The Practical Value of Biological Information for Research

Beckett Sterner*†

Many philosophers are skeptical about the scientific value of the concept of biological information. However, several have recently proposed a more positive view of ascribing information as an exercise in scientific modeling. I argue for an alternative role: guiding empirical data collection for the sake of theorizing about the evolution of semantics. I clarify and expand on Bergstrom and Rosvall's suggestion of taking a "diagnostic" approach that defines biological information operationally as a procedure for collecting empirical cases. The more recent modeling-based accounts still perpetuate a theory-centric view of scientific concepts, which motivated philosophers' misplaced skepticism in the first place.

1. Introduction. Until recently, there has been general skepticism among philosophers about the practical scientific value of biological information as a concept. Criticism has appeared to pin biological information from multiple directions. Biologists regularly ascribe information to living systems in a way that seems to imply that information has semantic properties, such as representing the environment. Yet the obvious home for the concept, mathematical information theory, explicitly excludes semantic meaning from its scope. Simultaneously, biological information has become detached from its historical origin as a key concept in the Central Dogma's account of protein synthesis, and even in this role its value has been questioned (Sarkar 1996).

Received June 2013; revised August 2013.

*To contact the author, please write to: Field Museum, 1400 South Lake Shore Drive, Chicago, IL 60605; e-mail: bsterner@uchicago.edu.

†Chris DiTeresi provided pivotal feedback on a late draft of this article. My thanks to Bill Wimsatt, Bob Richards, Ken Waters, Alan Love, Scott Lidgard, and Rosa Cao for their comments and suggestions and to everyone in the philosophy of biology and history and philosophy of science groups at the University of Chicago over the years. Also, thanks to Erin Barringer-Sterner for her lovely help with the design of fig. 2. This research was supported in part by an NSF graduate research fellowship and postdoctoral grant SES-1153114.

Philosophy of Science, 81 (April 2014) pp. 175–194. 0031-8248/2014/8102-0005$10.00
Copyright 2014 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

Even if information does have theoretical value for explaining protein synthesis, others have argued that this would only allow it a minor role in biology at large (Griffiths 2001; Godfrey-Smith 2007). In stark contrast to such a conclusion, biologists have rapidly expanded their use of biological information over the past several decades: there is positional information in developmental biology, information processing in cell biology, information in signaling or communication between organisms, and so on (Maynard Smith 2000; Taylor, Tishby, and Bialek 2007).

A more positive view has emerged over the past several years from reconsidering biological information as an exercise in scientific modeling (Godfrey-Smith 2007; Levy 2011; Griffiths and Stotz 2013; also see Sarkar 2005). While Godfrey-Smith and Levy allow information a positive practical role, they limit its meaning to describing a certain kind of predetermined causal structure, such as the sender-receiver schema. On their view, the possibility of semantic content is irrelevant or incoherent within the reconstructed practice and can thus be ruled out on conceptual grounds alone. In contrast, I will argue for an alternative view in which the existence of semantic capacities is an empirical topic for investigation rather than a logically prior issue of coherence. What we need, then, is an account of the concept whose practical value is motivated by its very neutrality with respect to plausible theories of the semantic aspect of information.

Carl Bergstrom and Martin Rosvall have made a related claim with their proposed transmission account of biological information (Bergstrom and Rosvall 2011a, 2011b). They have argued that the concept can be given a clear, literal meaning without recourse to any notion of semantics. Their most recent statement of the account is that "An object X conveys information if the function of X is to reduce, by virtue of its combinatorial properties, uncertainty on the part of an agent who observes X" (Bergstrom and Rosvall 2011a, 198). The transmission account combines an information-theoretic analysis of signals as reducers of uncertainty with the restriction that the signal serves an evolved function.

However, Bergstrom and Roscall's account has met with objections (Godfrey-Smith 2010; Maclaurin 2010; Stegmann 2013; but also see Shea 2010). The biggest problem for our purpose here is that the transmission account does not address how biological processes might come to acquire semantic content. Having the evolved function of reducing uncertainty in a statistical sense may thus be an "available and coherent" meaning for biological information, but it appears irrelevant to what is arguably the most pressing issue in clarifying the concept (Maclaurin 2010, 193).

In fact, this negative conclusion precisely inverts the reason why a view like Bergstrom and Rosvall's has practical value. Biologists do not have to begin with a perfect account of how and why biological processes have semantic properties in order to fruitfully investigate positional information

https://doi.org/10.1086/675679 Published online by Cambridge University Press

or information processing, for example, as potentially semantic in nature. Instead, what they need is the practical ability to collect and study empirical cases that are relevant to theorizing and testing claims about if, how, and why these phenomena exhibit semantic properties. If we can find a definition of biological information that facilitates this practical work of data collection for the sake of future theorizing, then the concept can have a positive role in research independently of any particular view of semantics. In fact, the ultimate value of the concept in this regard would be to offer a logically neutral starting point for empirically informed theorizing about naturalized semantics.

Historically, the possibility of this practical role for biological information was obscured by a kind of theory-centrism that assumed that the concept must succeed in the manner of a traditional philosophical account in order to be philosophically interesting. Certain aspects of this theory-centrism continue to operate within the recent modeling-based approach, that is, a top-down understanding of the semantics of models. Thoroughgoing alternatives to theory-centrism are particularly valuable in the case of biological information, given its potential to serve as a mediator between the physicalist language of naturalism and concepts such as intentionality, representation, and meaning (Godfrey-Smith 1996; Millikan 2004; Skyrms 2010).

2. Bergstrom and Rosvall's Transmission Account. The biggest philosophical challenge behind understanding biological information is to explain how biological processes can acquire semantic properties such as being "about" an external state of the environment and being "correct" or "incorrect" about this state—or to definitively disprove that such properties exist. Although useful in many other ways, Claude Shannon's mathematical theory of information does not help us in settling this issue: information theory characterizes the features of a causal process at a strictly syntactic rather than semantic or pragmatic level (Shannon and Weaver 1949). Hence, we can talk about information in the world as the existence of merely physical correlations between states, but there seems to be a gap between this sense of information and the way some biologists use the term. Bergstrom and Rosvall's transmission account provides something in between: it offers no explicit grounds for ascribing semantic properties as part of biological information, yet it incorporates the idea that biological information must be functional in an evolutionary sense in order to be practically relevant. The question is, what do we gain from this new middle ground? Bergstrom and Rosvall point briefly to an answer, but some of the objections raised to their work show the need for a more thorough justification.

As a quick preliminary, recall that mathematical information describes the average capacity of a transmission channel to send information, usually

https://doi.org/10.1086/675679 Published online by Cambridge University Press

represented as sequences of symbols. These signals help the receiver of the transmission discriminate between possible states of the sender. More generally one can also measure the degree to which one signal can be substituted for another given their probabilities of occurrence. Mathematical information on its own is not adequate to ground semantic properties for information: one must go beyond the probabilities of syntactic combinations of signals to model the meaning of these signals.

Shannon's notion of an information channel is built on top of causal specificity. If the receiver's output behavior does not vary as an effect of the sender's signals, then there is no point to determining the channel's information capacity or any other measure. In the converse direction, if we find that input variations in the channel are a good guide to variations in output, then this indicates the presence of causal specificity, presuming we already know that the input signals causally influence the output. (Information theory also goes beyond causal specificity in many ways, of course.)

Bergstrom and Rosvall's transmission account in revised form is as follows: "An object X conveys information if the function of X is to reduce, by virtue of its combinatorial properties, uncertainty on the part of an agent who observes X" (Bergstrom and Rosvall 2011a, 2011b). They describe this definition as a "transmission" account because it focuses on the sending and receiving of signals across generations of cells. Genetic inheritance therefore qualifies as transmitted information in this context, but so do epigenetic methylation patterns. The word "uncertainty" here means the predictability of the actions of the receiving agent, so that maximum uncertainty would mean a uniform distribution of probabilities for all output states given any input state. In this way, uncertainty is an objective, nonsemantic property rather than an intentional state of the agent about its own actions. Also, Bergstrom and Rosvall mean function in an evolutionary sense, that is, having evolved in the past because sending and receiving the signal was adaptive for organisms within the lineage.

What I want to stress here with respect to the transmission account is that it offers one way to pursue a more general approach to conceptualizing biological information. Other interpretive schema besides the sender-receiver template are workable within this more general approach, and there are also different possible tactics for turning the account into a practical tool (Stegmann 2013). Both of these aspects require further study, but my focus here will be limited to giving a more thorough motivation for why something like Bergstrom and Rosvall's account matters to the larger debate over biological information. Maclaurin's critical response, in particular, demonstrates the need for this prior work to clarify the potential value of the framework in the larger debate (Maclaurin 2010).

Maclaurin's concern centers on why the transmission account should matter for people who are primarily interested in whether and how bio-

https://doi.org/10.1086/675679 Published online by Cambridge University Press

logical processes have semantic properties. Maclaurin gives an argument for why the transmission account is not relevant, using an analogy between genomes in a cell and libraries in a university: "We are convinced libraries are repositories of information because they are clearly designed to transmit that information. Principles of grammar, syntax, cataloguing and physical library design are clearly designed to maximise the ability of readers to extract from a library the information they want" (Maclaurin 2010, 192). However, "Bergstrom and Rosvall's interpretation of biological information will seem recondite to many. Readers care deeply about the content of books. . . . But the cataloguer's art is one that only a librarian could love." Moreover, "systems used by library cataloguers have no effect on the semantic information contained in their books" (193).

The crucial error in Maclaurin's analogy is his presumption that we already know how to read (i.e., understand the meaning of) the contents of a library. If we already know how to read, and what we really want is to understand what the books in the library are about, then the way that the books are organized matters only through its impact on our ability to find and access particular books. However, this is not the situation we are in with respect to biological processes such as quorum sensing in bacteria or epigenetic modifications like methylation. It would be more accurate to assume that we do not even know what human language is. In this context, identifying and studying the workings of things like libraries could be incredibly useful for testing and articulating claims about the properties of natural language.

This is exactly the fix Bergstrom and Rosvall recommend in suggesting that we "take the perspective of the proverbial anthropologist from Mars" (Bergstrom and Rosvall 2011a, 197). They write: "We freely admit that these processes are not the most interesting thing to be found within the library—the anthropologist won't learn all that much about humanity by studying the Dewey Decimal system—but these processes are highly diagnostic of a library" (197). In other words, the transmission account is properly understood as a means to an end rather than an ultimately complete explication of biological information. In the remainder of this article, I aim to clarify and provide principled grounds for why this alternative approach to biological information is in fact a positive contribution to the project of naturalized semantics rather than a distraction or dead end.

3. Empirical Strategies for Fixing the Scope of Biological Information. What could be the role of something like the transmission account for the larger project of naturalized semantics? Bergstrom and Rosvall suggest that "it simply offers the tools to diagnose biological information" wherever it can be found (Bergstrom and Rosvall 2011a, 197). But what is a "diagnostic tool" for biological information, and why is it helpful? To answer

https://doi.org/10.1086/675679 Published online by Cambridge University Press

these questions, it will help to first back up and recognize that the transmission account is offering an operational definition of biological information. That is, the meaning of the concept is specified in terms of a procedure for judging whether some empirical phenomenon falls under biological information as a kind. Moreover, there is no a priori way to calculate the extension of biological information based on this procedure because we lack a way to enumerate all the possibly relevant biological processes. As a result, biologists must go out and determine empirically which cases pass the test and which do not.

Although Bergstrom and Rosvall do not discuss it, there are in fact multiple ways to give an operational definition of biological information. In order to get clear about what the diagnostic approach is in particular, it will help to contrast it against a common alternative. Hence, in this section, I will lay out two alternative strategies one could use to give an operational definition. In both cases, the procedure is empirical in the sense that the only way we can find out the extension of biological information is to go out and apply the procedure to different phenomena. The two strategies differ, however, over whether collecting and organizing the cases we find should go hand-in-hand or be kept separate. I will discuss similarity to a paradigmatic exemplar as an instance of collecting and organizing simultaneously and identification by general diagnostic traits as an instance of holding them separate.

Probably the most common way of judging whether a new phenomenon is a case of biological information is to compare it to genetic information. If the two are similar in the right way, then the new case counts as an instance of the kind. We can call this the *exemplar-based* approach to operationally defining biological information. In order to give an explicit account of the exemplar-based approach, one would need to specify the exemplary case (e.g., protein synthesis), its relevant properties, and how to judge the similarity of a new case with respect to those properties.

By contrast, the *diagnostic* approach demarcates biological information using a procedure for sorting cases based on the outcomes of various test measurements. If the necessary properties are present, then the case is a positive instance. Otherwise, it is excluded. Think, for example, of identifying species of bacteria by staining them with different chemicals. Those that behave one way under a bevy of tests can be sorted into one species, while others will go elsewhere, depending on their particular combination of results. The procedure consists of the bevy of tests and rules associating specific results with particular species.

Using protein synthesis as an exemplar, however, seems increasingly inadequate. Biologists have adapted biological information to apply to epigenetics, gene networks, cell signaling, development, and other cases that on a concrete material level share little in common with protein synthesis.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

Judgments about similarity would have to operate at a higher level of conceptual abstraction, such as "is like a code," that are already contentious and hard to define. Indeed, our understanding of protein synthesis itself has become increasingly complicated and contextualized (Sarkar 1996).

In consequence, an empirical approach to biological information would be better off not giving any privileged status to genetic information. That is, it is better to avoid a strategy for collecting cases that implicitly organizes them in terms of similarity to protein synthesis.[1] Instead, the diagnostic approach offers a fruitful alternative.

4. Background Principles for a Diagnostic Approach. How can the diagnostic approach be pursued in practice? At a minimum, we need a diagnostic procedure and a principled standard for evaluating the procedure's success. In a situation where possible theories are highly contentious, it is better to find a neutral ground for the procedure. The value of an empirical approach in this circumstance will be precisely its ability to proceed without assuming a prior theory of the phenomenon. Indeed, its aim will be to systematically collect the data we believe are relevant for theorizing. In this manner, the diagnostic approach to biological information could serve as a means to investigating the semantic properties of biological processes without presupposing a theory of semantics.

In order to be valuable, a diagnostic approach should provide a means for comparing and testing claims about semantics in biological systems without presupposing any particular theory. Under a naturalistic view, any theory of semantics should allow us to determine whether a given system possesses semantic capacities or not based on its physical properties.[2] The utility of the diagnostic approach, then, is to provide a collection of cases within which different views of semantics can differentiate themselves. Ideally, the collection should be comprehensive enough to test the full scope of these views but narrow enough to cut out cases that are irrelevant to all of them.

The heuristic I will use to find this balance is to pick out properties that are necessary for semantic content but not yet sufficient (on any nondeflationary view). Each necessary property forms a diagnostic criterion that filters out some irrelevant cases but does not differentiate among possible accounts any further. One could thus criticize these criteria on the grounds

1. My position thus converges with advocates for the causal parity thesis but proceeds on different grounds and does not deflate the value of information in general (Griffiths 2001).

2. This does not have to involve a totalistic reduction of semantics to the language of physics. See Wimsatt (2006) for a discussion of how we can maintain a localized, methodological commitment to physicalism while remaining agnostic about the broader relationships between disciplines.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

that they are not strict enough, that they are too strict, or that they in fact possess a bias toward one theory over another.

The first criterion is that one cannot have biological information without causal specificity. James Woodward offers a useful account of causal specificity for our purposes (Woodward 2010). Think of the relevant causal system as consisting of a set of input variables and a set of output variables. The initial causes are connected to their ultimate effects by various causal processes in between. In basic terms, causal specificity is a matter of degree, depending on the structure of the mathematical mapping relating input states to output states. There are two major dimensions: specificity is maximal if every input state maps to a distinct output state and if every output state is mapped to by some input state. (Mathematicians would say the function is one-to-one and onto.) If there are some overlaps in the output or some states are not covered, then causal specificity decreases.

To be clear, I am not attempting a reduction of semantics to causal specificity: the connection is based on the weaker claim that a physical system that realizes semantic properties must exhibit causal specificity under those relevant conditions. For a system to have semantic properties, it must have activities or states that we can say are "about" something else. For example, is a signaling protein in a cell that controls the cell's motion toward a food source "about" the surrounding chemical environment? If the physical system does not respond differently to what it is about than what it is not about, then aboutness is a distinction without a difference.

In requiring causal specificity, several other properties are implied about the causal system. For one, we should be able to describe the system in terms of input and output variables, that is, a set of initial causes and a set of downstream effects. Obviously, there is also a directed causal relationship running from input to output causes.[3] In practice, biologists typically focus on causal relationships that exhibit specificity and that are also mechanisms (Bechtel and Richardson 1993), but it is not necessary in principle to constrain the scope of biological information to exclude one-off events. Nonetheless, for simplicity, I will assume going forward that instances of biological information refer to the actions of mechanisms.

Still, causal specificity alone is obviously too broad. What can be added to narrow it down? An answer can be found through considering another question, "Why would having semantic properties matter for living systems?" Under a nondeflationary view, if there are things like aboutness and representation to be found systematically in living systems, they are not there merely incidentally. For instance, consider Godfrey-Smith's notion of a "fuel for success" in the evolution of the mind (Godfrey-Smith 1996). In

3. Cyclical causal systems could have the same variables as input and output but index their values by time.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

other words, we would expect that semantic properties would predominantly be found in living systems because of what they do for those systems rather than because of other processes such as random drift.

An obvious addition, then, is to require that the causal specificity be found in an adaptive biological mechanism, that is, a mechanism whose operation contributes to the organism's fitness under some selection regime (not necessarily historical). Moreover, the effects of the mechanism should be adaptive in part because of the causal specificity it exhibits. The organism should be more fit in its environment because the mechanism does different things under different conditions.

Now, adding adaptiveness might raise a couple of concerns. One is that we are making a substantive assumption about semantic theory, so the diagnostic approach is not as unbiased as it seemed. Another is that adaptiveness is difficult to demonstrate empirically and irrelevant to the practices of many molecular and cell biologists.

The first concern is not a problem because adaptiveness is serving as a diagnostic trait rather than as part of an explanation of why the system has semantic properties. Asserting that capacities for semantic reference should be evolutionarily valuable does not imply that this evolutionary value is what makes them semantic. By contrast, teleosemantic theory uses evolutionary function in part to define what it means for a state of an organism to refer to another object (Millikan 1984, 2004). Alternatively, an externalist, causal theory of representation (e.g., Dretske 1981) would need to explain why the capacity for representation is generally valuable for organisms. As a result, we can use adaptiveness as an indicator for relevance to semantics without a priori committing to an explanation for why semantic capacities are valuable.

The second concern is also not as big a problem as it might seem. The most important point is that contributing to fitness under some selection regime is not the same as adaptationism. A biologist would not need to demonstrate that the mechanism in question evolved primarily due to natural selection and is functionally optimal. Adaptiveness, as I use it, is more general than historically selected effects; instead, it is determined relative to a selective regime that can be historical or ahistorical (Walsh 1996). In order to identify a case of biological information, one only needs to show that the mechanism contributes to fitness under a stipulated selection regime, "the total set of abiological and biological (including social, developmental and physiological) factors in the environment of the trait which potentially affect the fitness of individuals with the trait" (Walsh 1996, 564). The case will then be indexed to the conditions under which that selection regime obtains. For example, it would be legitimate to call a process informational if it contributes to fitness today even though it evolved for other reasons in the past (i.e., the process is co-opted for new use). Finally, regarding the need for evidence that adaptiveness depends on the causal specificity of

https://doi.org/10.1086/675679 Published online by Cambridge University Press

the mechanism, one can show this experimentally by disrupting or altering the mechanism and measuring whether reproductive success varies over the sample populations. Somewhat weaker evidence is also possible through simulating or modeling the effects of this intervention (see the example from systems biology below).

When such evidence is lacking, it does not mean that research on the case must stop. Biologists' ability to study the mechanistic organization of a process does not depend on its categorization as informational (see Bechtel and Richardson 1993). As a research program, the study of information without knowledge of adaptiveness can be understood as the study of general kinds of causal specificity that are known to occur in living systems and are assumed to be relevant as such to their fitness in the wild. This strategy can discover new cases of information that evolutionary theory did not anticipate, for example, through exhaustively investigating the causes of variation in gene expression. Under my view, this use of biological information is a heuristic that allows research to proceed in the absence of evidence about adaptiveness. It represents a practical characterization of the larger, mutually supportive though partially autonomous relationship between evolutionary and causal role functions. I believe the stronger account using adaptiveness can make sense of the meaning and importance of the weaker notion while supporting inquiry into the evolution of semantics in a way that the weaker notion could not. However, a broader discussion about the inferential and practical relationships between evolutionary and causal role function concepts is outside the scope of this article.

In sum, any instance of biological information should at least exhibit the following property: we should be able to identify a biological mechanism that contributes under some selection regime to the fitness of the organism of which it is a part and does so in part because there is a causally specific relationship between its input and output states. Any procedure that actually implements the diagnostic approach should be at least as narrow as this standard. The next section discusses some possibilities.

5. Implementing the Diagnostic Approach. The principled grounds I gave for the diagnostic approach are more general than the transmission account in that they do not depend on a particular mathematical formalism for quantifying specificity. In this section, I discuss an example of what would count as demonstrating a case of biological information based on causal specificity alone. Without question, one could also use information theory, but the diagnostic approach is not strictly constrained to choose among particular mathematical formalisms (Stegmann 2013).

The example I will use is from ongoing research in systems biology on causal patterns, known as motifs, in gene networks (Alon 2007a, 2007b). For our purposes here, much of the background to the science will not matter

https://doi.org/10.1086/675679 Published online by Cambridge University Press

(for further discussion about motifs and systems biology more broadly, see Boogerd et al. [2007]). What is important here is the behavior of a particular kind of causal interaction (a motif) among three genes that is called a coherent feed-forward loop (FFL). See figure 1 for a picture of the eight different kinds of FFLs drawn as networks. The relevant FFL motif for our discussion (see top-left diagram in fig. 1) is characterized by the way that one gene, X, increases the expression levels of two other genes, Y and Z. Gene Y is also able to increase the expression level of Z, but in order for Z to turn on both X and Y must be active. The two-step connection from X to Y and Y to Z is the "feed-forward" part of the pattern, while the fact that X and Y both increase the activity of Z makes it "coherent."

In a sequence of papers, Erez Dekel, Shmoolik Mangan, and Uri Alon characterized the causal behavior of coherent FFLs as "sign-sensitive delay elements" within gene networks (Mangan and Alon 2003; Mangan, Zaslaver, and Alon 2003; Dekel, Mangan, and Alon 2005). I will explain the phrase in a moment. They have also shown using computer models that coherent FFLs can be adaptive under a general class of environmental conditions. In other words, they argued that coherent FFLs have a distinctive functionality for the fitness of cells based on their causally specific response to external

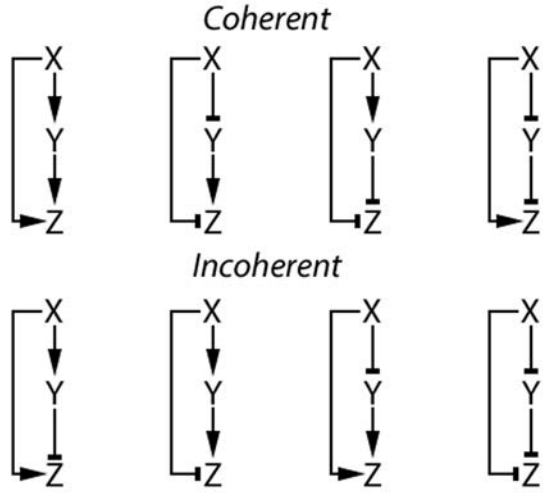


Figure 1. The eight possible feed-forward loops. Possible causal interaction networks between three genes in a cell. X, Y, and Z depict genes, while edges describe their interactions. Edges with an arrow indicate that one gene increases the activity of the other, while an edge with a line indicates that it decreases activity. Coherent FFLs have the same net effect along the direct edge from X to Z as the indirect path through Y: positive, negative, negative, and positive, respectively from left to right. The direct and indirect paths have opposite effects in the incoherent FFLs.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

conditions. Moreover, they explicitly conceive of motifs such as FFLs as carrying out "information processing" within the cell.

The description of coherent FFLs as sign-sensitive delay elements is based on an analogy from electrical circuit elements. Instead of current running through the wires of a circuit board, one has flux in gene expression coursing through networks of gene interactions. The structure of the gene network—that is, which genes influence the expression of other genes and how—determines the way that a change in conditions outside the cell would dynamically affect the internal expression of its genes over time.

Although Alon and his colleagues used mathematical modeling to demonstrate the behavior of coherent FFLs as sign-sensitive delay elements, it is possible to explain the basic insight qualitatively. Recall that X affects Z indirectly through Y and also directly without mediation. Also, both X and Y must be active for Z to turn on. This implies that if some environmental change turns on the expression of X, it will take some time before X will have turned on Y such that both are present in sufficient amounts to activate Z. The FFL therefore exhibits a temporal delay with regard to changes activating X. Alternatively, when X is turned off, its expression level will decay quickly and Z will turn off quickly even though Y is still present. Hence, the FFL responds without delay to off-switches in X. This makes it a "sign-sensitive delay element."

In Dekel et al. (2005), the authors show using computer simulations that coherent FFLs are adaptive when it is costly for a cell to act on short bursts of change in the environment. For example, imagine that gene Z in the FFL produces one or more proteins that digest a sugar molecule into useable energy for the cell. The environment around the cell varies over time stochastically, so there will be times when the sugar is present only for a short period and others when it is around for longer. Producing these proteins is only beneficial when the sugar is present for a substantial period of time; otherwise the cell will expend more energy on making the proteins than it will gain through their activity. Because the coherent FFL exhibits a time delay for activation of X, short bursts of change will not last long enough to activate Y and thus Z. The delay acts as a filter to save the cell energy and benefits the cell when short bursts are common.

Thus, it is precisely the distinctive sort of causal specificity exhibited by coherent FFLs with regard to environmental change that makes them adaptive. Although quantifying the causal specificity of coherent FFLs was not strictly necessary, mathematical information theory is just as applicable in this class of cases as for Bergstrom and Rosvall's original transmission schema. It is worth pointing out that the informational status of coherent FFLs does not depend on an evolved signal in the sense of Bergstrom and Rosvall, since the motifs can also respond to nonevolved environmental changes such as food sources. In light of this, the diagnostic approach is

https://doi.org/10.1086/675679 Published online by Cambridge University Press

properly more general than the classical sender-receiver scenario. Eva Jablonka has advanced an account of information along this alternative line that depends only on the receiver being able to use the input to adaptive effect without the need for an evolved sender (Jablonka 2002).

6. Escaping Theory-Centrism. The error underlying the historical skeptical view of biological information can be localized to a kind of theory-centrism. The error lies in assuming that any philosophically interesting account of biological information must succeed in the manner of a traditional theoretical account. If we drop that assumption and recognize the viability of an operational approach, new possibilities open up not only for what counts as a philosophical account of biological information but also for what makes the concept of biological information philosophically interesting.

Theory-centrism is a well-recognized problem for philosophy of science at large. It involves the assumption, coming out of logical empiricism, that scientific knowledge is "encapsulated in scientific theories" and moreover "that scientific theories are to be understood as axiomatic systems to which the methods of logical analysis could be applied" (Giere 1999, 33). While the traditional view of theories has expanded somewhat to include model logic, the basic presumption behind theory-centrism is the same: in order for biological information to have value for biology, it must have an irreplaceable role within the inferential structure of some scientific theory.

The scope of theory-centrism about biological information is plain: one need only survey the extensive efforts to defend or criticize the concept as it figures in either the Central Dogma (Sarkar 1996; Godfrey-Smith 2000; Oyama 2000; Griffiths 2001; Stegmann 2005) or Shannon's mathematical theory of information (Griffiths 2001; Pfeifer 2006; Jantzen and Danks 2008). It has simply been assumed that these two theories provide the best resources for defending or destroying the legitimacy of biological information. See Griffiths and Stotz (2013) for a related discussion.

Giere's description of theory-centrism is a composite: it is a dual claim about where scientific knowledge can be found and how it can be analyzed. Both components of theory-centrism are misleading for estimating the value of biological information. From the first claim, we are tempted to conclude that a concept cannot have any scientific value if it does not appear in a theory because the aim of science is the production of knowledge encapsulated within theories. Even if we avoid this error, the second claim might still lead us astray: we might think that finding value in biological information outside theory implies an opposition to theorizing because the concept would thus also be outside the domain of our methods of logical analysis.

The first mistake has been treated extensively within the philosophy of science over the past few decades. Pivotal insights for the field have come from studying models (Cartwright 1983; Morgan and Morrison 1999; Wim-

https://doi.org/10.1086/675679 Published online by Cambridge University Press

satt 2007), experiments (Hacking 1983; Mayo 1996), and data collection (Star and Griesemer 1989). Crucially, our understanding of a concept's value for scientific research has expanded outside the context of justification to include processes in the context of discovery.

The role I have described for biological information places it in what might usefully be called the arena of data collection practices (Star and Griesemer 1989; O'Malley 2011; Shavit and Griesemer 2011; Leonelli and Ankeny 2012; Strasser 2012). Scientific knowledge in this context is embodied in the methods scientists use to collect data, any methodological theories they might have, and how they structure the data for communication and analysis. Traditionally, philosophers have conceived of data collection as driven entirely by the empirical testing of theories, but this is again a theory-centric view. Scientists regularly gather data in an exploratory fashion and for the sake of constructing phenomenological models despite the conventional hypothesis-testing picture of the research process (O'Malley 2011; Strasser 2012).

The second worry is that defining biological information in terms of a procedure for collecting cases might appear to block any sort of theorizing about the phenomenon itself. Under a theory-centric view, this would seem to put a hold on theorizing until all data collection was complete. This is wrong in a couple ways. First, scientific concepts can play multiple simultaneous roles in research. Theorizing about the nature of speciation, for instance, does not halt while biologists collect more data about the biogeography of species. Second, there is more to theorizing than theory: the process of theorizing includes building, organizing, and revising conceptual entities besides theories, such as models, isolated statistical and explanatory generalizations, conceptual notations, and mathematical tools (Weisberg 2007; Griesemer 2012). In this regard, theorizing can progress in the absence of any preexisting systematic theory. By contrast, the methods of logical analysis presumed by theory-centrism would limit theorizing to activities such as deducing consequences from a preexisting theory, testing the logical consistency of the theory against counterexamples, revising the axioms or semantics of a theory, and defining a new theory in terms of its axioms and semantics.

Figure 2 summarizes the distinctions we need to ameliorate the theory-centric prejudice against biological information. The portion of biologists' research that is accurately described by a theory-centric view is only part of the work of biological theorizing and is an even smaller part of the work of biology as a whole. A proper evaluation of biological information needs to examine all of the roles that the concept plays in biological research, not merely the theory-centric ones.

In this regard, figure 2 does not address the dynamic interplay between data collection and theorizing (broadly understood). Moreover, this interplay happens at two levels: within biology itself and between biology and

https://doi.org/10.1086/675679 Published online by Cambridge University Press

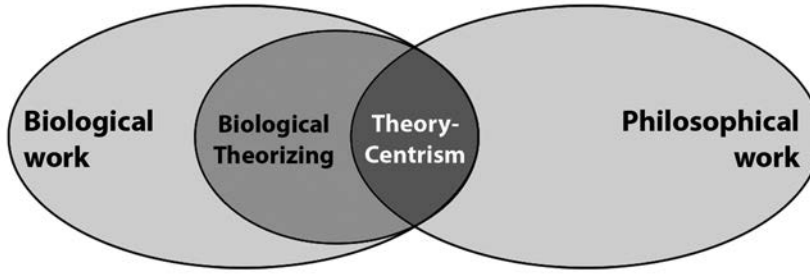


Figure 2. Overlap between practices in biology and philosophy. The diagram illustrates a limited theory-centrism, in which we recognize that concepts can play a valuable role in biologists' work outside the activity of theorizing, and that biological theorizing is not contained within the traditional philosophical view of theorizing. However, the diagram still poses the question whether there is substantial overlap between philosophical and biological work outside a theory-centric view.

philosophy. As a group, philosophers have had two distinct motivations for analyzing the biological information. The first is understanding why biological information is an increasingly popular idea across almost every field of biology. The second is whether biological information can be used as a resource for giving philosophical accounts of semantic reference, consciousness, and meaning, among other things. I suggest that philosophical problems in this cluster of topics can indeed be informed by biological data in the absence of systematic theories if we take inspiration from the broader process of theorizing in science.

7. Theorizing with More than Theory. Understanding biological information operationally requires new ways of approaching theorizing between biology and philosophy. I have already discussed how there is more to the work of theorizing than manipulating or proposing whole theories. To give a more positive account of this surplus, we need to explicitly bring in other scientific entities, such as data or models, into the process. Several authors have recently done this for biological information based on the practice of scientific modeling (Godfrey-Smith 2007; Levy 2011; Griffiths and Stotz 2013). However, I will argue that the modeling-based view has maintained certain flawed elements of theory-centrism. I will clarify these lingering aspects using a distinction from systems biology between top-down and bottom-up methods for theorizing.

Historically, both positive and negative assessments of biological information came in the form of top-down theorizing: one would draw on some predetermined theoretical concept (e.g., the sender-receiver schema in information theory) and then apply it uniformly via some principle (e.g., isomorphism) to characterize a whole domain of phenomena. The validity of

https://doi.org/10.1086/675679 Published online by Cambridge University Press

the theory would then be determined by its coherence over possible counterexamples within the stipulated domain. By contrast, a bottom-up approach begins with the formulation of one or more local characterizations about a subset of cases, which might come in the form of models, generalizations, or comparisons. These characterizations are then expanded inductively to test their adequacy over a broader range, which may involve revising or replacing them. The structure of the characterizations is not specified in advance and is often generated through abductive inferences. Crucially, bottom-up theorizing does not presuppose the existence of a universally true theory and may result in multiple general characterizations that only cover the domain as an aggregated set.

The first holdover from a theory-centric context is the lingering assumption that positive arguments for semantic content in information must come in the form of a top-down theory. At this point, the main contender for a positive semantic view comes from teleosemantics. The development of teleosemantics, however, has proceeded in a top-down fashion from Ruth Millikan's initial presentation of a complete theory to subsequent modifications in light of counterexamples (Millikan 1984, 2004; Shea 2013). As I will describe below, however, bottom-up approaches are also possible.

The second holdover transfers a top-down view of the semantics of theories to the semantics of models. As things stand for biological information, this issue comes in two parts: (*a*) defining the sender-receiver framework and (*b*) fictionalism about models. Godfrey-Smith's model-based view of biological information treats it as a loose exercise in mapping Shannon's sender-receiver schema onto biological systems (Godfrey-Smith 2007). In Godfrey-Smith's view, the scope of information talk is determined in advance by the structure of this schema and the meaning of "sender" and "receiver" as concepts. This assumes that the scope of information as a model is fixed from the top downward by our prior choice of interpretive principle.

The alternative, however, is that we determine our set of cases first, for example, by the diagnostic approach, and judge our models by how they fit with these cases. I take Bergstrom and Rosvall to be following this bottom-up approach when they wrote (in response to Godfrey-Smith's criticisms on this issue) that "when structures become more complicated than can be represented by the basic telegraph schema, we can keep using [information] theory. Just as the theory has extended beautifully to more complicated scenarios in telecommunications engineering and computer science, we expect the same for application to biological heredity and development" (Bergstrom and Rosvall 2011a, 200). Such expansion in the face of new cases would not be possible if biological information were simply equivalent to a certain schema and interpretive principle. (I have of course already argued that the diagnostic approach depends on causal specificity more fundamentally than information theory per se.)

https://doi.org/10.1086/675679 Published online by Cambridge University Press

Fictionalism about information in effect uses a top-down theory of modeling practices to foreclose the possibility that scientists could investigate the literal truth of semantic capacities in living systems (Levy 2011). In response, Griffiths and Stotz have pointed toward Mary Hesse's argument that models have open-ended, neutral analogies to their targets as well as negative and positive analogies (Griffiths and Stotz 2013). Indeed, the basic point of the diagnostic approach is that the evolution of semantic capacities can be an empirically open question rather than something settled by a logically prior reconstruction of modeling practice.

Moving forward, investigating the evolution of semantics empirically requires a positive method for theorizing on top of the diagnostic criteria. I will suggest a comparative approach as a complementary alternative to modeling (liberated from theory-centrism). Leading biologists have already been using a bottom-up, comparative approach to theorizing about biological information for over a decade (Maynard Smith and Szathmary 1995; Jablonka 2002; Jablonka and Lamb 2005). For positive examples of bottom-up modeling work, see Taylor et al. (2007), Skyrms (2010), and Shea, Pen, and Uller (2011). Comparative theorizing focuses on discovering relational principles between cases. Let us say case 1 has property X and case 2 does not. The comparative approach looks for differences among other properties of the two cases that explain the contrast. The explanation is anchored concretely in the relationship between these two cases and does not depend on a similar explanation holding for other comparisons. However, analyzing new cases in light of this local explanation can lead to a more general framework or model.

Here is an example of a philosophical problem about biological information posed in a comparative mode. It is widely remarked that people are more comfortable ascribing semantic properties to information carried by neurons than to mechanisms outside the nervous system. Why? Perhaps neurons possess representational capacities missing in other mechanisms. In response, we could investigate the same kind of functional behavior in an organism possessing a nervous system and one without it. Both *E. coli* and *C. elegans* (a species of nematode) exhibit chemotaxis, the behavior of moving up concentration gradients toward sources of food and down gradients away from toxins. *E. coli* obviously does not possess a nervous system, being a single-celled organism. *C. elegans*, however, has one of the best-documented nervous systems of any species. Both species move toward food using similar strategies: periods of swimming in a straight line followed by periods of rotation that change their direction of motion.

Biologists have demonstrated, however, that certain neurons are essential to chemotaxis in *C. elegans*, while *E. coli* carries out its behavior using signaling proteins embedded in the cell membrane. Does *C. elegans* represent the concentration gradient of food molecules differently than *E. coli*? Does

https://doi.org/10.1086/675679 Published online by Cambridge University Press

E. coli represent the gradient in any sense at all? What about their internal causal structure accounts for any differences?

The comparison provides a way for various notions of representation to differentiate themselves empirically. In keeping with the principles I gave above, the crucial question is what must be added to basic causal specificity and adaptiveness in order for a process to count as representation. Is there a certain kind of causal specificity involved, such as isomorphism? Are there distinctive benefits that representation generates compared with other cases of biological information? By comparing cases we can test whether a supposedly distinctive benefit of representation can indeed only be acquired through the associated kind of causal specificity. We can also generate new hypotheses in a bottom-up fashion by looking for generalizations connecting mechanistic structure and adaptive functionality across cases. We might find that biological information as an operational kind is host to a diversity of nested, independent, and crosscutting relationships between varieties of causal specificity and adaptive benefit.

8. Conclusion. I have argued that biological information is a kind of empirical phenomenon that can be identified using diagnostic procedures, and I have described how to theorize about information using a bottom-up, comparative method. This stands in contrast to the classical top-down, theory-centric approach as well as recent accounts based on scientific modeling that have continued some aspects of the theory-centric tradition.

Criticism of the transmission account as irrelevant to questions about semantics has things exactly backward: it is precisely because we can define biological information in a minimalist fashion that it can become the subject of empirically-informed theorizing that addresses these questions in a neutral manner. I have sketched a way that one could rethink philosophical theorizing so that inquiry into the semantic aspect of biological information can continue even when full-blown positive theories are missing or unsatisfactory. In this manner, theorizing in the comparative mode could reunite the philosophical commitment to naturalism with the empirical ethos of the naturalist.

REFERENCES

Alon, Uri. 2007a. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Vol. 10. Boca Raton, FL: CRC.

———. 2007b. "Network Motifs: Theory and Experimental Approaches." *Nature Reviews Genetics* 8 (6): 450–61.

Bechtel, William, and Robert C Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

Bergstrom, Carl T., and Martin Rosvall. 2011a. "Response to Commentaries on 'The Transmission Sense of Information.'" *Biology and Philosophy* 26 (2): 195–200.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

———. 2011b. "The Transmission Sense of Information." *Biology and Philosophy* 26 (1): 159–76.
Boogerd, Fred C., Frank J. Bruggeman, Jan-Hendrik S. Hofmeyr, and Hans V. Westerhoff, eds. 2007. *Systems Biology: Philosophical Foundations*. Amsterdam: Elsevier.
Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. New York: Oxford University Press.
Dekel, Erez, Shmoolik Mangan, and Uri Alon. 2005. "Environmental Selection of the Feed-Forward Loop Circuit in Gene-Regulation Networks." *Physical Biology* 2 (2): 81–88.
Dretske, Fred I. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
Giere, Ronald N. 1999. *Science without Laws*. Chicago: University of Chicago Press.
Godfrey-Smith, Peter. 1996. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
———. 2000. "On the Theoretical Role of 'Genetic Coding.'" *Philosophy of Science* 67 (1): 26–44.
———. 2007. "Information in Biology." In *The Cambridge Companion to the Philosophy of Biology*, ed. David L. Hull and Michael E. Ruse, 103–19. Cambridge: Cambridge University Press.
———. 2010. "Senders, Receivers, and Genetic Information: Comments on Bergstrom and Rosvall." *Biology and Philosophy* 26 (2): 177–81.
Griesemer, James R. 2012. "Formalization and the Meaning of 'Theory' in the Inexact Biological Sciences." *Biological Theory* 7 (4): 298–310.
Griffiths, Paul Edmund. 2001. "Genetic Information: A Metaphor in Search of a Theory." *Philosophy of Science* 68 (3): 394–412.
Griffiths, Paul Edmund, and Karola Stotz. 2013. *Genetics and Philosophy: An Introduction*. New York: Cambridge University Press.
Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.
Jablonka, Eva. 2002. "Information: Its Interpretation, Its Inheritance, and Its Sharing." *Philosophy of Science* 69 (4): 578–605.
Jablonka, Eva, and Marion J. Lamb. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic*. Cambridge, MA: MIT Press.
Jantzen, Benjamin, and David Danks. 2008. "Biological Codes and Topological Causation." *Philosophy of Science* 75 (3): 259–77.
Leonelli, Sabina, and Rachel A. Ankeny. 2012. "Re-thinking Organisms: The Impact of Databases on Model Organism Biology." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 29–36.
Levy, Arnon. 2011. "Information in Biology: A Fictionalist Account." *Noûs* 45 (4): 640–57.
Maclaurin, James. 2010. "Commentary on 'The Transmission Sense of Information' by Carl T. Bergstrom and Martin Rosvall." *Biology and Philosophy* 26 (2): 191–94.
Mangan, Shmoolik, and Uri Alon. 2003. "Structure and Function of the Feed-Forward Loop Network Motif." *Proceedings of the National Academic of Sciences* 100 (21): 11980–85.
Mangan, Shmoolik, Alon Zaslaver, and Uri Alon. 2003. "The Coherent Feedforward Loop Serves as a Sign-Sensitive Delay Element in Transcription Networks." *Journal of Molecular Biology* 334 (2): 197–204.
Maynard Smith, John. 2000. "The Concept of Information in Biology." *Philosophy of Science* 67 (2): 177–94.
Maynard Smith, John, and Eors Szathmary. 1995. *The Major Transitions in Evolution*. New York: Oxford University Press.
Mayo, Deborah G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
———. 2004. *Varieties of Meaning: The 2002 Jean Nicod Lectures*. Cambridge, MA: MIT Press.
Morgan, Mary S., and Margaret Morrison, eds. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.
O'Malley, Maureen A. 2011. "Exploration, Iterativity, and Kludging in Synthetic Biology." *Comptes Rendus Chimie* 14 (4): 406–12.
Oyama, Susan. 2000. *The Ontogeny of Information: Developmental Systems and Evolution*. 2nd ed. Durham, NC: Duke University Press.

https://doi.org/10.1086/675679 Published online by Cambridge University Press

Pfeifer, Jessica. 2006. "The Use of Information Theory in Biology: Lessons from Social Insects." *Biological Theory* 1 (3): 317–30.
Sarkar, Sahotra. 1996. "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology." In *The Philosophy and History of Molecular Biology: New Perspectives*, 187–232. Norwell, MA: Kluwer.
———. 2005. "How Genes Encode Information for Phenotypic Traits." In *Molecular Models of Life: Philosophical Papers on Molecular Biology*. Cambridge, MA: MIT Press.
Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
Shavit, Ayelet, and James R. Griesemer. 2011. "Transforming Objects into Data: How Minute Technicalities of Recording 'Species Location' Entrench a Basic Challenge for Biodiversity." In *Science in the Context of Application*, ed. Martin Carrier and Alfred Nordmann. New York: Springer.
Shea, Nicholas. 2010. "What's Transmitted? Inherited Information." *Biology and Philosophy* 26 (2): 183–89.
———. 2013. "Inherited Representations Are Read in Development." *British Journal for the Philosophy of Science* 64 (1): 1–31.
Shea, Nicholas, Ido Pen, and Tobias Uller. 2011. "Three Epigenetic Information Channels and Their Different Roles in Evolution." *Journal of Evolutionary Biology* 24 (6): 1178–87.
Skyrms, Brian. 2010. *Signals: Evolution, Learning and Information*. Oxford: Oxford University Press.
Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39." *Social Studies of Science* 19 (3): 387–420.
Stegmann, Ulrich E. 2005. "Genetic Information as Instructional Content." *Philosophy of Science* 72 (3): 425–43.
———. 2013. "On the 'Transmission Sense of Information.'" *Biology and Philosophy* 28 (1): 141–44.
Strasser, Bruno J. 2012. "Data-Driven Sciences: From Wonder Cabinets to Electronic Databases." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 85–87.
Taylor, Samuel F., Naftali Tishby, and William Bialek. 2007. "Information and Fitness." Unpublished manuscript. http://arxiv.org/abs/0712.4382.
Walsh, Denis M. 1996. "Fitness and Function." *British Journal for the Philosophy of Science* 47 (4): 553–74.
Weisberg, Michael. 2007. "Who Is a Modeler?" *British Journal for the Philosophy of Science* 58 (2): 207–33.
Wimsatt, William C. 2006. "Reductionism and Its Heuristics: Making Methodological Reductionism Honest." *Synthese* 151 (3): 445–75.
———. 2007. *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.
Woodward, James. 2010. "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology and Philosophy* 25 (3): 287–318.

https://doi.org/10.1086/675679 Published online by Cambridge University Press