral and social
levels. The dominant approach today is the Copenhagen school (e.g., Hoffmeyer 1996), which takes the semiotics of C.S. Peirce as its starting point. Again, since there are many controversies involved that would take much space to represent, let alone resolve, I will be brief. Peirce believed that pragmatic issues were the basis of meaning, in particular what expectations about the world are attached to a given idea in a way that might guide action. On Peirce’s full-fledged view signs are an irreducible whole of three parts, one what we would normally call the symbol, the object (which corresponds roughly to the intensional reference), and the interpretant. This whole is embedded in a system of interpretance that connects to expectations and actions. If these ideas are to be applied to biological systems, the interpretant has to be within the organism, or more accurately, within the relevant biological system. He considers the sunflower, whose flower tends to face the sun very reliably. The direction the flower faces, then, is a good sign of the direction of the sun. However it is not a sign for the sunflower, since there is nothing in the sunflower that makes use of the information in the sign. The effect is a tropism caused by the size of the sunflower, its rapid growth, and the induction of growth inhibiting hormones by sunlight. The direction of the flower itself plays no functional role for the sunflower. Peirce did not know the explanation, but inferred correctly that the direction the flower faces was not a sign for the sunflower. On the other hand, he did not rule out that there could be genuine biological signs.

If we consider DNA as a sign of (at least some aspects of) traits, we need to find an appropriate interpretant within the organism to complete the trinity. As described early in this chapter, genetic information is expressed if differences in the genes make a difference to the traits expressed, no matter how small. This expression is functional on either the organizational role or the etiological accounts. The best candidate for the interpretant in this case is the other coding and catalytic processes involved in epigenesis. If this idea can be made out coherently, then there is a good case that DNA contains information about the phenotype of the organism for the organism itself, rather than from merely external view of some anthropomorphizing observer. And this meaning would be about in the semantic sense, with epigenesis providing the pragmatics. John Winnie (2000), though not in the biosemiotic tradition, suggests that parameters whose effects on the components of the system contribute to the likely performance of the system exhausts
the “semantic” aspect. In a living system the performance is the contribution to viability of the system, which is subject to selection. This is not very different from Peirce’s idea.

Summary and conclusions

The first section showed how information theory can be used descriptively in biology in the case of the genes. This descriptive use is also explanatory to some extent, and invokes substantive information, but in no way that is specifically biological. However, the talk of biologists, and the distinction created by Transition 3, suggests that biological information involves something more than this. It is relatively easy to introduce notions of transmission, control and guidance as substantive, somewhat less easy to convincingly introduce a need for the substantive use of information codes, and much less easy at this time to justify substantive notions of meaning and semantics, though biosemiotics is highly suggestive.

Much of what has been said in this chapter about genetic information applies, mutatis mutandis, to other forms of biological information, such as molecular communication, communication in the nervous system, immune system, hormones, pheromones, and behavioral transmission between organisms. There are special aspects of each case, but most of the arguments justifying the use of information concepts in a substantive way carry through to these cases. I hope that the portrait I have given of genetic information is helpful in extending the ideas to other cases.

References


Information in Biological Systems – John Collier


http://www.library.utoronto.ca/see/pages/SEED_Journal.html

http://www.library.utoronto.ca/see/pages/SEED_Journal.html


Information in Biological Systems – John Collier


Information in Biological Systems – John Collier


Figure 1. Nesting of major kinds of information