

Steps to a science of biosemiotics

Terrence W. Deacon

University of California, Berkeley, 2015

“Biosemiotics could not be reconciled with science, and the only way to introduce meaning in biology is the new approach that became known as Code Biology.”

— Marcello Barbieri

Introduction

The quote that opens this essay makes two claims. First, that biosemiotics as currently conceived is not a science and will never even be reconcilable with the biological sciences. And second, that shifting to a perspective based on a code metaphor will be able to successfully introduce the concept of meaning into biology. In this essay I will agree that the *current* state of biosemiotics indeed remains more in the domain of the humanities than in the natural sciences, however, I will argue that this doesn't preclude the possibility that its foundational principles can be made consistent with well-established principles in biology, chemistry, and physics. But I will also argue that the code metaphor is incapable of producing the result that is hoped, and in fact its superficial appearance of compatibility with reductionistic biology is likely to be a significant impediment to the development of a non-trivial concept of meaning (and with it the entire domain of teleo-functional concepts) in biology.

In summary, my claim is that the current use of phenomenologically derived semiotic terminology to rename and re-describe well-analyzed molecular and cellular processes guarantees that current biosemiotic theories will remain squarely within the humanities. In this form biosemiotic theories are unlikely to provide new insights that can contribute to research at the lab bench or even in theory. Additionally, I argue that the code metaphor only provides the illusion of teleo-functional explanation, when instead it perpetuates a form of methodological dualism that cryptically assumes that meaning is imposed by an unacknowledged outside mentality. This unacknowledged homunculus interprets biological mechanisms as though they are uninterpreted code-elements. This logic could equally well be applied to the interactions between the components of a computer or automobile engine. In this respect, the dilemma that undermines both efforts to legitimize biosemiotic theory is their failure to directly address or even acknowledge their dualistic assumptions while at the same time arguing that they can provide a scientific semiotic foundation for biology.

In this essay I argue that we ultimately need to re-ground biosemiotic theory on natural science principles and abandon the analogy with human level semiotics, except as this provides clues for guiding analysis. But to overcome the implicit dualism still firmly entrenched in the biological sciences requires a third approach that is neither phenomenologically motivated nor based on a code analogy. This approach must preserve the centrality of the concept of interpretation (that is ubiquitous in the phenomenological domain) and yet base it in biophysics and mathematical information theory. To accomplish this we must undertake a thorough re-examination of information theory to determine how it can be extended to deal with issues of real reference and functional significance. I argue below that this requires showing how the concept of entropy (as it is differently defined in thermodynamics and the information sciences) can be used to explain the relationship between information, meaning, and work. And that it also requires a radical expansion of dynamical systems theories to explain the physics of intrinsically end-directed processes. Together these developments are necessary to account for the capacity to

interpret immediate physical conditions as representing other displaced or as yet unrealized possible conditions. This is a necessary first step to making biosemiotics compatible with the information theoretic perspective that is currently dominant in physics, chemistry, and molecular biology.

How 'information' lost its meaning

In 1948 the term 'information' split in two, with one version remaining in public vernacular referring to what that which conveys news about the world and the other version reserved for technical usage referring to only those properties of a signal medium that make it able to provide this capacity. The second, technical coining of the term was a product of a brilliant effort to develop a method for mathematically analyzing and measuring signal processing properties by the Bell Labs scientist Claude Shannon. As he remarks in the introduction to his report in the Bell System Technical Journal:

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.” (p. 397)

The exact words he uses are significant. As the title (*A Mathematical Theory of Communication*) indicates he is not so much interested in information as in its communication: i.e. the transmission of a signal from one place to another. He makes this clear when he points out that considerations of meaning and reference are “irrelevant” to the problem he is considering. So when in the remainder of the paper he defines information in terms of a signal's statistical properties and the uncertainty of receiving any particular signal among those that are possible, he has explicitly set aside any discussion of semiotic properties. His interest was entirely focused on the “engineering problem.”

Shannon's analysis demonstrated that semiotic functions that are implicit in the common usage of the term 'information' can be set aside when considering the physical details of the transmission, storage, modification, and analysis of signals. This technical usage became the default standard interpretation during the subsequent development of the information technologies that currently envelop us. Besides its engineering usefulness, however, another more fundamental reason explains why it became accepted as the primary definition in the sciences. This is because it is entirely consistent with the methodological dualism that has been the tacitly understood operating principle of the natural sciences since the Enlightenment. Although seldom recognized as such, it is currently assumed to be an inviolable criterion for qualifying an explanation as scientific. To be scientific an explanation cannot make use of any mentalistic assumptions as determinative principles. This can be called methodological dualism because it simply brackets the use of explanatory accounts that assume the contribution of teleological, normative, or semiotic properties. Even when these concepts are invoked in discussions of biological functions or human behaviors they are presumed to be placeholders for incomplete physical accounts of the detailed mechanisms.

Two additional historical developments insured that this technically reduced conception of information became a primary explanatory concept in biology as well as in the cognitive and neurosciences. These were the discovery of the molecular basis for genetic inheritance and the

development of digital computing. Both developments have contributed the basis for the some of most significant and ubiquitous 21st century technologies. And both have shifted attention to material-mechanistic correlates of information at the expense of its semiotic properties. Because these less easily defined attributes can often be ignored this conception of information has largely overshadowed the semiotic conception in essentially all technical contexts.

When in 1956 (and again in 1970) Francis Crick proposed what he called the “central dogma” of molecular biology, he specifically used the term ‘information’ (though sometimes in scare quotes) to describe what was transferred from DNA to RNA to protein structure. He also specifically described the DNA molecule as a “passive template” for storing this information. This usage was entirely consistent with Shannon’s account of the “fundamental problem of communication” as involving reproducing a “message” construed as a signal pattern. It avoids any mention of reference, meaning, or even function (as did Shannon’s use), and yet was analogically consistent with the use of such text based terminology as code, transcription, translation, and copying, which soon became standard in the field. Though ultimately metaphorically derived, these descriptors of biochemical relations within the cell carry no necessary semiotic baggage, and none suggest a role for interpretation. Not surprisingly these terms are also relevant to cryptography, where the meaning of an encrypted message is unknown but where its string of characters may be carried by different media, or transcribed from form to form in the process of attempting to translate it into an interpretable form. Indeed, many of Shannon’s insights arose in the context of his work with cryptography during the Second World War. In that context analysis of the structure of a signal is the critical problem.

Throughout the war Shannon and his colleagues were involved in analyzing methods of encryption used for US communications. Parallel efforts in Britain to decipher the famous German Enigma code led to the development of computing technology that would later make extensive use of his theory. In order to reliably decipher this ingenious code, British intelligence assembled a veritable army of mathematically sophisticated analysts aided by this first digital computer. The vast majority of this analysis involved manipulations of a completely uninterpretable string of alphanumeric characters. During this process, it was known that the coded messages contained instructions for German commanders in the field and U-Boat captains. The challenge was to work out the elaborate and constantly shifting permutations that the Enigma machine produced. The vast amount of work to decipher these coded messages was taken up analyzing the astronomical combinatorial options that this encryption machine generated. This needed to be completed prior to even beginning to decipher its content. In contrast, reading the finally decrypted message was a trivial final step that any German speaker could accomplish. By analogy, Shannon’s contribution can be seen as concerned with analyzing this pre-interpretive phase, but applied to any given communication process, encrypted or not.

There is an interesting parallel with respect to the analysis of genetic information. In the many decades since Watson and Crick first described the structure of the DNA molecule, molecular biologists have continued to discover unexpected new kinds of “meanings” conveyed by DNA nucleotide sequences. Although the central dogma was originally defined in terms of information flow from DNA to RNA to protein structure, it turns out that perhaps as little as 1% of the total genome directly corresponds to the amino acid sequences constituting proteins, and yet, as much as 80% gets transcribed in to RNA molecules. These RNA molecules appear to have other regulatory functions (e.g. RNA interference effects, etc.) many of which may still be unrecognized. In many respects, then, analyzing the functional “messages” stored and transmitted in DNA molecules continues to be a decryption exercise that almost certainly hides

new surprises. There has even been bizarre speculation that a “message” might actually have been incorporated into the genetic code by ancient aliens in the form of highly improbable mathematical symmetries among codons and amino acids (e.g. shCherbak and Makukov, 2013).

The similarities between decryption analysis and DNA research go further than this. A variety of ingenious pattern-recognition algorithms have been developed that have helped to identify sequences within the non-protein-coding regions of the genome that likely convey functional information—even though the function may remain unknown. Implicit in Shannon’s original analysis was a means for correcting errors introduced by signal degradation. This involved the recognition that signals conveying a message (e.g. a functional signal) if sent redundantly will share a common pattern of statistical constraints due to their common source, whereas the noise that corrupts these signals will tend to exhibit statistical independence. Inverting this logic and considering natural selection as an error-correcting process that eliminates corrupting mutational noise over time it is possible to consider redundancy a clue to functionally useful information. For this reason the presence of conserved DNA sequences (i.e. sequences that are relatively invariant between genomes) can provide a clue to the location of functional signals. Thus, for example, the discovery of conserved non-coding sequences is for this reason presumed to be evidence of their functional significance (i.e. their meaningful contribution to organism survival and reproduction). This redundancy is an important clue that there is a definite relationship between signal properties and their functional correlates.

Recognizing this linkage between signal redundancy (more generally exhibited as pattern or statistical constraint) and function provides a first step toward the reintroduction of meaning into the theory of information. In many respects this involves a figure/ground shift in emphasis that is nicely captured by the following quote from Gregory Bateson.

“What I have tried to do is to turn information theory upside down to make what the engineers call “redundancy” but I call “pattern” into the primary phenomenon. . . . if a pattern is that which, when it meets another pattern, creates a third – a sexual characteristic exemplified by moiré patterns, interference fringes and so on – then it should be possible to talk about patterns in the brain whereby patterns in the sensed world can be recognized.” — Gregory Bateson (from a letter to John Lilly on his dolphin research, 10/05/1968, cited in Harries-Jones, in press)

This figure/ground reversal reflects two important aspects of Shannon’s analysis. First, his theory demonstrated that it is possible to correct any level of noise (below complete noise) corrupting a signal, by introducing the requisite amount of redundancy into the signal. Again it doesn’t matter whether the signal is interpreted or not. Accurate signals and noisy signals have distinct statistical properties that show up when it is possible to compare independently sent replicas of the same signal. This redundancy and coding theorem is related to the problem of reference because it is based on the realization that messages are conveyed by constraints in signal properties. So long as some level of statistical variety is reduced in the signal being received the receiver has reduced some uncertainty about what could have been sent. And reduction of variability produces pattern. This analysis also provided a way to determine how compressible a given signal is, since conversely this involves removing unnecessary redundancy. Second, it points out that information theory analyses tend to focus on the possible signal variety (or signal “entropy”) to measure information rather than on signal constraint, because of the

engineering importance of determining what could *possibly* be communicated rather than any particular instance of a communicated message.

This curious inversion actually reflects two complementary aspects of Shannon's measure of information: the information carrying capacity of a communication medium versus the amount of information conveyed by a particular received message using that medium. Indeed, they are necessary complementary aspects of information. As Robert Ulanowicz (2014) recently argued, the possible variety of a signal medium and the eliminated variety of a particular message are of opposite sign. Indeed, one of the early arguments among the two most important information theorists at the time—Claude Shannon and Norbert Wiener—was over the question of whether the measure of information is more accurately represented by a positive or negative sign. Ulanowicz implies that both are right. Each necessarily implies the other. Interpreting a signal message necessarily involves a comparison. In this context Bateson's point is that when information about something is conveyed it is the pattern (and thus the constraint on signal variety) that matters. Ulanowicz simply points out that this is implicit in Shannon's theory, though often overlooked. Information is neither intrinsic to a particular signal medium nor to a particular conveyed pattern that constitutes a given communication. Information is a *relational* property. This fact can be ignored if one's interest is not to interpret the reference conveyed by a given communication, but rather to analyze message carrying capacity in general, as is the interest of an IT engineer. But if our interest is to analyze the relationship between signal properties and what they communicate, understanding the basis of this complementary relationship becomes the critical challenge.

A tale of two entropies

Following a suggestion that apparently came from the mathematician John von Neumann, Shannon used the term 'entropy' to describe the measure of a communication medium's capacity to carry information; i.e. its possible variety of different states. The choice of a term that already had a long history in thermodynamic theory was motivated by the fact that this quantity was best specified by a mathematical expression that was identical to the expression proposed by Ludwig Boltzmann to describe the variety of possible "microstates" of an idealized gas in a container isolated from outside influences, i.e. its thermodynamic entropy. Entropy in this sense is a measure of the relative mixed-up-ness of these molecules. It was shown that the expression $\sum p_i \log p_i$ provided an unambiguous measure of the global sum of these possible states. Since determining the relative variety of states is also required to assess the informational capacity of a communication medium, that same expression applies to both. But beyond this abstract similarity the two concepts of entropy are associated with quite dissimilar processes and phenomena. This has led to many complaints that Shannon made an unfortunate choice of terminology. In the years since the publication of Shannon's theory considerable controversy arose between those who believe that the homonymy provides useful insights and those who think that it confuses incompatible properties. Though I agree that it is indeed a mistake to argue that these two uses of 'entropy' refer to the same property in these two quite dissimilar domains, and I also agree that thermodynamic entropy cannot be reduced to a special case of information entropy (as some have suggested e.g. Ben-Naim, 2008). I will nevertheless argue that it is the relationship *between* these two entropies—precisely because of their considerable differences—that plays the critical role in the explanation of reference and significance. I believe that bridging these two domains is the key to providing a sound physical science foundation for biosemiotics, and for semiotic theory in general.

To begin, we need to identify what distinguishes informational from thermodynamic entropy. For one thing, there is no informational analogue to Clausius' theorem that the total entropy of an isolated physical system can only increase during any transformation from one state to another (the 2nd law of thermodynamics). Information entropy is not generally understood in dynamical terms where one state spontaneously gives way to another. Nor is informational entropy associated with the concepts of energy and work. So they should definitely not be confused for naming the same property. I will therefore distinguish them by calling Shannon's measure *signal entropy* to distinguish it from *thermodynamic* entropy. But the idea that there is a link between information and thermodynamics has a long history. Ways of demonstrating such a linkage have been proposed in many forms from Maxwell's famous demon to Landauer's argument about the thermodynamic cost of information erasure. However, these approaches focus primarily on the energetic and thermodynamic "costs" of manipulating physical markers or taking measurements, and how this might or might not violate the 2nd law. But the possibility that reference and significance might also depend upon the relationship between these two entropies has largely gone unnoticed.

The concept of thermodynamic entropy can, of course, be applied to the physical features of an information medium. This is because an information-bearing medium is a physical medium, and any physical medium capable of conveying information must be able to exhibit different states. Any physical process, structure, or system that can be analyzed onto component states, each able to be assigned a probability of being exhibited, can in this way be described in terms of both its signal entropy and its thermodynamic entropy. The two measures will not be the same, even were the thermodynamic constant k in the Boltzmann formula to be ignored. This is because there are usually vastly more physical or thermodynamic differences of state possible than are ever actually used for their information-conveying role.

Shannon's analysis measured the amount of information potentially provided by a received signal (i.e. a message) as a function of the amount of uncertainty that the received signal removes. This can be measured as a difference between the prior or potential signal entropy of the communication medium and the signal entropy of the received signal.¹ This exemplifies the relational nature of information as Shannon characterized it. It means that information is not a simple intrinsic property of a signal or other physical sign, it is a relational property.

Interestingly, this relationality is a feature that both concepts of entropy share. The 3rd law of thermodynamics likewise asserts that thermodynamic entropy is a relative measure: a difference between states of a system. The relational nature of thermodynamic entropy wasn't fully appreciated until 1906 when Nernst augmented thermodynamic theory by establishing an absolute reference point at 0 degrees Kelvin. So both Shannon's analysis of information and thermodynamic theory depend on comparing current entropy with respect to some reference entropy state; i.e. what is minimally and/or maximally possible for a given system.

A received signal that exhibits constraint in its information entropy is more predictable even if it is not reduced to a single fixed value. Thus even a noisy signal reduces uncertainty by some amount, so long as it is not fully random and has a value of information entropy that is lower than the maximum possible. Noise, which is the corruption of a signal, increases uncertainty by reducing the constraint that was initially present in the signal. But noise is often the result of physical degradation of the conveying medium. This link between thermodynamic degradation of the communicative medium (or sign vehicle) and degradation of its information-

¹ It is not a mere coincidence that this is also the logic behind Bayesian analysis.

bearing capacity is evidence of an intrinsic interrelationship between signal entropy and thermodynamic entropy.

The distinction between signal and noise is not an intrinsic property of the signal medium, however. It is a distinction determined by an interpretive process. This is also an important clue to how an information-bearing medium can be linked to some non-intrinsic object, event, or property; i.e. to what it can be *about*.

Noise that is introduced into a signal changes the signal entropy of the message by increasing it and thereby reducing the difference between the maximum potential entropy of the medium and the reduced entropy of the message. Often, as in the case of radio transmission, signal degradation is due to an increase in thermodynamic entropy of the conveying medium as a result of interference with the transmission or reception process. These are factors extrinsic to the information-bearing medium itself, but which affect its physical attributes. Of course, the distinction between signal and noise is only a matter of perspective. For someone trying to repair a broken communication device, like a cell phone or television, the noise can become a signal about the source of the problem. Both the signal and the noise reflect the effects of something external to the communication medium itself.

The essential point is this: every medium for storing or conveying information is constituted physically and its distinguishable states are physical states. So any change in that medium's statistical physical properties (e.g. its thermodynamic entropy) also has the potential of changing its informational properties.

This is not a necessary relationship, since (as noted above) the distinguishable states used to convey information in any given case are inevitably a very small subset of the total range of different states that the physical medium can assume. However, because of its physicality, any change in the informational entropy of a given medium must necessarily also entail a change in its physical statistical properties. And, following the strictures of the 2nd law of thermodynamics, any physical medium will only tend to be in an improbable state if it has in some way been driven away from its more probable state by the imposition of physical work or else prevented from achieving it by some extrinsic restriction. In other words, the relationship between the most probable state of the medium and the observed state at any particular moment is a reflection of its relationship to its physical context. Its intrinsic statistical properties are therefore indirect clues to factors that are extrinsic to it and to which it could thereby refer.

Recall that according to Shannon's analysis the measure of information in a message is proportional to the reduction of the uncertainty exhibited by a received signal compared to the intrinsic uncertainty implicit in the potential entropy of the communication medium. So the reduction of the physical entropy of an information medium that otherwise obeys thermodynamic laws is proportional to a reduction of its Shannon entropy by some proportionality constant that determines what portion of its physical entropy is used for information purposes.

The relationship is also invertible. If the same medium is in its most probable state—e.g. in equilibrium with its surroundings or in its maximum entropy state—it can provide information about the *absence* of a given specific referent because no extrinsic influence has modified it. This is the logic of a burglar alarm not triggered. This relation to the ubiquitous thermodynamic tendency to change toward the most probable—e.g. equilibrium—state also implies that any change of signal entropy can refer. So even a medium that is maintained in a state that is predictably far-from-equilibrium—such as a medium that is undergoing constant predictable change—can provide reference to an influence that *increases* its entropy. This is the case for the repair-person using noise as a sign of equipment malfunction. If the signal carried by the

malfunctioning device is predictably constrained, as for example if carrying a spoken voice, its de-constraint (and thus increase in entropy) due to electrical interference is the source of reference to the source of interference. This link can also be a source of highly sensitive instrumentation in which its unstable far-from-equilibrium state is easily disrupted. The sensitivity of a metal detector takes advantage of this sensitivity. Electricity constantly flowing through a coil produces a stable electromagnetic field that can be easily disturbed by the presence of a conducting metallic object.

The fact that either an increase or decrease of entropy can provide reference indicates that the presence or direction of change isn't as important as the susceptibility of the medium to such an influence. So the capacity to refer is intrinsically related to the concept of physical work.

What constitutes an interpretive process?

The fundamental defining attribute of any representational (i.e. semiotic) relationship is that it involves a physically immediate sign vehicle (i.e. signal) that affects the behavior of an interpreting system with respect to some phenomenon that is displaced in space or time or abstraction and to which the sign vehicle is also in some way linked. This mediating relationship is not any intrinsic property of the sign vehicle, but (as shown above) a relational property between the sign vehicle and that to which it refers. But not all referential relationships involve a direct physical link between an information medium and an immediate physical object of reference. As the indirectness of the causal chain leading to a change in some informing medium increases or becomes more distributed, reference can likewise become quite abstract and displaced from physical immediacy. This doesn't mean that reference is lost or even ambiguous in such cases, but it does demonstrate that mere physical connectedness is not alone a sufficient determinate of reference. It is merely the source of the *possibility* of reference. The determination of which aspect of a complex causal chain is relevant (i.e. the significant object of reference) is a function of an *interpretive process*. Indeed, any physical medium or sign vehicle can be about almost anything else in any respect depending on the nature of the interpretive process. It is the organization of the process of interpretation that determines both the object of reference and the significance of that information.

Thinking about semiotic processes in human terms, the philosopher Charles Sanders Peirce argued that the mental processes that interpret signs are also signs. This requires that an explanation of interpretation cannot find its grounding in the concept of mind. From a Peircean point of view, then, Saussure's distinction between signifier (sign vehicle) and signified (the corresponding mental concept) simply begs the question of the nature of interpretation. In this view, a code or correspondence theory of meaning (such as is often drawn from the Saussurean paradigm) thereby simply maps one signifier to another—differences to differences—leaving what is signified by both, undefined. In many respects, then, code biology draws the implicit conclusion from this situation that a signifier-signifier relationship is not sufficient to constitute interpretation. But by denying interpretation to processes at the biochemical level on these grounds it also implicitly forces one of two conclusions about semiosis in general: either there must be some special self-interpreting feature that distinguishes between signifieds and signifiers in cognition or else interpretation is an epiphenomenal gloss of a purely physico-chemical process. In contrast, the Peircean approach locates interpretation in the process, not in the properties of any particular component of that process. But this requires us to identify what sort of process, as distinct from the typical physico-chemical processes can produce semiotic relationships.

In order to avoid treating thoughts as intrinsically meaningful termini of an interpretive process Peirce developed a concept of sign that involved three components: a sign vehicle (representamen), an object of reference, and an interpretant sign. This famous triadic conception of the sign relation (which has roots in medieval philosophy) avoids cryptically appealing to an unexplained mental homunculus to ground interpretation, but it introduces other challenges. Simply describing the signification relationship in terms of component sign elements or phases offers no account of what special form of process organization is necessary and sufficient to account for interpretation.

One way around the lack of an explicit account of the physical process constituting interpretation—occasionally found in Peirce’s own writings—is to describe the process of interpretation as “sign action.” Unfortunately, this merely begs the question by making the sign-relation itself a sort of quasi-agent or homunculus. Sign vehicles are qualities, artifacts, or events. They don’t “act” or do the work necessary to get themselves interpreted. And this role cannot be assigned to the interpretant either, since it too is just a sign vehicle, no more special than the sign vehicle that initiates the process. Peirce’s effort to avoid invoking an ineffable mental process that is intrinsically meaningful has inadvertently induced certain of his commentators to find cryptic ways to reintroduce it, such as assuming that signs themselves “act.” Indeed, it may even have motivated Peirce’s own later tendency toward a kind of objective idealism that locates “thought ... in the work of bees, of crystals, and throughout the purely physical world” (CP 4.551).² It is no wonder that many contemporary semioticians have embraced some form of dual aspect theory or some version of pansemiosis, where every physical process is also semiotic in character.

So although Peirce’s analysis avoids the fallacy of appealing to the unanalyzed black box concept of mind in order to ground interpretation, it nevertheless leaves us with the equally perplexing problem of explaining what sort of physical process constitutes an interpretive process. Peirce hints at an answer when he argues that what he calls the *final interpretant* is, in effect, a *habit* formed in response to the sign. To be more specific, he ultimately considers the generation of habit to be the basis for semiotic significance in general. It is, he says, the “essential function of a sign to render inefficient relations efficient—not to set them into action, but to establish a habit or general rule whereby they will act ...” (CP 8.332) In terms of the referential relationship described above, this means that the constraint embodied in a sign medium must in turn constrain the capacity of the interpreting system to perform adaptive work with respect to whatever the signal is about. It is therefore some distinctive form of habit or regularity intrinsic to an interpreting system that is critical.

The fact that reference involves a comparative relation between entropic states—both thermodynamic and informational—provides the next step in this analysis. Peirce insightfully captured this insight by recognizing that semiosis involves something like a comparison between signs in order to give rise to new information about an object of reference. Though it is seldom described in exactly these terms, his core proposal describes a process whereby the reference (object) of a given sign vehicle is generated when an additional sign vehicle (an interpretant) is produced. This allows us to draw a parallel to Bateson’s (quoted above) analogy between interpretation and a moiré pattern (see Figure 1).

² I will use the standard designation of location in the volumes of collected papers of Peirce where CP indicates the Collected Papers followed by volume number and paragraph number.

INSERT FIGURE 1 HERE

Thus, combining Peirce and Bateson, the mental version of an interpretation process might be described as follows: a perceived pattern (a sign vehicle) is interpreted by the generation of a mental pattern (an interpretant sign) that interacts with the perceived pattern to generate a third pattern (an additional sign that Peirce calls the *immediate object* of the sign). The question that remains is how the perceived pattern and the generated pattern link to something in the world that the perceived pattern represents (which Peirce calls the *dynamical object*). If this link can be established we will have made progress in determining how a physical process can provide reference to a displaced or yet to be realized feature of the world.

Interpretation requires one more critical property: it must in some way be addressed to a recipient and/or contribute to some end-directed process or function. In commonplace non-semiotic considerations of an interpretation process we typically ask: What determines that X (the sign vehicle) is about Y (the referent) *for* Z? In human experience Z is some agent with a mind, but more generally Z is some *beneficiary* of this information in the broadest sense, e.g. contributing to some function or purpose attributed to that beneficiary. It is an easy evasion to say that a beneficiary must be a person or organism, but as Peirce understood, this just postpones the explanation. What is it about living agents, such as persons, that makes them beneficiaries and endows them with interpretive power?

To qualify as a beneficiary an interpretive system must somehow exhibit both need and fragility. But what does this entail? Just being incomplete or physically fragile isn't sufficient. Nor is the simple tendency toward degradation that is the hallmark of the 2nd law of thermodynamics. But thermodynamics is relevant. A system that is susceptible to degradation *and is additionally organized to counter this tendency* is by definition organized with respect to beneficial or threatening conditions. This requires the capacity to *compare*. As discussed above, comparison is a defining feature of a message in information theory as well as of Peirce's and Bateson's conceptions of semiosis. For Shannon it is the uncertainty reduced by the receipt of a message in comparison to the maximum possible signal entropy (uncertainty) of the channel or medium that defines how much information is conveyed by a given message. For Peirce the comparison between a received sign and an interpretant sign produced in response to it generates reference to an object. For Bateson information involves the comparison of one pattern to another to generate a third.

Comparison requires a metric of some sort. Without this, comparisons cannot be evaluated. As we noted above, the so-called 3rd law of thermodynamics (mentioned above) was introduced to provide such a metric of comparison determined with respect to an unambiguous threshold: absolute zero. For an interpretive system the reference state for comparison must be internally generated. It must therefore have a "preferred" state with respect to which divergence from it can be assessed. A dynamical system, like an organism, (or a mind) that needs to actively maintain itself so that it is able to perform the work required to compensate for the effects of extrinsic disturbance or degradation will necessarily have a minimal threshold level of organization. This threshold is the ultimate reference state for semiotic comparison. Since a dynamical system also requires material and energetic resources from its environment in order to compensate for disruptive tendencies, a self-maintaining dynamical system must in some way implicitly negotiate exchange of resources across a self/other interface. This makes features of its environment potentially relevant to its persistence, and their potential influence measurable with

respect to this threshold value.

Autogenesis

To test the hypothesis that these properties are both necessary and sufficient to determine a process of interpretation we need a way to precisely compare processes that minimally include or exclude them, without introducing any extraneous unexplained features that might cryptically introduce what we hope to explain. So if we are to avoid simply assuming interpretation as a black box and instead hope to explain exactly how this capacity is generated, in both cells and brains, we need a completely explicit and simplified approach that leaves no relationship unaccounted for. In other words we need an account of the interpretive process that is more basic than even the simplest living cells, just not so simple that semiosis is absent. If we can then show that removing any of these attributes also eliminates semiosis we will have established a firm physical grounding for all higher order semiotic analyses.

To address this need I introduce an exceedingly simple thought experiment in the form of an empirically testable molecular system termed an *autogen* (defined as a unit system exhibiting a process I have called *autogenesis*; Deacon, 2012). Though far simpler than even the simplest bacterium, I believe that it enables us to identify the minimal necessary and sufficient conditions required to define interpretation in any form.

Autogenesis consist of two component molecular processes: 1) a small set of molecules that are reciprocally catalytic such that each member of the set interacts with substrate molecules to generate one of the other catalysts in the set resulting in all being generated; 2) molecules that tend to spontaneously self-assemble into a structure, similar to a microtubule or viral capsule able to contain other molecules within its interior; and 3) linkage between processes 1 and 2 such that the self-assembling molecules (2) are produced as byproducts of the reciprocal catalysis (1). These properties together create a high probability that (4) self-assembly of a container will tend to encapsulate a sample of the very catalysts that together make self-assembly likely.

These are well-known molecular processes found ubiquitously in living cells as well as in many inorganic processes. The chemistry and physics of these chemical processes have been well-studied both theoretically and empirically. So there are no unusual or mysterious properties assumed in this model system, which makes it empirically testable such that its predicted properties are falsifiable.

INSERT FIGURE 2 HERE

Reciprocal catalysis and self-assembly are each self-organizing dissipative processes. Each generates increasing local constraints as they also generate entropy more globally. The key feature is that the respective constraints generated by each self-organizing process serve as the permissive and regulatory boundary conditions for the other process. The self-assembling “side-product” produced by reciprocal catalysis thereby maintains local concentrations of the interdependent catalysts by enclosing them. Although the self-assembling “side-product” molecules produced by reciprocal catalysis will precipitate out of the local environment as they become bound to the growing container they are continually replaced in solution by ongoing reciprocal catalysis. Conversely, container formation prevents diffusion of the interdependent catalytic molecules, and since proximity is a critical boundary condition for reciprocal catalysis, being enclosed together in a self-assembling container preserves this proximity constraint even though catalysis is thereby halted by lack of free substrates. In conditions where there is a

sufficient concentration of substrates to maintain persistent catalysis and where the rates of catalysis and self-assembly are not too discordant, the regions of most rapid self-assembly will tend to be co-localized with regions of most rapid catalysis resulting in autogen formation. And if a closed inert autogen has its integrity disrupted, such that free catalysts are again introduced into an environment with free substrates, the process will be re-initiated, leading to self-repair and possible duplication of the original (i.e. replication). Thus, in a supportive environment this simple configuration will exhibit a tendency to generate, preserve, and reconstitute the very constraints that make this process likely to persist.

Such a system is literally autogenic because it changes its dynamical state in response to disruptive influences in such a way that it reconstitutes the state it was in prior to disruption. As a result, these interdependent processes collectively preserve and can replicate a higher-order formal (synergy) constraint, in the form of their reciprocally supportive and limiting boundary constraints. Since new molecules and new energetics will be involved in each cycle of dissociation and self-reconstitution, this higher order constraint is not any specific physical or chemical product. It is in this sense a kind of memory of the formal relation between the component processes that constitute the integrated unity of an autogen. The formal constraint embodied in and transmitted from one molecular medium to another is in this respect information that effectively “re-presents” and “re-members” the whole of which it is a part.

An autogenic system simultaneously completes a thermodynamic work cycle in order to complete a cycle of self-reconstitution (see Figure 3). In chemical thermodynamic terms the reciprocal catalytic process is endergonic in that it must be driven by extrinsic energetic gradients likely generated by catalytic lyses of complex substrate molecules. In contrast the self-assembly process is exergonic, in that heat (entropy) is produced into the environment as the capsid molecules coalesce into a lower energetic state. So completing a cycle of self-repair is a process with an intrinsic end toward which the system develops and which it preserves against dissolution. It initiates work to regenerate and preserve the formal constraint that makes this work possible.

INSERT FIGURE 3 HERE

A key ingredient (as noted above) is that the system needs to be organized to persistently maintain this organization against the constant tendency toward dissolution. Autogenesis achieves this by being both self-generative and self-repairing. Its “preferred” target state therefore has a sort of self-referential organization: it acts to preserve its capacity to preserve this same capacity. Indeed, I think that we can confidently describe this process as end-directed, or teleological in some general sense. Autogenesis also creates a dynamically preserved self/other distinction, in which the internal reproduction of self-reconstituting constraints creates a background with respect to which specific contextual factors can become significant. The transmission and preservation of the self-referential information that provides the basis for extrinsic conditions affecting this process to be assessed. The fixed reference point of a completed autogenic work cycle thereby enables “comparison” of relative usefulness and accuracy of correspondence. Assessment occurs, however, at a “population” level by processes resembling natural selection that differentially preserve better-fitted variants. In this way we can use formal and simulated versions of autogenesis to develop a measure of relative significance, in the form of “work saved.”

Though at first this model system may appear too simple and reductionistic, its implicit

end-directedness and self-preservation supply the critical properties needed to explain the dynamics of semiosis, as noted in the previous section of this essay. It is a periodically dynamic system that is maintained far from equilibrium that when disturbed away from its target state by some extrinsic influence initiates work to reconstitute that state utilizing material and energy from its environment. What is maintained over time through successive phases of disruption, reconstitution, and possible replication, is the synergy constraint between an autogen's component self-organizing processes, irrespective of the changing material substrates that embody this constraint. Indeed, this higher order constraint persists whether the autogen is inert and closed to its environment or dynamic with component structures and processes distributed within the local environment. With respect to a Shannonian conception of the information in a message, this constrained state of the physico-chemical "medium" compared to the range of relationships and processes that would otherwise be possible is the "information" making maintenance of autogenic organization far less uncertain. With respect to the interpretive analysis provided above, the dynamical process of reconstituting autogenic integrity responds to the external disruption (sign) by generating an intrinsically initiated action (interpretant) which results in the re-production of the intrinsic target state (immediate object) that embodies the adaptation of the autogenic system with respect to supportive environmental conditions (dynamical object). The higher order synergy constraint thereby provides reference to relevant aspects of the environment in terms of this autogenic property; the beneficiary of this interpretive process. This provides what amounts to a lowest level minimal interpretive ground upon which higher order forms of referential relationships can be built. None of these semiotic properties are realized if this systemic synergy is absent. Semiosis is thereby defined without reducing its defining referential and normative properties to simple chemistry and physics and without needing to explain them away.

Variations on a minimalistic interpretive process

In order to address more complex referential relationships we can also explore two modifications of the autogenic model system. The first is what can be described as an *adaptive autogen*. This involves an autogen with a shell whose surface tends to bind external catalytic substrate molecules and where the integrity of the shell becomes more fragile with increasing numbers of bound substrates (depicted in figure 4). Populations of autogens of this type, with varying binding and fragility levels, will tend to differentially preserve those variants that tend to break open in environments with optimal substrate concentrations. This provides selective referential information about two major classes of environmental conditions and their relative significance to system ends (i.e. self-repair and self-preservation).

INSERT FIGURE 4 HERE

A second variant of autogenesis involves template-based information such as in DNA (see Figure 5). Template-assisted autogenesis becomes important as autogenic complexity increases. This is because as component molecular diversity increases the number of possible cross-reactions between them (whether in catalysis or self-assembly) increases geometrically. Since only a highly constrained subset of these reactions will be conducive to the reconstitution of autogenic integrity, the increasing diversity of alternative interaction possibilities decreases viability. Limiting this variety thereby becomes the critical prerequisite to the evolution of increasing autogenic complexity, and with it increasing the variety of possible adaptive

responses. A template molecule can provide this essential additional constraint.

Figure 5 depicts the way that differential binding of catalysts to a linear polymer composed of heterogeneous monomers (e.g. nucleic acids) can regulate a complex network of catalytic autogenic reactions. The monomers comprising the template polymer would be additional diverse side products which each bind to one another so as to compose a heterogeneous linear polymer. In my 2012 book *Incomplete Nature* I envision an intermediate (unexplained) evolutionary step between simple and complex autogenesis in which variant nucleotide molecules are generated in the catalytic process and serve as energy transferring molecules via their pyrophosphate bonds (as they—e.g. ATP, GDP, etc.—currently do in all living cells). This is depicted in 5A. The energy captured (e.g. from activated phosphates produced near volcanic vents) and transferred by these molecules, could serve to accelerate catalytic reactions in an autogenic system, so that energy would not need to be extracted via the lysis of substrate molecules. This would enable utilization of a wider diversity of types of substrates and a more synthesis-based autogenic system. Nucleotide binding into a polymer (5B) could evolve as a means of temporarily inactivating this energetic function and storing nucleotides in the inert autogenic phase, ready to be depolymerized during the dynamic phase and again used to ferry energy to appropriate catalysts. But the random heterogeneous polymerization of different nucleotides will create variable 3-dimensional structure that will tend to differentially bind free catalysts. Figure 5C depicts the way that differential proximity of catalytic binding on such a linear polymer might bias catalytic interactions by releasing catalysts in a preferential order. And Figure 5D depicts a catalytic network of interactions distinguishing which reactions contribute to autogenic self-reconstitution and which compete. Favorable interactions could be biased by template-based sequential release of selective catalysts bound to distinct positions on the template molecule. Thus the constraint of binding proximity (on the template) would contribute to constraints affecting the probabilities of different catalytic interactions. Though such polymers would be initially structured at random, populations of autogens of this type with different linear polymeric structure, will compete for substrates. As a result a form of natural selection will occur that tends to differentially preserve those autogens with templates that minimize the probability of unproductive molecular interactions. Ultimately this will favor templates that replicate as well. Though this is not a template like the genetic code, which determines catalyst structure, it may be a critical simpler precursor providing sequence specific selection on polymeric structure that could later be recruited for more complex semiotic functions.

INSERT FIGURE 5 HERE

In these two more complex forms, the autogenic process is provided with information referring in different ways. In the adaptive autogen different levels of substrates in the immediate environment induce the autogen to respond in correspondingly different ways with respect to their differential beneficial or detrimental effects. Thus these different dynamical interpretant states effectively *indicate* to the system whether it exists in a beneficial or non-beneficial context. In the template-assisted autogen information preserving the complex system of constraints that preferentially maintain autogenic capacity has been offloaded from molecule-intrinsic catalyst-to-catalyst interaction constraints onto structural characteristics of the template molecule. The physical structure of the template in this way serves as a sign vehicle representing the preferential network of chemical interaction constraints that constitutes

autogenesis and can thereby reproduce and preserve this entire complex of constraints. As a system of structural relations (among individual structure-catalyst correlations) that corresponds to another system of catalytic interaction relations, it parallels *symbolic* representational relationships. The point is not to imply a direct mapping between molecular and cognitive indexicality and symbolicity, but rather to suggest that a common hierarchical logic of semiotic modes of reference may apply at all levels. Using these three variants of a simple model system we can begin to reconstruct the necessary features of the different modes of reference that roughly correspond to iconic, indexical and symbolic reference, phenomenologically—although in their most minimalistic forms. Being able to precisely identify the thermodynamical-informational-dynamical nature of reference, significance, and interpretation on a molecular level, and show how this can both preserve the essential properties and distinctions of the major semiotic categories, provides a solid foundation for a science of biosemiotics that is neither of merely philosophical interest nor merely an epiphenomenal gloss for simple chemistry and physics.

Finally, it should be noted that the determination of the relative significance of the information obtained by these autogenic processes is assessed by the relative minimization of work per reconstitutive cycle. Thus the probability of persistence of an autogenic form is a function of thermodynamic efficiency in some sense. This is the basis for evolution. Biological evolution can in this respect be understood in terms of a process that tends to optimize the relationship between information and the work that it organizes to preserve itself.

In conclusion I believe that these simple model systems provide a first step toward re-legitimizing the concepts of reference and significance that have so far been excluded from the natural sciences. Demonstrating that an empirically realistic simple molecular system can exhibit interpretive properties is the critical first step toward a scientific biosemiotic theory. A better understanding how interpretive dynamics can emerge from simpler chemical and physical processes should also point to new ways to study biological, neurological, and even social processes.

References:

- Barbieri, Marcello (2014) From Biosemiotics to Code Biology. *Biological Theory* 9:239-249.
- Bateson, Gregory (1968) Quoted in Harries-Jones, P. (in press) *Upside-Down Gods and Gregory Bateson's World of Difference*. New York: Fordham University Press.
- Ben-Naim, Arieh (2008) *A Farewell To Entropy*. World Scientific Publishing Co.
- Crick, Francis (1956): On Protein Synthesis. *Symp. Soc. Exp. Biol.* XII, 139-163.
- Crick, Francis (1970). "Central dogma of molecular biology.". *Nature* **227** (5258): 561–563.
- Deacon, Terrence (2012) *Incomplete Nature: How Mind Emerged from Matter*. New York: W. W. Norton and Co.
- Harries-Jones, Peter (in press) *Upside-Down Gods and Gregory Bateson's World of Difference*. New York: Fordham University Press.
- Peirce, Charles Sanders (1931) *Collected Papers of Charles Sander Pierce*. C. Hartshorn and P. Weiss (eds.) Cambridge, MA: Harvard University Press.
- Saussure, Ferdinand de (1983) *Course in General Linguistics*. C. Bally and A. Sechehaye (eds.). Translated by Roy Harris. La Salle, IL: Open Court.
- Shannon, Claude (July and October 1948) The Mathematical Theory of Communication. *Bell System Technical Journal* 27: 379–423, 623-656.
- Shannon, Claude and Warren Weaver, 1949. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.
- shCherbak, Vladimir and Maxim Makukov (2013) The “Wow! signal” of the terrestrial genetic code. *Icarus* 224: 228–242.
- Ulanowicz, Robert (2011) Towards Quantifying a Wider Reality: Shannon Exonerata. *Information* 2, 624-634.

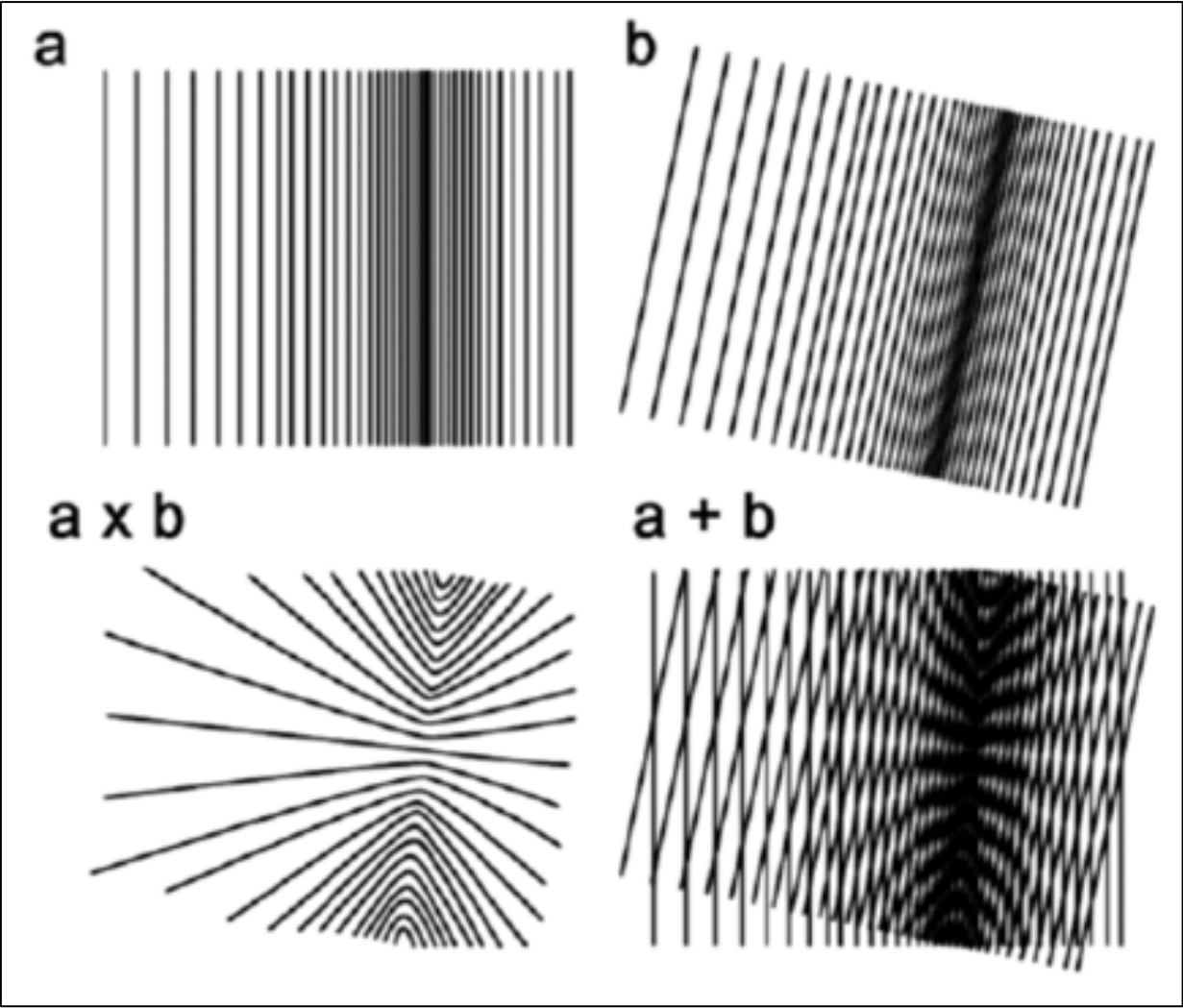


Figure 1. Moiré pattern due to the shifted juxtaposition of two regular patterns. Any two of the patters (including the interaction pattern) will generate the third.

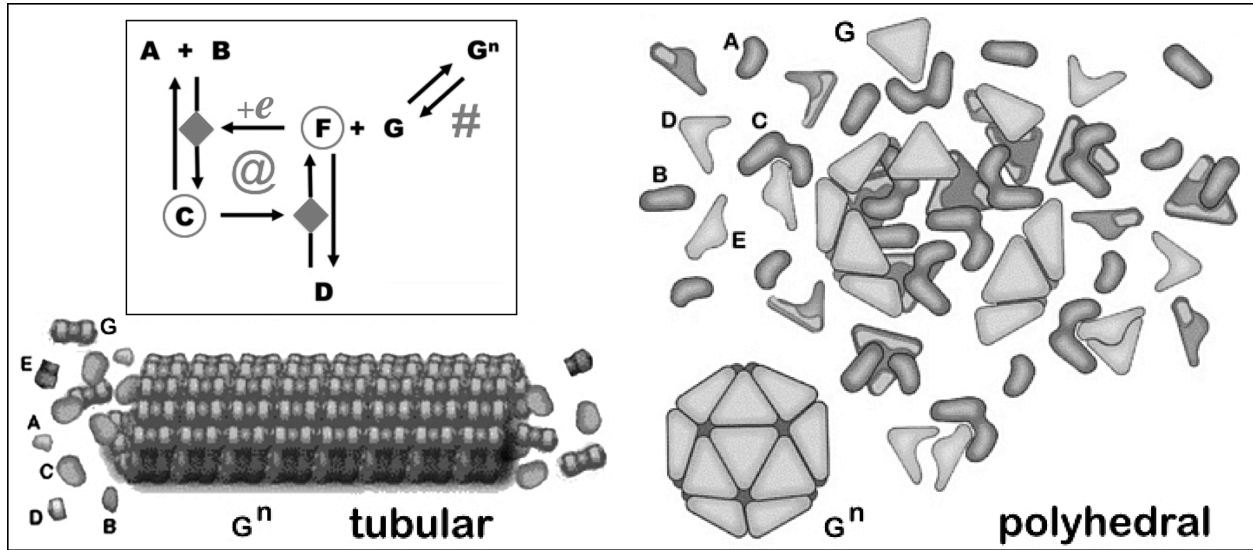


Figure 2. Two possible variants of a simple autogenic system are illustrated: a tubular form (lower left) and a polyhedral form (right). The general chemical reaction network for simple autogenesis is shown in the box in the upper left. Symbols in the reaction network diagram are as follows: Letters = molecules; O = catalysts, @ = reciprocal catalytic cycle, # = capsid shell self-assembly, +*e* = energy liberated from broken covalent bond, arrow = direction of reaction, diamond = catalyzed reaction.

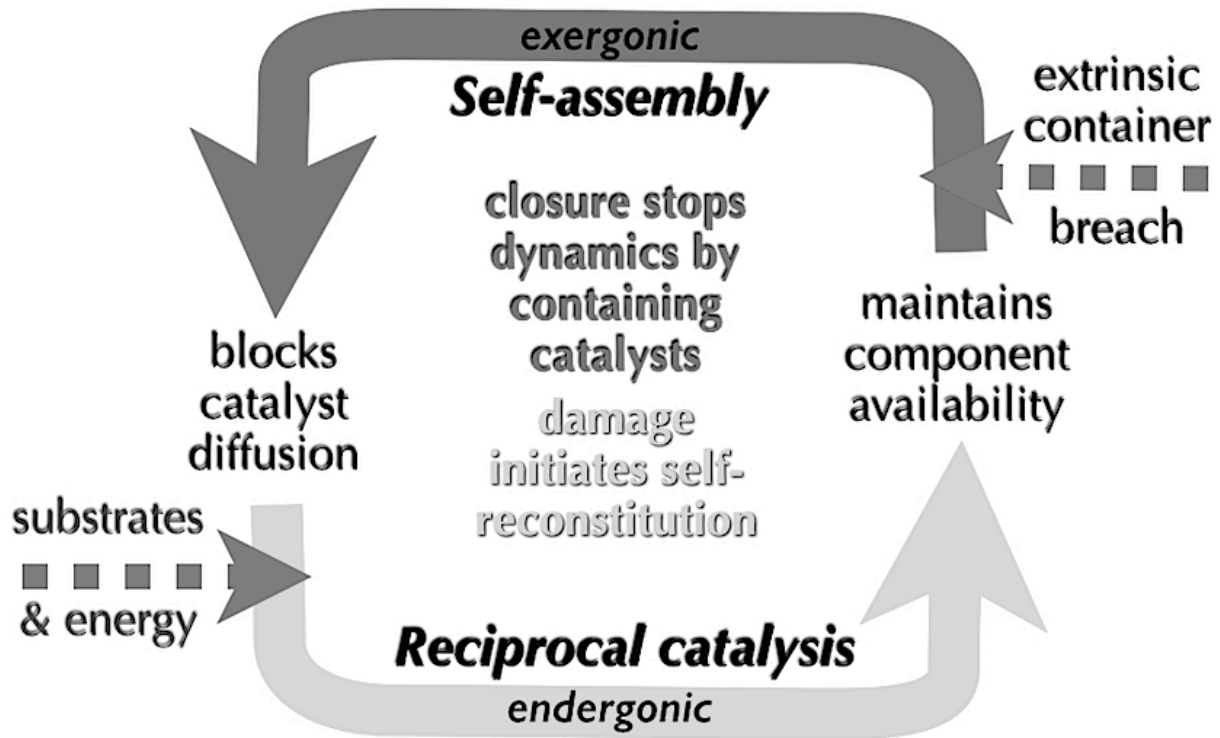


Figure 3. Diagram of an autogenic work cycle showing the synergy between the constraints generated by each component self-organizing process and the coupling of an endergonic (energy using) and exergonic (entropy increasing) processes.

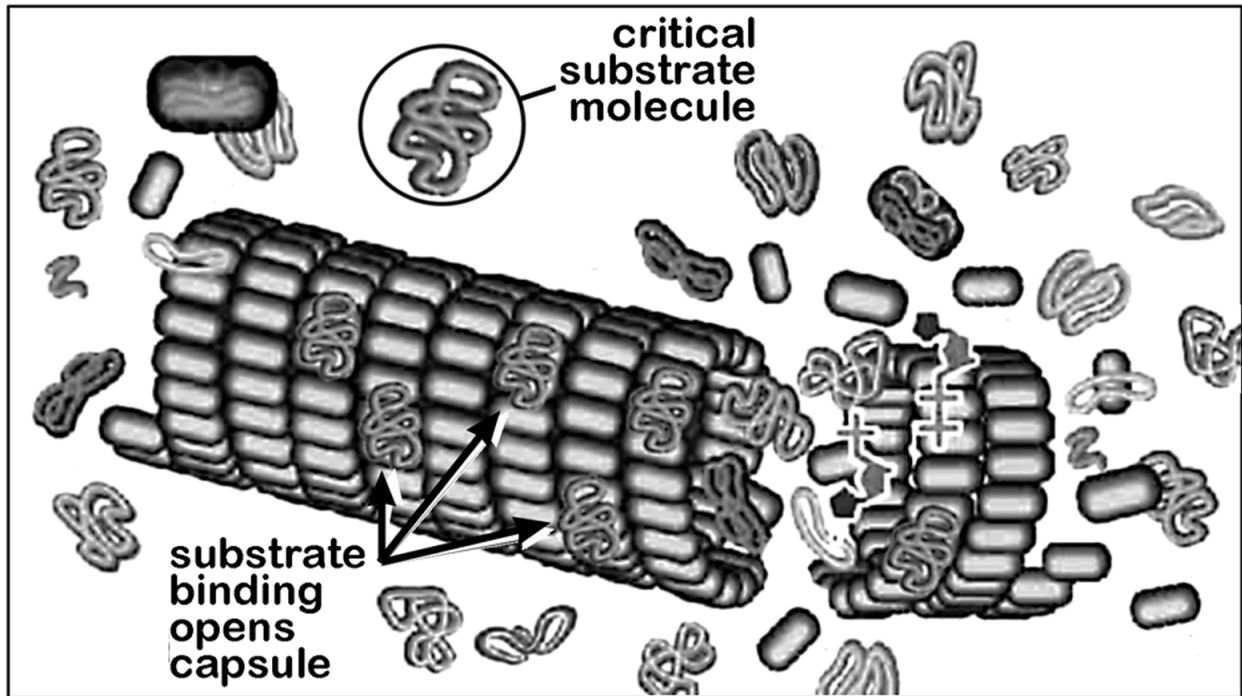


Figure 4. The logic of an adaptive autogenic system. An autogen is shown breaking open in response to critical substrates binding on its surface, and thereby weakening it. Once the capsid shell disassembles substrates will be released and made available for catalytic interactions along with other substrates in high concentration in the immediate environment. Thus, the extent of surface binding becomes an indication of the environmental potential for successful autogenic reproduction. See text for additional explanation.

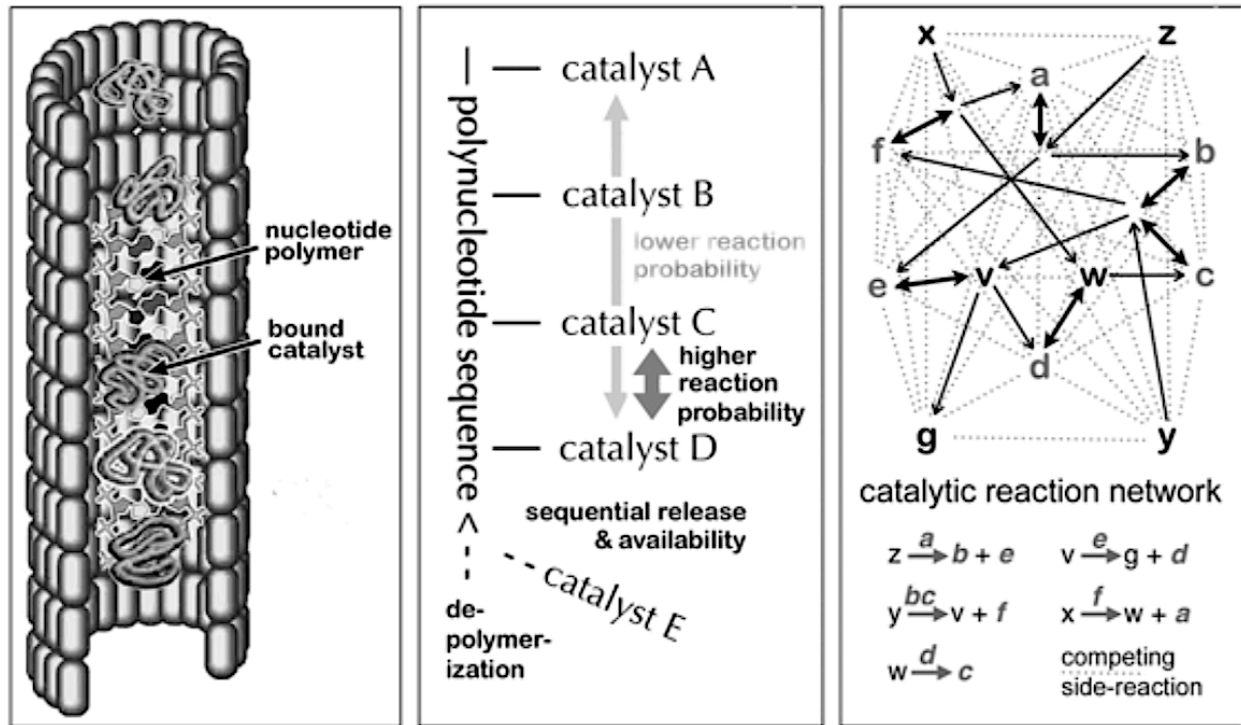


Figure 5. The logic of a template based autogenic system. A. An autogen that produces nucleotides to capture and deliver energy in the form of phosphates to aid the energetics of catalysis. B. Catalysts bound to specific loci along a polynucleotide template molecule within a tubular autogen. The polymerization of free nucleotides allows them to be “stored” during the inert phase for later use, and inactivating energetic phosphates that could otherwise provide potential damage. C. Sequential release of weakly bound catalysts from the template molecule showing how the spontaneous order of release can influence the probability of specific catalyst interaction. D. Above: Comparison of the full network of possible catalyst interactions (dotted lines with those that specifically contribute to autogenic reconstitution (solid arrows). Below: “Equations” for five catalytic reactions that contribute to autogenic reconstitution. Letters over each reaction arrow indicate the molecules that catalyze the reaction. See text for additional explanation.