Biosemiotics. Commentary on the target article by Terrence Deacon

Data and context

Thomas E. Dickins
Faculty of Science and Technology
Middlesex University
London, NW4 4BT, U.K.

t.dickins@mdx.ac.uk
http://tomdickins.net/

Abstract Deacon presents a fascinating model that adds to explanations of the origins of life from physical matter. Deacon’s paper owes much to the work of Howard Pattee, who saw semiotic relations in informational terms, and Deacon binds his model to criticism of current information concepts in biology which he sees as semantically inadequate. In this commentary I first outline the broader project from Pattee, and then I present a cybernetic perspective on information. My claim is that this view of information is already present within biology and provides what Deacon seeks.

Keywords Epistemic cut · information · data · context · idealization · analogy

Introduction: The epistemic cut and information

Deacon’s target article is dedicated to Pattee’s 1969 paper about communication within biological systems (Pattee 1969). In that paper, and others, Pattee develops a biosemiotic perspective for tackling questions in theoretical biology. A principal contribution has been the concept of the *epistemic cut*, which describes what must be done to develop an objective understanding of a system. For Pattee, systems are material, and their physics is dynamic, whilst understanding is expressed in symbolic terms. Pattee asks how anything expressed in linear symbolic terms can objectively represent a dynamic system.

The epistemic cut is an issue of measurement (Pattee 2001). If we are to measure the initial starting conditions of some parameter within a dynamic system, S, then the measurement we choose to use, M, is also describable in terms of the same fundamental laws as S. If we were to treat S+M as a compound, due to this commonality of governance, M would lose its function as a metric and a new measure would be required. Put another way, we cannot describe the function of measurement, M, in the terms of dynamical laws. Something must be done to cut the link between S and M, such that M can provide understanding. For scientists, the epistemic cut creates a point from which to observe. It is an arbitrary decision with respect to reality and conveys no ontological distinctions.

In his 2001 paper Pattee also discusses how the epistemic cut is a feature of nature. For example, in fixed bodies such as crystals, geometric forces vastly reduce the dynamic possibilities for atoms within the structure, producing rigid objects. Other constraints are flexible and enable articulated assemblies of rigid objects, directing available energy to do work. These are biological mechanisms (Bechtel and Bich 2021). There are also assemblies of less rigid objects that are described by Pattee as labile, e.g., biopolymers. Biopolymers include the polynucleotides RNA and DNA as well as polypeptides which are folded to construct three dimensional proteins. These constraints act to control the overall system by adding degrees of freedom, by adding different possible outcomes. This may appear counter intuitive, but Pattee discusses how such variables change rates and ranges of response in system variables to achieve control. This technically adds more options to the system.

Pattee discusses the role of molecules in signalling to a system the requirement to turn <on> or <off>, a situation he characterizes in terms of the mechanism of a switch (Pattee 1969). He notes that this mechanism can be accounted for symbolically in terms of a conditional architecture, but only makes functional sense in the context of a larger system of constraints and he draws an analogy between this larger system and a language. His project is to understand the origins of this language during the emergence of life, and he focuses his definition of life around replication.

The epistemic cut is another way of capturing the relationship between thermodynamic and cybernetic information (Avery 2012). The thermodynamic concept of information is related to entropy in physical systems, such that maximum disorder equates to zero information. This is related to measurement, famously through Maxwell’s Demon who required information about the speed of particles to sort the fast and slow into separate chambers. Maxwell’s overall system can be in different states of orderliness, as a function of the energy flow within it, but the establishment of order (opposed to entropy) carries an energetic cost which we might conceive as a constraint (Deacon 2017). Avery also characterizes life as a set of constraints in biological systems that direct available energy to do work, creating order in a universe otherwise predisposed to entropy. Pattee’s conditional switching architectures embody cybernetic information, which is sometimes also termed semiotic information. This becomes intuitively clear when we think of switches as signalling <on> or <off>, and codons as directing amino acids to form a polypeptide chain.

For many in the biological sciences, semiotic usage is an analogy that derives its utility from a perceived qualitative similarity between symbols and DNA code (Maynard Smith 2000). Whilst Pattee adopted Peircean semiotic terms to categorize constraints he was firmly of the opinion that all semiotic relations were informational (Pattee 2013). This meant Pattee was looking for a relationship between thermodynamic and cybernetic information to explain minimal life. For Pattee this is not an analogy, but a research strategy.

Deacon's information

Deacon's target article is a contribution to the broad project articulated by Pattee and he begins by critically inspecting information concepts in biology (Deacon 2021). He claims that genetic information lost the property of *aboutness* when it was reformulated as a template pattern to be copied, with the action of copying regarded as interpretation. Deacon suggests that this outcome was acceptable for a materialist science that did not want to introduce more metaphysical versions of information, enabling information to be understood in terms of Shannon's theory of communication (Shannon 1948). As Deacon notes, Shannon's problem was one of reproducing a message transmitted via a communication channel to a high level of fidelity. The content of the message is irrelevant to this task. This, Deacon claims, is a concept of replication with near identity to that of Dawkins' selfish gene (Dawkins 1989). He further notes that Shannon did not see his work as information theory, because information is *about* something, drawing a clear distinction between communication and information.

> But Dawkins makes no such distinction. Unlike Shannon's "engineering problem," however, the "biological problem" cannot be adequately addressed without taking into account the function of molecular information. A physical pattern by itself is not *about* anything. The sequence of nucleotides in a DNA molecule is just a molecular structure considered outside the context of a living cell. For structure to be *about* something there must be a process that interprets it. And not just any process will do.
>
> So, is replication such a process? (Deacon, 2021: 2)

Deacon is in alignment with Pattee. A switch is not really a switch without a broader, systemic context, and DNA cannot be a sign-vehicle unless it refers to something. But Deacon has treated Dawkins with some brevity here. Dawkins developed the replicator-vehicle distinction to emphasize the transmission of information – what he referred to as the *idea* of the organism, the design principles – across generations. DNA has the properties of copying fidelity, fecundity and longevity making genes the fundamental units to enable natural selection (Williams 1996). To that end DNA replicates both within and between organisms. However, Dawkins was also very clear that genes acted to catalyze development and were not to be regarded as blueprints for the construction of an organism but instead as early stage, necessary conditions for development (Dawkins 1989: 240). What this means is that genes enter a developmental system that responds to those inputs, and others, in a systematic fashion. Thus, Dawkins' model includes both reference (the idea of the organism) and interpretation (protein synthesis and further downstream developmental processes). Whilst he did not package his distinction semiotically his view is consilient with such usage.

Adopting a principle of charity, Deacon's point is best understood as a call to inspect fundamental assumptions. Dawkins is clearly not antagonistic, at least in theory, to a semiotic take, but this does not mean we have a full view of how replication might deliver meaning and it is to this that Deacon directs his attention.

Deacon (2021: 3) emphasizes the importance of interpretation for meaning. He does this with his central dogma of semiotics which states that:

> Any property of a physical medium can serve as a sign vehicle of any type … referring to any object of reference for whatever function or purpose because these properties are generated by and entirely dependent upon the form of the particular interpretive process that it is incorporated into.

This is a semantic view of information, hence the emphasis upon interpretation and aboutness. But this is not completely divorced from Shannon's theory of communication and nor is information entirely about semantics, especially if we adopt a more formal approach to cybernetic information which is appropriate given Pattee's project.

Floridi gives the example of a computer awaiting the outcome of a fair coin toss (Floridi 2010). Prior to the toss the computer is in a state of data-deficit, which Shannon termed a state of uncertainty. What uncertainty means is that the system can be in *n* states, but a precise state, *S*, has yet to be determined. This will be determined by an input or datum. In this case the input will be the outcome of the toss. Tossing the coin produces an amount of information that is a function of the two equiprobable outcomes, <heads> or <tails>. This is, in this case, 1 bit of information, and is equal to the data deficit it removes. (It is technically a measure of uncertainty. Think of the number of *yes* or *no* questions needed to determine which side up the coin had landed after a toss. <Is it heads?> <No.> This interaction resolves the uncertainty.)

As Floridi notes, Shannon's basic idea was that information can be *quantified* in terms of the reduction of uncertainty, and this quantification helped to resolve his engineering problem. But quantification does not tell us what information is. This becomes clear when we realize that one can receive two equal amounts of information about two entirely separate objects. (Compare asking <can I have potatoes with that?> and <is it a girl?>. Both can be answered with a <yes> or <no>, yielding 1 bit of information.) Knowing the number of bits of information received does not help us to understand what role the information might play. Floridi enforces this point using a general definition of information which states that:

Information = data + meaning

This notation implies that data must conform with the semantics of the system it enters to be considered informative, thus information is in fact the functional outcome of a relationship between data and semantics. From this Floridi characterizes the quantification of information as:

Information – meaning = data

And he concludes that Shannon's mathematical theory of communication is in fact a theory of data communication.

Floridi reinforces his general definition by looking at the role of queries:

<is it a girl?> + <yes>

This has the format of query + binary answer (1 bit). The binary answer is a datum that, in Floridi's terms, unlocks the information contained within the query. Floridi's favored definition of factual semantic information states that something can only be considered so *if and only if* it is constituted by well-formed, meaningful, and veridical data (2010: 50).

A reason to separate data from information is the fact of cryptography (Boisot and Canals 2004). We might download a data set to find it useless due to encryption. Only once we have the key can we de-encrypt it and then find the data informative. That informational value will come because of rendering the data usable within a particular context. Given this idea, I prefer to replace Floridi's *meaning* with the term *context* in the general definition of information:

Information = data + context

This now appears closer to Pattee's view of the context of the larger system, and by default Deacon's central dogma of semiotics. But what this view does is to strip away the commonly adopted objectification of information. Floridi's definition does this also. Under this view, information is not something to be transmitted but rather the outcome of a relationship between data and context. Information is a function, not a thing. I believe much recent criticism of informational concepts in biology has been down to a colloquial use of information that reifies it and fails to understand its functionality. So, for example, when a biologist talks about genetic information it is readily assumed that she is discussing the gene as a total source of structure, in the sense of a blueprint. Combine this with the fact that genes are replicated, and one quickly assumes standard theory is committed to a hermetically sealed view of the gene as the source of all – its own data and its own context. But, as I have already pointed out, Dawkins who is often portrayed as the arch gene-centrist by critics of evolutionary theory, was committed to a systemic, contextual view of the role of the gene. For Dawkins, genes are data that make no sense outside their context. When in context, the overall biological system can be informed.

This view contains an important commitment. For data to be informative, the system[1] into which it is inputted must be prepared to respond to it. Input alone is insufficient; there must be a systemic effect by which I mean state change. Here Shannon's view of uncertainty is again relevant. We can conceptualize the role of data as reducing the uncertainty of the system. At the level of constructing a polypeptide chain there is uncertainty about the next amino acid to be added, and that is resolved by the arrival of the ribosome at the appropriate codon. That codon is data, its context in this case is the chain building process, and the resultant polypeptide chain is the outcome of that functional relationship. This view requires a theory of design to explain the regularities and relationships, and that is a key role of evolutionary theory. The data are not about anything, no information is transmitted; but one might say that the resultant state change is a consequence which we might term meaningful. The biological system, so described, is well formed, and contains veridical data.

Biosemiotics scholars might not wish to fully embrace the view I have extracted from Floridi[2]. One reason for this is the notion that the *states* of a system can only be interpreted as observer dependent (i.e., something measured, following Pattee). Pattee sought to draw a direct link between measurement and control, and my reading of his work is that he saw control as naturally arising. This implies that there is no requirement for an observer to determine states, but rather for control variables so to do. Natural scientists will try to observe real states and of course those efforts may not precisely match reality. To that end, evolutionary theorists would look to evolutionary processes to supply an account of persistent control and as is widely known, it is not uncommon to use agent metaphors to package natural selection as a designer and hence a kind of observer. Again, Dawkins has addressed the explicit removal of natural theology from natural philosophy, which was a removal of agency as a metaphysical concept (Dawkins 1986). Evolutionary theory has precisely no need for it, but cybernetic information captures the statistical outcomes of evolution by natural selection, or rather natural selection is a method of creating biological information (Avery 2012)[3]. This should not be read as an exclusive statement, as other natural processes may also achieve this, and this is what both Pattee and Deacon are focused upon.

[1] Here I am using system and context interchangeably.
[2] I am indebted to the editor and an anonymous reviewer for raising these concerns.
[3] See also Dickins, T.E. (Forthcoming.) *The Modern Synthesis: Evolution and the organization of information*. Springer.

Another sticking point might be the notion of context which is more usually interpreted as something external to the organism, and that is captured by sign relations that impact upon the internal economy of the organism. The view of context I am adopting is simply that of a biological mechanism, which can be at any scale, understood as a set of processes that direct available energy to do work. Again, this is in accord with Pattee, and the relation between thermodynamic and cybernetic considerations. Organisms consist of mechanisms linked in a heterarchical manner, but we can regard organisms as unitary systems within an external context when doing certain kinds of science. In keeping with ethological (and ecological) perspectives, the organism sits within an *umwelt* determined by the evolved nature of its various mechanisms and their relations. This is hardly anathema to biosemiotics. But my derived view also places all biological mechanisms, at all levels within an *umwelt* that consists of potential data interactions with other mechanisms. Thus the outputs of one mechanism can be the data for another.

A third concern is that of clarifying what data are because under the current account it might appear entirely dependent upon systemic context. My view is that *data* are stimuli that are either usable or unusable. Both kinds are *potentially* usable physical stimuli emanating from the world, but processes such as evolution are required to realize that potential by creating systems that can take such stimuli as inputs. This is not unrelated to the minimally mechanistic view from Boisot and Canals (2004). They advocate a filter view, for example a semi-permeable cell membrane permitting certain ion transfers and not others can be seen as a filter. Such filters control which stimuli can enter a system, and thus which stimuli can affect state changes. Passing the filter makes stimuli into *usable data*. The existence of the filter is an outcome of material process and selection. Thus, data are everywhere but not all data are usable due to the absence of appropriate contexts. This again reinforces the idea that information should be regarded as the outcome of biological processes, and that those processes are synonymous with meaning, or rather meaning is nothing more than this. Here I am in full alignment with Deacon, context is all and in explaining the origins of context we explain which data have been adopted by biological systems and their informative role.

Conclusion: Back to the cut

The preceding discussion about cybernetic information does not undo the ambitions of Pattee or Deacon, and nor was that my intention. But it does call into question Deacon's assertion that molecular replication is problematic for the origins of information. He (2021: 4) makes this statement:

> The problem with the "naked replicator" approach… is that replication isn't *about* anything, nor does it contribute to anything except increasing numbers of similar objects. And although there can be something analogous to "selection" eliminating modified sequences that fail to replicate, the "external" environment does all the work. Replicating molecules are passive artifacts. They don't actively adapt to their environment, and so their structure does not contain or acquire information about the environment and they not have any intrinsic disposition to correct "errors" because the very concept of error has no intrinsic meaning. There just is what gets copied and what doesn't, and whether something gets copied or not is only interpretable as success or failure from an external observer's point of view.

Clearly, I would not use the term *information* in this statement, but might replace it with *data*, because information is not something to be harvested. Moreover, data is perhaps best understood as potentially usable stimuli (Boisot and Canals 2004) and I might repackage the statement to the effect that molecular replicators do not acquire stimuli from outside of their own replication. All this to one side, however, I most certainly would not base a theory of biological information on replication[4]. Replication enables

[4] And to the best of my knowledge no one in fact does this. Instead, the tendency in evolutionary biology is to associate a colloquial view of information with the gene as an analogy, with no formal commitment to a theory of information. This is not a theory of information, and as an analogy allows only one direction of epistemic travel. This works, but only within limits (Maynard Smith 2000).

data to be preserved, and this plays a key role in facilitating natural selection. Once this is clarified Deacon's concerns evaporate.

But Deacon's ambition runs deep, and his last sentence above returns us to Pattee and the epistemic cut. This final criticism from Deacon reveals his interest in locating information within biological systems, and in this way locating meaning within molecules. He aims to replace what he sees as the informational assumptions of replicator theory with something more resolutely grounded in material reality.[5]

Deacon's proposal is a model system based on virus structure that uses autocatalysis and self-assembly, which he claims as a variant of crystallization. Importantly the two processes create the conditions for one another. This enables some damage repair and the creation of an autogenic workflow for energy between the processes. Immediately you will note constraint, in keeping with Pattee's arguments, and Deacon cites Kauffman's work on this concept. From this fascinating model he draws five irreducible, emergent properties, which I extract below:

1. *Individuation*: the maintenance of a self/non-self distinction
2. *Autonomy*: the maintenance of boundary conditions
3. *Recursive self-maintenance*: the system repairs and replicates its own boundary conditions
4. *Normativity*: its disposition is to do 1-3 but it can fail
5. *Interpretive competence*: it represents its own boundary conditions anew and so reproduces its conditions of existence.

He sees this system as a ground-zero semiotic process that interprets "the most basic semiotic distinction" between self and non-self, such that disruption to integrity is "a sign of non self and the dynamics that ensues and reconstitutes the stable state is the creation of an interpretant which actively reconstructs this self/non-self distinction." For Deacon this is a version of iconicity, the fundamental Peircean semiotic process incorporating distinction. This model forms the basis for further semiotic derivations, but this is where I shall focus my remaining comments.

A pervasive problem for biosemiotics is the inevitable question to such proposals – *for whom is the sign produced?* This might feel facetious but in fact relates to the kind of concerns associated with homuncular functionalism and the problems of the "Cartesian theatre" (Dennett 1991). Here we can again revert to Pattee and note the similarity between the sign of failed integrity *and* data that simply flips a switch (which has functionality within the broader context of the system). There is a causal energetic story about the production of the data, and one for its effect within the system. We then see Deacon's account for what it is, and that is a neat model system for the minimal conditions for life, under some definitions. But it is not an account of information; rather, it is a model that conforms to the theory of cybernetic information as a relation between data and context. Deacon wanted context to be to the fore, and it is; but it always has been within the kind of information theory at work in evolutionary biology.

As with Pattee, I think information is a fundamental aspect of the physical world. It follows that information reveals itself as we do science. The structure of evolutionary biology conforms to information theory for that reason, but this does not mean biologists have typically made efforts to formally account for information. Instead, the looser, analogical uses of the term have been used effectively as idealizations to capture causal complexity. Idealizations can contain untruths, and perhaps inaccuracies, to deliver

[5] We should remember that his main assumption is that replicator views of information disposed of the concept of aboutness. He laid this at Dawkins' door, which I hope to have dissuaded the reader from committing to, but in my parse of information and data I did note that data are not about anything. This was in order to place the burden of information on the functional relationship between data and context. This may be where Deacon is heading, but I don't want him to ride the back of straw biologist.

scientific understanding, and this is perhaps why an informal usage is so pervasive (Potochnik 2020). Nonetheless, Dawkins and others have given the detail of the role of genetic data in developmental contexts in a way that is entirely consilient with the more formal view. Given this, I feel Deacon’s claim about the denuding of information is wrong.

What remains? Both Deacon and Pattee are really focused on the emergence of life and the role of replication in that. Deacon has directly tackled the issue of the epistemic cut from a materialist biological perspective and presented a plausible model for this purpose. That model can be expressed semiotically, but it does not have to be. For both Deacon and Pattee semiotics is more than mere analogy and are technical items that capture key kinds of constraints put upon dynamic systems to do work. This is more than idealization for them. I am currently reserving judgement on this issue. One reason for my caution is a sense that the adoption of Peircean semiotic terms runs counter to the nominalist tradition in science. Cybernetic approaches to information provide useful formalisms for thinking about systems, but the full array of data + context relations is unknown. Seeking the specific relations of semiotics within biology runs the risk of essentialism, or assuming the joints of nature rather than naming them once uncovered (Popper 1945). It is potentially limiting.

Acknowledgement

I am grateful to Paul Cobley, the editor and an anonymous reviewer for critical and useful engagement with a draft of this commentary. They have improved it, but all errors remain my own. I declare no conflicts of interest and that I received no funding to support this work.

References

Avery JS (2012) Information Theory and Evolution, Second Edition. World Scientific Publishing, Singapore
Bechtel W, Bich L (2021) Grounding cognition: heterarchical control mechanisms in biology. Philos Trans R Soc B Biol Sci 376:20190751. https://doi.org/10.1098/rstb.2019.0751
Boisot M, Canals A (2004) Data, information and knowledge: Have we got it right? J Evol Econ 14:43–67. https://doi.org/10.1007/s00191-003-0181-9
Dawkins R (1989) The Selfish Gene, Second Edition. Oxford University Press, Oxford
Dawkins R (1986) The Blind Watch Maker. Penguin Books, London
Deacon TW (2017) Information and Reference. In: Dodig-Crnkovic G, Giovagnoli R (eds) Representation and Reality in Humans, Other Living Organisms and Intelligent Machines. Springer International Publishing, pp 3–15
Deacon TW (2021) "How molecules became signs”. Biosemiotics 1–21
Dennett DC (1991) Consciousness Explained. Allen Lane: The Penguin Press
Floridi L (2010) Information: A Very Short Introduction. Oxford University Press, Oxford
Maynard Smith J (2000) The concept of information in biology. Philos Sci 67:177–194. https://doi.org/10.1017/CBO9780511778759.007
Pattee HH (1969) How Does a Molecule Become a Message? Dev Biol Suppl 3:1–16. https://doi.org/10.1007/978-94-007-5161-3_3
Pattee HH (2001) The physics of symbols: Bridging the epistemic cut. BioSystems 60:5–21. https://doi.org/10.1016/S0303-2647(01)00104-6
Pattee HH (2013) Epistemic, Evolutionary, and Physical Conditions for Biological Information. Biosemiotics 6:9–31. https://doi.org/10.1007/s12304-012-9150-8
Popper KR (1945) The Open Society and Its Enemies. Routledge
Potochnik A (2020) Idealization and many aims. Philos Sci 87:933–943. https://doi.org/10.1086/710622
Shannon CE (1948) A Mathematical Theory of Communication. Bell Syst Tech J 27:379–423
Williams GC (1996) Adaptation and Natual Selection: A Critique of Some Current Evolutionary Thought, Third. Princeton University Press, New Jesey