

# Information in Biology: A Fictionalist Account<sup>1</sup>

Arnon Levy

Department of Philosophy, Harvard University

FORTHCOMING IN *NOÛS*

## 1. Introduction

A striking feature of contemporary biology is the prominence of semantic concepts, especially concepts that have to do with information and communication. Biologists speak of codes and signals, transcription and translation, even of editing and proof-reading. Such notions have a home in descriptions of interactions among intentional agents, but they are nowadays routinely invoked in descriptions of metabolism, physiology and development. Recent decades have seen great strides in our understanding of the chemical and physical basis of life, but the tendency to view biological processes through an informational lens has anything but subsided. This is a puzzling situation: What is biological information? What does it mean to say that cells communicate or that molecules are signals? Does the appeal to informational notions have a genuine theoretical role in biology, and if so what is it?

---

<sup>1</sup> I wish to pay special thanks to Peter Godfrey-Smith for many discussions of information in biology, and for reading a number of earlier versions of this paper. I have greatly benefited from exchanges with Yemima Ben-Menachem, Liz Camp, David Haig, Ned Hall, Daniela Helbig, Nick Shea and Michael Weisberg and from written comments from two anonymous referees for *Noûs*.

Increasingly over the past decade or so, philosophers and philosophically-minded biologists have discussed these questions (Godfrey-Smith & Sterelny, 2008, provide a recent review). For the most part, debate has centered on the role of inherited information in developmental biology – the common but controversial idea that genes code for or supply information for the making of an adult. Information and related notions are certainly prominent in developmental genetics, but appeals to such notions occur regularly in other parts of biology. The role of non-genetic factors in development, the activity of hormones in adults, the exchange of factors between nearby cells – I discuss these examples below – as well as a variety of other cell-level processes are typically explained in terms of signaling and communication.

The central goal of this paper is to provide a general view of informational notions in biology, one that takes into account their full range of application. The account I'll develop is a fictionalism of sorts, but one that is intended to illuminate the function of informational concepts in biology rather than deflating them. On the one hand I claim that given the range and character of appeals to informational notions, the most plausible interpretation is that they are fictional – metaphors rather than descriptions that are grounded in genuine semantic properties of cells and macromolecules. However, I will also argue that appeals to information bear theoretical weight by allowing us to reason via a fiction about real causal properties. On this view, invocations of information in biology are not literal descriptions but they are not rhetorical flourishes either. They play a genuine role in biological understanding.

One important feature of informational language in biology is that its fictional status often passed undetected. Informational language is what we might call a *liminal* metaphor – one that operates near the threshold of the noticeable. Liminality is not an uncommon feature of metaphorical language, especially in science. It can have utility when the goal of metaphorical description is not primarily aesthetic. By backgrounding the fact that one is not talking literally one gains a better grip on the insights provided by a metaphor. Nevertheless, I suspect that liminality accounts for some of the confusion surrounding

information, and for mistaken interpretations of molecular biology, on which I comment at the end of the paper. In this regard, I shall attempt to walk a tight line between taking information seriously as playing a real theoretical role and refraining from taking it too seriously by assigning it a foundational status in biology.

Such a stance raises questions about the status of non-literal devices in science more generally. While a direct engagement with these questions goes beyond the scope of the present discussion, I think it is an important and largely neglected fact that there *is* a spectrum of literality in scientific discourse, and that a concept's position on this spectrum matters. A secondary goal of this paper is to call attention to a class of scientific concepts that resides somewhere in between the fully literal and the merely ornamental.

I begin, in the next section, by reviewing the main themes of the existing debate over information. Section 3 describes some cases of informational explanations outside of genetics. In sections 4 and 5, I argue that existing accounts are unable to handle these cases, and I motivate a fiction-based view. Sections 6 and 7 describe the account I favor in more detail. In closing, I address ways of taking information too seriously, and connect the case of information in biology to the more general issue of the role of fictions in science.

## **2. The State of the Debate**

In thinking about development, professionals and laymen alike commonly treat genes as providing the information for adult form. In contrast, other factors are usually thought of as raw materials or as background conditions. Susan Oyama's book *The Ontogeny of Information* (1985), prompted a number of philosophers to argue that this dichotomy is untenable and that there is no distinct sense in which genes

carry developmental information. There seems to be a general agreement that genetic coding, the mapping of DNA base triplets onto the amino acids that constitute proteins, is a legitimate and important theoretical concept. The critics' claim is that a richer, semantic sense of genetic information cannot be rescued from biological usage, nor is it necessary for explaining development. Indeed critics generally think that appeals to information are detrimental. Sarkar (1996) and Griffiths (2001) have argued that lack of care in the use of informational concepts leads to widespread misunderstandings of the explanatory structure of molecular biology: It encourages the belief that phenotypes can be “read-off” genes in a bottom-up manner. Furthermore, some hold that attributing an informational role to genes lends spurious support to genetic determinism (Griffiths, 2006).

There exists a highly developed mathematical theory of information, pioneered by Claude Shannon (1948). But most authors who have written on the topic agree that Shannon's notion is not the relevant one here. It is worth recounting – in a brief and non-technical way – why. Shannon's theory allows that anything can be a source of information so long as it has a range of distinct states. One state carries information about another provided that the two are correlated. Information theorists then say that the two states are connected by a *channel* along which signals are transmitted. Intuitively, an information channel allows the receiver to learn about the state of the sender by consulting the signal. Information theory provides quantitative measures for the amount of information contained in a signal, the capacity of channels, the efficiency of particular coding and transmission schemes and related matters. These tools can be useful in biology, especially in bioinformatics, where large amounts of data pertaining to genes and proteins are analyzed. But in these contexts the apparatus of information theory is used as data analysis tool: Bioinformaticians treat an available data set as carrying information about some process or structure of interest – the structure of a protein, for instance, or the topology of a regulatory network. They are not using information in an explanatory account, as a way of saying what genes do or how they do it. And for good reason. Genes carry Shannon information but so does any

other factor that reliably affects protein structure (such as temperature). Genes may carry more information, but if information is understood along Shannon lines, their role isn't qualitatively different. In contrast, the use of informational terminology that is under debate is meant to distinguish genes from other developmental factors. Genes are said to carry developmental information whereas food, also a crucial ingredient in development, doesn't. In accounting for this explanatory use Shannon's notion is of no avail.<sup>2</sup>

Accepting that Shannon-information isn't the way to go, "advocates" of information have mostly opted for a teleosemantic account (Sterelny, Smith and Dickison, 1996; Maynard Smith 2000; Sterelny 2000; Shea 2007). This view seeks to ground inherited information in natural selection. It relies on the idea that we can think of products of natural selection as having a function, and furthermore that under certain conditions functions can ground ascriptions of content. On the teleosemantic view, genes carry information about (or for) the structure they encode in virtue of being selected for producing that structure.<sup>3</sup>

Of course, many biological structures have a function but are not information carriers – wings are for flying but they do not carry information about flying. So some further element has to be added to this picture. Maynard Smith (2000) and Sterelny (2000) argue that informational factors are those selected structures that exhibit an arbitrary relation to their effects. Modeled after the relation between words and their referents, the idea is that to have genuine semantic content the structure of a putative signal must be largely unconstrained by the structure of the message; just as the structure of 'table' is largely unconstrained by the structure of tables. If the structure of the 'message' is too closely connected to the structure of the 'signal,' we have chemistry or physics and not semantics. It is hard to make this

---

<sup>2</sup> Information theory is put to use in some parts of neuroscience, especially in computational neuroscience (Nelson 2007; Rieke et al. 1996) but these uses are beyond the purview of the present discussion.

<sup>3</sup> The teleosemantic view can be seen as an instance of a more general teleosemantic theory of content, as developed by Millikan (1984) and others. But most authors in the present debate have not explicitly drawn on the more general framework.

notion of arbitrariness sufficiently clear. One worry is that any effect may seem arbitrary if there are enough causal links separating it from some salient cause (Godfrey-Smith, 2000). Another is that the kind of contingency that is supposed to ground arbitrariness (especially in the paradigm case of the genetic code) isn't essentially different from the historical contingency characteristic of any product of natural selection (Stegmann, 2004).

Nicholas Shea (2007) has suggested a more complex teleosemantic view which shifts the locus of function to the level of the inheritance system as a whole – in the case of genes, to the function of the system of genetic inheritance. Shea argues that informational factors in development are those factors that, on an evolutionary time-scale, carry Shannon-information about their outcomes *and* are elements of an evolved system of inheritance. An inheritance system is an evolved structure consisting of a sender, a consumer and a range of intermediates that coordinate the states of sender and consumer. In such a system intermediates carry (at least a rudimentary form of) content, and it can give rise to ascriptions of information. The heart is not part of an inheritance system so, like most evolved structures, it isn't an information carrier. Genes, on the other hand, as well as some kinds of non-genetic factors such as chromatin marks (Jablonka and Lamb, 1995) are intermediates in an evolved system of inheritance, so we can treat them as carriers of inherited information.

It appears that the teleosemantic view, in one version or another, is the most promising option on the table. But I'll argue it cannot be the whole or even the main story about information in biology. We have seen that the teleosemantic view arises in the context of genes and development. As I've noted, informational notions occur in many other areas of biology. Paying attention to this broader range of cases tells against the teleosemantic account. I'll explain why in section 4; first let me describe a few key examples.

### 3. Information Outside of Genetics

The first example comes from developmental biology, but is not directly related to genes. It concerns morphogens and “positional information” (Wolpert, 2006). Many animals exhibit a basic division of the body into axes: dorsoventral (back-belly), anterior-posterior (mouth-anus) and lateral (left-right). Axis formation is a very early step in the assignment of spatial identities to the body’s different parts. A basic question facing developmental biologists, here as in other early pattern formation processes, is: How do different parts of the embryo differentiate in a location-specific manner? Or, as it is sometimes put: How do embryonic cells know where they are? One important mechanism involves the delivery of positional information by morphogen gradients. A morphogen is typically a transcription factor. Its impact, the particular set of developmental genes that it regulates, depends on its concentration in the immediate vicinity of the cell or within it. In axis specification, morphogens act by forming a concentration gradient along the relevant axis – e.g., highest near the dorsal end and decreasing ventrally. As different genes get transcribed in response to different concentrations of the morphogen, cells along the gradient develop into distinct types and an axis is formed.<sup>4</sup> Thus, cells know where they are because they are sensitive to morphogen concentrations, and these map onto location. This important developmental mechanism is typically described by saying that morphogens are signals that supply embryonic cells with *positional information*.

Next, consider hormones. Some hormones act in development, but many don't. In explanations of hormonal activity the language of information and communication is ubiquitous. Hormones are practically defined as signals that coordinate the activity of non-neighboring cells. Typically they are produced by a gland somewhere in the body, say in the brain or in the pancreas, then secreted into the

---

<sup>4</sup> This is perhaps the simplest case. Complex gradients, and interactions between the gradient and cell-level activities are also important. In addition, responders are sometimes nuclei within a multi-nucleated cell (syncytium) – a common situation early in insect development – but the principle is the same.

blood stream. Hormones bind to receptors at the target cell and exert their effects, typically metabolic regulation, either directly or via modulating transcription. The action of these receptors is often described in terms of “perception” or “interpretation”.<sup>5</sup> Once bound to a receptor, a hormone activates a chain of reactions that regulates metabolism either directly or via “signal transduction cascades” that affect transcription.

The last example concerns local cell-cell communication. Gap junctions are molecular ring-shape structures that are embedded in membranes, creating a narrow fluid-filled channel that connects the cytoplasm of neighboring cells (Evans et al., 2002). Gap junctions enable ions and other small molecules to diffuse between adjacent cells. Or, as this is often described, they allow signals to travel between cells, and information to be shared or transferred (Alberts et al. 2002, 835). Note that gap junctions are too narrow for “sequence” molecules such as RNA, enzymes or structural proteins to move through them; the signals are typically small charged species such as ions or peptides. Gap junctions play a part in coordinating the activities of cells within a tissue by allowing rapid transmission of non-genetic signals.

Though my description of these examples has been brief, I think it enables us to point to two important features of the language of communication and information. First, it is ubiquitous, occurring in various parts of cellular and molecular biology. DNA and the role of genes in ontogeny are described in informational terms, but so are phenomena quite far removed from genes or development. Second, information is typically invoked in the course of offering *mechanistic explanations*. A precise notion of mechanistic explanation is not required here (see, e.g. Machamer et al., 2000); suffice it to note that information and related intentional notions usually figure in explanations that show why a certain organ

---

<sup>5</sup> For instance, in a recent review of hormonal regulation of development in plants, Chow and McCourt state that to understand the role of simple organic molecules that serve as hormonal signals in plants, “first requires an understanding of how they are perceived” (2006, 1998).



or cellular structure exhibits a certain property or behavior by describing its internal organization and the ordered interaction among its constituent parts. Mechanistic explanation is proximal, i.e. the explanans are non-historical facts that have to do with the causal relations between elements of the system being explained – it is an explanation that appeals to what’s “under the hood.” As we shall see in the next section, both the ubiquity of informational language and the fact that it is typically employed in mechanistic explanations are directly relevant to an evaluation of the teleosemantic view.

#### **4. The Inadequacy of the Teleosemantic View**

Shannon’s notion will not, as noted, explain the role of informational language in biology. Nor, of course, will a simple reference to function: It won’t distinguish hormones or genes from other evolved body parts. This is what drives more sophisticated teleosemantic accounts to posit further conditions. The result is to narrow down the list of potential informational factors. But the sophisticated accounts are motivated by cases of inheritance, and it is doubtful that they can be made to fit the wider set of cases in which informational descriptions are employed. Shea’s account requires that the informational factors be part of a system of inheritance. But while some morphogens are maternally inherited, others are not. Hormones are not typically inherited, and neither are gap-junction-mediated signals. In terms of evolved functions, these factors seem more like the heart than the genome. If we turn instead to Sterelny’s and Maynard Smith’s notion of arbitrariness – which rests on shaky grounds anyway – it is unclear whether it applies to the relation between signaling molecules and receptors, often a subtle structural match that underlies a specific biomechanical transformation in the receptor. And one may worry that the appearance of an arbitrary connection between, say, insulin and the message it carries (roughly: increase glucose metabolism) is an artifact of the causal distance – the number of links in the causal chain separating them. Nor does it seem that the relation between a flow of ions through a gap

junction, and a resultant change in a cell's electrical potential, meets the arbitrariness criterion. This is unsurprising, as these do not appear to be the kinds of cases Sterelny or Maynard Smith had in mind – they were generalizing from features of DNA.

Another, deeper, source of trouble for the teleosemantic view has to do with the way it portrays the *type* of explanations in which information figures. On the teleosemantic view informational notions occur in functional explanations – which proceed by appeal to what informational factors are *for*. By ‘function’ is meant evolved function, i.e. the effect for which the factor in question was selected. Importantly, such functional explanations are distinct from proximal-mechanistic explanations. Shea is explicit on this point. He contrasts the role of information carriers according to the teleosemantic account with “stage-by-stage description[s] of the causal processes” underlying the phenomenon in question (2007, 317). Thus, if the why question addressed by appeals to information is something like: ‘Why does gene g give rise to protein p?,’ and the answer given by biologists is that g carries the information for p, then on the teleosemantic account this is a way of saying (roughly speaking) that g was selected for giving rise to p.<sup>6</sup> Similarly, in the case of morphogens the explanation for why a morphogen m specifies the dorsoventral axis is said to be that it carries positional information which is interpreted by cells. On the teleosemantic account this is a way of saying that m's selected function is to cause dorsoventral differentiation. Now, it is very likely true, at least in many cases, that genes and morphogens have evolved functions. But it appears that biologists do not use informational notions to describe these functions. For the most part, “stage-by-stage” mechanistic descriptions are exactly where one finds appeals to communication and information. Genes give rise to proteins *in the here and now* by supplying information in the form of a coding sequence – this is a way of pointing to the properties of

---

<sup>6</sup> On Shea's more subtle account this is a way of saying that g is an intermediate in a system of inheritance that carries some relevant bit of Shannon-information. The basic point I am making, however, holds for this account too. What makes something an inheritance system is its evolved function. Furthermore, the signal carries Shannon-information in virtue of correlations that exists on an evolutionary time scale, as Shea emphasizes.

genes in virtue of which they play a key part in producing proteins. In saying that morphogens supply positional information to embryonic cells, developmental biologists take themselves to be describing the mechanism of axis formation (Wolpert et al., 2002, a widely-used developmental biology textbook, is very explicit on this point. See pp. 19-20). Furthermore, the evidence sought to confirm that a particular morphogen is a positional signal – e.g. that its absence results in abnormal axis formation – also suggests that a proximal mechanistic explanation, rather than a distal selection-based one, is at work. Thus, a central idea behind teleosemantic accounts, the idea that informational descriptions explain via implicit reference to evolved function, seems to ill-match the role information plays in many actual cases.

Taken together, I think these arguments cast serious doubt on the teleosemantic view.

## **5. Motivating a Fiction-Based Account**

There are at least three motivations for thinking that informational language in biology should not be treated literally. The first lies in the fact that a literal understanding of information does not seem forthcoming. Shannon-information is not relevant to the problem, whereas the favored view in the field, the teleosemantic view, loses much of its plausibility when the full range and mechanistic character of informational discourse comes into view. Of course, it is possible that a different literal account will eventually be offered. But as matters stand this does not seem likely.

Second, as Alexander Rosenberg (1986, 2006) has noted, ascriptions of information in biology appear, at least in some cases, to differ from ordinary intensional contexts in that they are not opaque: They allow truth-preserving substitution of co-referring terms. Rosenberg makes this claim as part of a defense of a reductionist, gene-centered view of development. Questions about reductionism are largely orthogonal to the present discussion, but I think Rosenberg's observation can be recruited for my

purposes as well. It suggests that the description of genes (and possibly other factors) as content-bearing is a thin one, and not a full-blown ascription of intentional properties.

The third motivation is textual: The wording of explanations invoking information often suggests that information is used non-literally. As an illustration, consider the following quote from Alberts et al., the “king” of cell biology textbooks. The authors review the activity of Dorsal, a morphogen that specifies the dorsoventral axis in *Drosophila*. They sum-up by stating:

Thus, the regulatory DNA can be said to *interpret* the positional signal provided by the Dorsal protein gradient, so as to define a dorsoventral series of territories – distinctive bands of cells that run the length of the embryo (2004, 1184; italics in the original).

Notice how informational language is hedged by the use of italics, and by the phrase “...can be said to.” Such a tone is not uncommon, although rarely is a non-literal interpretation made explicit. This strongly suggests that something akin to a metaphorical mode of description is being employed.

This situation is not surprising – taking information talk to be non-literal is, in a way, the most natural way to take it. After all, why attribute to submicroscopic molecular structures the ability to mean anything, let alone to send and receive messages? The puzzle arises because biologists commonly talk this way, and appear to be talking seriously. Both observations, I think, are true: Information-talk is serious but it isn’t literally true. On the account I’ll offer, explanations involving information operate indirectly via a fiction. The activity of genes, hormones, morphogens and other factors is described *as if* it were a process of communication in which a sender transmits a signal that regulates a receiver’s behavior. Informational descriptions are telling us something about the causal role that genes, for instance, or morphogens, play. But they do so not by attributing full-blown intentional and semantic

properties to cells and molecules, but by using a schema associated with information and its communication (in their ordinary senses) so as to bring to the fore coarse-grained causal properties of the processes in question.

Since it is used in a fictional mode, the decision to apply an informational description is not a forced one: Genes or hormones are *not* objectively signals. However, the causal features which informational language provides access to are not fictional. The following sections will flesh out this fictionalist view in several steps.

## 6. The Pretense Theory of Fiction

An attractive way of understanding fiction and metaphor is found in Kendall Walton's pretense theory of fiction (Walton, 1990). Walton's theory is primarily aimed at explaining representational art, but it can be extended and applied to cover other fiction-involving uses of language such as metaphor and idiom (Walton, 1993; Egan, 2008). For present purposes I will not attend to the distinction between metaphor and related kinds of fiction (and I'll use the terms interchangeably). All we need is a way of understanding how a non-literal use of language allows one to track real, true facts.

The pretense theory models fiction on games of make-believe of the sort played by children – directed acts of the imagination, often coordinated among several individuals. In such games there is a set of rules – often implicit – that determines what is to count as true in the game. Walton calls these ‘principles of generation.’ Principles of generation instruct participants what it is that they are to pretend when playing the game. In particular, principles of generation often specify what one is to imagine in response to certain *non-fictional* facts. In the game of cops and robbers, for instance, there is a principle of generation that implies that if I point my index finger at you, my thumb sticking out and the

remaining fingers folded, while exclaiming “bang bang”, then, fictionally, I have shot you. If we are pretending that clouds are animals, then the shape of a cloud in the sky might make it the case that a reindeer is approaching. Thus, what’s true in the fiction may depend, in quite particular ways, on facts in the non-fictional world because the relevant principles of generation specify what participants in the game are to imagine in response to such facts.

One important upshot of this is that one can learn real-world facts by consulting fictions. If we are imagining that tree stumps are bears and you say “wow, check out that huge bear,” I can infer, without bothering to look, that there is a large tree stump around. Some metaphors work this way too: I may describe Italy as a boot in order to indicate that Lecce is on the heel. Or, to choose an example closer to the present discussion, I might think of an organism’s phenotype as lying in a fitness valley in order to reason about available evolutionary possibilities. In such cases, the fiction is constructed so that what is fictionally true corresponds to what is true simpliciter. Since the fiction employs a familiar set-up that makes it easier to handle, we use an indirect route and make fictional statements as a way of reasoning about the real world. To use a tracking metaphor, we use a fictional set-up to track non-fictional truths.

Information in biology, I propose, is a pretense of exactly this kind. Biologists metaphorically describe molecules and cells as engaged in communication and information sharing. Such descriptions invoke games of make believe in which participants are to imagine the relevant elements – genes, hormones, cells or whole organs – as if they were sending and receiving messages. Correctly read, this is a way of saying what these elements really are doing. Moreover, applying an informational schema is a way of organizing the causal facts and highlighting particular aspects of a process that have explanatory significance. The pretense is valuable because it enables one to compactly describe and reason about these explanatory facts. The next section will explain which types of causal facts I have in mind and how an informational description organizes them.

## 7. The Theoretical Role of Information

There is a generic kind of explanatory context in biology where one wishes to understand how the activity of one cell or structure is regulated by a distinct cell or structure, typically in a complex and adaptive manner. It is in these regulation contexts that one often finds informational language. For instance, many biology texts describe (up)regulation of glucose metabolism in something like the following way. Insulin is a signal that originates in the pancreas by a mechanism that is sensitive to increases in the concentration of glucose; it informs muscle cells that the level of glucose in the blood is high, and directs them to up-regulate the breakdown of glucose in response. On the present account this is a way of describing the coarse-grained causal structure of hormonal action by means of a metaphor according to which pancreatic  $\beta$ -cells communicate with muscle tissue *as if* the two were intentional agents. We are familiar with the general features of such interactions and have facility in reasoning with them. Likewise with morphogens, in gap-junction mediated signaling and in DNA transcription (I'll return in more detail to these cases below). Invoking an informational pretense consists in treating one element of a causal interaction as a sender, another as a recipient and an intermediate factor as a signal that informs the recipient of some state and/or induces in it an appropriate response.<sup>7</sup>

In particular, an informational description typically foregrounds a causal pattern with the following features:

Directionality. Communication involves designating a sender, a receiver and a direction of influence – from sender to receiver. The directionality in question might be spatial or temporal (most often, both). In biological cases signaling often occurs across an external-internal border, where the recipient, at least, lies within some enclosed space such as a nucleus, a cell or a particular organ. But in genetic regulation and in some cases of local positive or negative feedback loops, one spatiotemporal

---

<sup>7</sup> The distinction between a descriptive representation and an imperative one may be blurry or even nonexistent in simple types of communication systems (Millikan, 1995).

part of a process or entity might signal to another – like writing on one’s hand as a reminder to a future self. Thus treating some process as an instance of communication typically involves describing the location in time or space of the sender and implies that causal influence runs in a particular direction.

Connecting variation. Communication is a way of maintaining a correspondence between variable states of two distinct parts of a system – the receiver changes state as a function of a change on the sender’s end. Moreover, it does so in a pre-specified manner, according to what may be thought of as a rule of interpretation. Thinking in terms of an interpretive rule allows one to focus on the connection between the changes at the ends of the causal chain while de-emphasizing intermediate links. An informational description of hormones may highlight the connection between the level of a nutrient and metabolic activity by describing metabolizing cells as interpreting the hormonal signal according to a rule that specifies (e.g.) to elevate metabolism in response to a decrease in signal molecules. Such a description brackets much of the underlying detail of how the correspondence is maintained, but it gains a transparency in pointing to how an overall systematic connection is achieved.

Active vs. passive. Designating a certain process as involving communication or information transfer serves to highlight the active character of sender and receiver versus the passive character of the signal. Active and passive here are used metaphorically, I suppose. I intend them to refer to those parts of the system that undergo substantial change in contrast to those that stay relatively static. Knowing which parts of a system do not change is often very informative.

The activity of hormones illustrates this picture nicely. Typically, hormonal signaling molecules are sent by a gland in one part of the body, say in the brain or the liver. The signal is carried by the blood stream to its destination. Once bound to the recipient it either activates a “secondary messenger” or enters the cell itself, up- or down-regulating metabolism. Hormone molecules remain relatively unchanged in the process whereas the gland and the target tissue change states. Describing this as if it



were a case of signaling singles out the variation in the state of the sender (or the bodily parameter it is sensitive to, such as nutrient level) and how a corresponding metabolic activity occurs on the receiving end.

The morphogen case has a similar structure. The positional signal travels across the cell membrane (often, across the nuclear membrane as well) inwards to the receiver. It connects variation in the cell's location with changes in developmental fate. Metaphorically, this is described as the morphogen *informing* the cells as to their location, and the cells employing an *interpretation rule* to differentiate accordingly.<sup>8</sup> The fictional description tends to portray the morphogen as static while responding cells undergo substantial change.

DNA transcription shares some important features with hormones and morphogens, although it is a case somewhat unto itself. A signal in the form of a coding sequence is sent from the nucleus to the ribosomal machinery in the cell cytoplasm (or on the exterior of the nucleus). The ribosome then synthesizes a protein on the basis of the sequence, interpreting it in accordance with the genetic code. Here too the sequence stays largely intact while the ribosome is active, and here too there is correspondence between variation in sequence and the activity of the ribosome. But there are differences. The main one is that in the case of DNA it is not entirely clear who the sender is. Indeed it is not clear that there is a sender. We could designate the DNA itself as the sender and mRNA as the signal. But that would be to over-emphasize the differences between these molecules. We could speak of parents – or ancestors more generally – as senders. To some extent this might capture the role of inherited factors in development.<sup>9</sup> But treating ancestors as senders does not seem correct in the case DNA transcription in adults. It appears that the role of the sender is less significant in this case. Perhaps

---

<sup>8</sup> For some very explicit uses of this language see Gurdon & Bourillot, 2001; Ashe & Briscoe, 2006

<sup>9</sup> But not fully: maternally inherited proteins, morphogens included, are often contrasted with genetic material in being a form of maternal control over development.

this is because the signal is maintained in the cell throughout its life and used continuously. It appears more appropriate, within an informational description, to treat DNA as a repository of information and not so much as a sender, and indeed it is often so described. Notably, something similar occurs in the case of morphogens: The sender is often left unspecified, especially when morphogens are maternally transcribed and present in the egg. No doubt this has something to do with the fact that the sender is located, if at all, outside the organism in question. This does not mean that the sender is non-existent but perhaps explains why its role is de-emphasized. As it is far and unlocalized in time and space, its state is less relevant to understanding the mechanics of the system presently described. At any rate, these cases show that informational metaphors have considerable flexibility, and may stray from their paradigmatic form in a particular explanatory context.

Apart from foregrounding causal features of a specific biological system, informational descriptions have a bridging role too. As Jablonka (2002) emphasizes, describing a variety of systems – genes, hormones, morphogens – as engaged in information transfer allows one to compare and contrast their overall causal structure and behavior: what is the character of the signal, how is it sent, how does it exert its end-state effect, etc. These comparisons are useful in formulating hypotheses, conceiving of possible mechanisms, and communicating the big-picture among peers and in pedagogical contexts.

I have provided some indications of when informational language occurs and which features of a causal process it highlights. But it should be stressed that there isn't, on my view, a set of conditions under which informational language is uniquely appropriate. Many biological processes can be shoehorned into an informational mold – digestion could, if one really wanted, be described as a process in which food molecules are a signal received by the digestive tract. But biologists do not describe digestion this way. It is not easy to say why the language of information and communication is applied in some contexts and not in others. One could focus on cases of regulation or control, but I doubt that 'regulation' can be defined in a satisfying manner. Presumably, there is a mixture of reasons underlying

the decision to employ an informational description. To some extent it has to do with the structure and complexity of the phenomenon, but pragmatic considerations surely play an important role too.

Sometimes those aspects that are highlighted by an informational description are not the ones we care to highlight. Sometimes they are uninteresting, or already well-understood. These are good reasons but they mostly have to do with the interests of scientists rather than with objective properties of organisms.

That said, it is important to distinguish the reasons for opting for an informational set-up from the status of statements made within it. Once an informational description is found to be helpful and is put to use, there are definite standards of correctness within the fiction it invokes. These standards are answerable to the real features of the process being described. In glucose regulation the pancreas (and not the adrenal or pineal glands) is the sender; muscle cells (and few if any other cell types) are receivers; insulin (not one of the hundreds of other hormones in our body) is the signal; and the message is: ‘glucose levels are up, break glycogen’ (and not, say, ‘increase heart rate and reduce digestion’). This is significant, for it shows that *once a specific informational framework is in place, it allows us to capture objective properties* of the process in question. If one treats the process by which the pancreas controls glucose metabolism in informational terms, one is then obliged to designate the pancreas as a sender, insulin as the signal and so on. In other words, one may choose to view the process as an instance of intra-bodily communication, but it is not up to one what (informational) description the various elements should then receive. This is because the informational language serves as a way of pointing to the real (literally true) causal roles of those elements – the directionality of the process and other features described above. These features are independent of one’s choice of whether to employ an informational fiction, and they place constraints on claims made within it. Thus, while informational fictions are put to use on pragmatic and cognitive grounds, once they are invoked there is a right and a wrong in using them. In this sense, although biological information isn’t, as it were, really out there, employing a fiction according to which what is out there is sending and receiving messages is a way of tracking the real

causal goings-on.

## 8. Taking Information Too Seriously

My fictionalist proposal is motivated by the idea that even if we treat information non-literally we may still take it seriously and assign it a real role in biological understanding. But we shouldn't take it too seriously. If information is a metaphor then it is, after all, untrue that cells and molecules bear semantic content. Let me comment on two issues where this makes a difference.

The “metaphysics” of information. 20<sup>th</sup>-century molecular biology is sometimes taken to have made a fundamental ontological *discovery*, namely that genes and other informational factors constitute a distinct kind of entity populating the world. Indeed several eminent biologists have expressed such an attitude, suggesting a metaphysical split of sorts between the informational and the chemico-physical. G.C. Williams, for instance, speaks of a “codical domain” which exists alongside the material one.<sup>10</sup> In a similar vein, John Maynard Smith and Eörs Szathmari (1999) — in a nontechnical, more reflective version of their seminal work on major transitions in evolution — speak of “the dual nature of life...metabolic and informational” (p.11), and suggest that advances in molecular biology have given us a handle not only on inheritance and development, but also on metaphysics (they tie their discussion to the ideas of Aristotle, Descartes and Leibniz). Sydney Brenner, pioneer of molecular biology and Nobel laureate, has made similar claims. For instance, he states that in biological systems “in addition to flows of matter and energy, there is also flow of information” (1999, 1964).

For the most part, these claims about the ontological status of information are not defended in

---

<sup>10</sup> Williams sometimes speaks of “codices” as separate kinds of entities, at other times he seems to think of them as arising from a distinct level of description, and hints at a multiple realizability argument to buttress this idea (1992, ch. 2). If anything, this unclarity strengthens the present point.

detail. Information is accorded a prominent role in describing a total worldview, and the justification appears to lay in the prominence of informational notions within the practice of molecular and cellular biology. From a philosophical point of view, the justificatory gap is evident, and it is hard to shake the feeling that these biologists take an over-serious attitude toward information. If some form of material-informational dualism is a plausible view, it is not because of recent molecular biology. However, if one accepts that informational things exist – as, in some form or other, it appears a literal reading of information implies – then it is difficult to say why. The fiction-based view explains what is wrong with a metaphysics of information, and obviates the inference from the theoretical role of informational descriptions to the existence of informational “things”. Informational notions have theoretical significance, but this should not lead us to reify them. Describing hormones or DNA *as if* they are involved in informational transactions facilitates causal understanding. Williams, Maynard Smith and other information enthusiasts are illicitly taking the cognitive success of information as the basis for an ontological commitment.

Information and genetic determinism. Oyama (1985), Griffiths (2006) and others have criticized appeals to information in biology, in part out of a concern that informational discourse lends spurious support to genetic determinism. Griffiths, for instance, thinks that describing genes in terms of information and communication tends to bring with it unwelcome inference habits. For example, he holds that viewing something as content-bearing leads us to neglect the context sensitivity of its effects. In the genetic case, the thought is that treating genes as carriers of messages obscured the fact that the content of this supposed message depends (so Griffiths thinks) as much on the environment as on the specific features of the gene – its sequence etc. Now, perhaps there is some truth to the observation that we tend to treat semantic content as context-insensitive. But I think the kind of inferences we are inclined to make when describing something as having content depends greatly on the status of that description. Physicists occasionally speak of particles in intentional terms (“the particle doesn’t know

where it came from, it only cares about its present interactions”), but such locutions, while helpful at times, are clearly metaphorical. They do not tempt us to outlandish conclusions about the mental states of particles. Similarly, once we take a clear stance according to which informational locutions in biology are fictional,<sup>11</sup> our inclination to make some of the bad inferences that Griffiths and Oyama warn against should be weakened at the very least. One may resist genetic determinism while rejecting the inference from information to the inevitability of what’s in the genes.

## 9. Conclusion

My argument has been that in biology the language of information and communication is a liminal fiction, one that often escapes notice, but that it nonetheless bears theoretical weight by tracking a certain class of causal facts. It is this cognitive function that makes information and its cognates valuable concepts that contribute to biological understanding, and it is the generic character of the explanatory contexts in which they figure and their flexibility that accounts for their persistence and wide range of application. Now, to some readers the very notion that a fictional description could bear theoretical weight, or could be said to have explanatory value, will be hard to swallow. My response is that scientists employ a range of reasoning strategies that involve imagining away some elements of reality or pretending that things are different than they actually are: idealizations, simplifications, approximations. So long as we can make sense of their cognitive and epistemic contributions to scientific understanding (as I have endeavored to do in section 7), I see no in-principle grounds for excluding fictions and metaphors from the proper conceptual toolkit of science.

Philosophers of science often assume, sometimes implicitly, that scientific concepts come in two

---

<sup>11</sup> In the biological case the intentional discourse is more common and more central than in the physical case.

flavors – the literal, ontologically committing concepts, and the “merely” metaphorical ones that play an unimportant aesthetic role. My hope is that this discussion of information shows that there are scientific concepts that stand mid-way between the fully literal and the merely ornamental – concepts that have theoretical significance, but function indirectly via a pretense. Other examples come to mind – the treatment of energy as a “stuff” that can be transmitted, consumed or conserved; or the idea that natural selection is a force (as in “selection pressure”) – but there might be interesting differences between these cases.

Some recent work has suggested that fiction plays a central role in the context of scientific modeling (Godfrey-Smith, 2006; Frigg, forthcoming) and it would be informative to connect fictions of the sort I have discussed to these more regimented, mainstream uses of the imagination in science. The use of fictions in science raises many interesting questions, which I shall leave for another day. But it should be noted that an acknowledgement of the possibility of “serious” non-literality in science is by no means a blanket psychologization of explanation nor does it necessarily open the door for an overly permissive attitude towards standards of rational evaluation of scientific concepts. Acknowledging fictions is taking seriously the thought that as science is a cognitive enterprise, it will have recourse to representational devices that serve the cognitive ends of its makers and consumers. At the same time, it suggests that we have to be very cautious in making inferences from the fact that some set of concepts is theoretically valuable to philosophical conceptions about the structure of the natural world.

## REFERENCES

- Alberts, Bruce. Johnson, Alexander. Lewis, Julian. Raff, Martin. Roberts, Keith & Walter, Peter. (2002) Molecular Biology of the Cell 4<sup>th</sup> edition, New York: Garland Science.
- Ashe, Hilary L. & Briscoe, James. (2006) “the interpretation of morphogen gradients”, Development 133: 385-394.
- Brenner, Sydney. (1999) “theoretical biology in the third millennium”, Philosophical Transaction of the Royal Society, London, B, 354:1963-1965.
- Egan, Andy. (2008) “pretence for the complete idiom”, Noûs, 42(3): 381-409.
- Evans, W.H. and Martin, P.E.M. (2002) “gap junctions: structure and function”, Molecular Membrane Biology, 19: 121-136.
- Frigg, Roman. (forthcoming), “models and fiction”, Synthese.
- Godfrey-Smith, Peter. (2000) “on the theoretical role of ‘genetic coding’”, Philosophy of Science, 67: 26–44.
- Godfrey-Smith, Peter. (2006) “the strategy of model based science,” Biology & Philosophy, 21, 725-740.
- Godfrey-Smith, Peter. & Sterelny, Kim. (2008) “biological information,” The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), URL: <http://plato.stanford.edu/entries/information-biological/>
- Griffiths P.E. (2001) “genetic information: a metaphor in search of a theory,” Philosophy of Science 68: 394-412.
- Griffiths, (2006), “the fearless vampire conservator: Philip Kitcher, genetic determinism and the informational gene,” In Neumann-Held E.M. and Rehmann-Sutter C. (eds.), Genes in Development: Re-reading the Molecular Paradigm, Duke University Press, Durham, NC.
- Gurdon, John. and Bourillot, Paul, Y. (2001) “morphogen gradient interpretation,” Nature, 413:797-803.
- Jablonka, Eva. (2002), “information: its interpretation, its inheritance, and its sharing,” Philosophy of Science 69: 578-605.
- Jablonka, Eva. & Lamb, Marion, J. (1995) Epigenetic Inheritance and Evolution. Oxford, Oxford University Press.
- Maynard Smith, John. (2000) “the concept of information in biology,” Philosophy of Science 67: 177-194.
- Maynard Smith, John. and Szathmary, Eörs. (1999) The Origins of Life: From the Birth of Life to the Origin of Language, Oxford: Oxford University Press.
- Millikan, Ruth. (1984) Language Thought and Other Biological Categories, Cambridge: MIT Press.



- Millikan, Ruth. (1995) “pushmi-pullyu representations,” Philosophical Perspectives, 9: 185-200.
- Moussian, Bernhard. & Roth, Siegfried. (2005) “dorsoventral axis formation in the *Drosophila* embryo – shaping and transducing a morphogen gradient,” Current Biology, 15(21):R887-99
- Nelson, Philip. (2007) Biological Physics: Energy, Information, Life, New York: W.H. Freeman.
- Oyama Susan. (1985) The Ontogeny of Information: Developmental Systems and Evolution, Cambridge: Cambridge University Press.
- Rieke, Fred, Warland, David, de Reuter van Steveninck, Rob & Bialek, William. (1996), Spikes: Exploring the Neural Code, Cambridge, MA: MIT Press.
- Rosenberg, Alexander. (1986) “intention and action among the macromolecules”, in Rescher, Nicholas. (ed.) Current Issues in Teleology, Lanham, MD: University Press of America.
- Rosenberg, Alex. (2006) Darwinian Reductionism, Chicago: The University of Chicago Press.
- Sarkar, Sahorta. (1996) “decoding coding – information and DNA,” BioScience, 46: 857-864.
- Sarkar, Sahorta. (2000) “information in genetics and developmental biology: comments on Maynard Smith,” Philosophy of Science 67: 208-213.
- Shannon, Claude. (1948) “a mathematical theory of communication,” Bell Systems Technical Journal 27: 279-423,
- Shea, Nicholas. (2007) “representation in the genome and in other inheritance systems,” Biology & Philosophy, 22:313–331.
- Stegmann, Ulrich. (2004) “the arbitrariness of the genetic code,” Biology & Philosophy 19: 205-222.
- Sterelny, Kim. (2000) the ‘genetic program’ program: a commentary on Maynard Smith on information in biology,” Philosophy of Science 67: 195–201.
- Sterelny, Kim, Smith, Kelly C. & Dickison, Micheal. (1996) “the extended replicator,” Biology & Philosophy 11: 377-403.
- Walton, Kendall. (1990) Mimesis and Make-Believe, Cambridge, MA: Harvard University Press.
- Walton, Kendall. (1993) “metaphor and prop-oriented make-believe”, The European Journal of Philosophy, 1: 39–57.
- Williams, George C. (1992) Natural Selection: Domains, Levels and Challenges, New York: Oxford University Press.
- Wolpert, Louis. et al. (2006) Principles of Development, New York: Oxford University Press.