

The pattern which connects pleroma to creatura: the autocell bridge from physics to life

Terrence Deacon

*Department of Anthropology and Helen Wills Neuroscience Institute,
University of California, Berkeley, CA 94720*

[deacon@berkeley.edu]

and

Jeremy Sherman

*Expression College of the Digital Arts
1830 Sonoma Ave.*

Berkeley, CA 94707

[js@mind-readers.com]

Keywords: Gregory Bateson, origins of life, protocell, autocell, autocatalysis, self-organization, evolvability, teleology, emergence

Running title: From pleroma to creatura

Abstract: By his own standards Gregory Bateson was unsuccessful in his life-long quest to explain how the informational or living realm (creatura) could emerge out of the energetic or physical realm (pleroma). Drawing upon recent insights in self-organization theory, the authors suggest a missing link connecting the realms; a simple spontaneously-arising, non-living, yet evolvable molecular system called an “autocell” consisting of the reciprocal linkage between an autocatalytic cycle and a self-assembling encapsulation process (modeled on viral encapsulation) where the molecular constituents for the capsule are products of the autocatalysis. Autocells are shown to have the rudiments of individuality, end-directedness, function, and valuation; thus bridging the critical initial gap between pleroma and creatura.

Introduction

In a previous paper in this volume, Ty Cashman calls our attention to Gregory Bateson's abiding, yet ultimately unanswered question: What connects creatura to pleroma, the realm of life to the realm of non-life, the informational realm to the energetic realm, mind's realm to matter's realm? This is not, however, a division that the majority of scientists and philosophers today recognize as particularly troublesome. Since Descartes, Western traditions have focused on the split between soul and body, consciousness and biology. But to Bateson, evolution and learning were contiguous and the distinguishing features of mind were evident in all living processes. For him, the critical “epistemic cut” was identified between energetic and informational processes, and since life has at its base an informational character the crucial distinction must be drawn at the life / non-life transition. Though he never directly attempted to address questions about this transition, understanding the pattern of causality that connects non-life to life, physics and chemistry to biology, is implicitly the core mystery that stands behind his larger enterprise. While many, if not most of our colleagues remain more interested in the emergence of consciousness, and treat the emergence of biology from chemistry as a nearly solved problem, we believe that Bateson's insight points to an unrecognized fallacious presumption in this viewpoint: that life is reducible to mere chemistry and physics and information to patterns of energy exchange. In this essay we show that a careful and complete modeling of the transition from pleroma to creatura can clarify what these reductionist assumptions miss and provide an unambiguous connection linking the energetic to the semiotic realm.

Among Bateson's many methodological innovations, one of the most important is a focus on relationships rather than things, distinguishing the material-energetic realm of physics from the informational realm of life and mind. Information is a relationship between things in which one thing is shaped over time by another thing. Any "piece" of information is an irreducible relationship between a sign vehicle, a sign source, and a sign interpreting process. "The difference that makes a difference," makes that difference to something about something. A piece of information is therefore three things with relationships in their interstices. We, as signal recipients, detect a difference in the sign vehicle. The difference we detect makes a difference to us—it changes us. It makes a difference to us with respect to something that we take the sign vehicle to be about—a sign sender, the difference we make is with regard to the sender.

This is intuitive enough. It becomes counterintuitive when we recognize that so long as we think of sign senders and sign recipients as like us, we are smuggling black boxes into our schema—homunculi who send and interpret messages. To avoid this preformationist assumption, we must recognize, following the philosopher Charles Sanders Peirce, that what interprets a sign is not merely a locus—some being outside the which itself constitutes a sign vehicle interpreted by yet another sign recipient, and so on. Therefore an informational relationship is not a relationship in isolation. It is part of an information system, a vast network of interrelated relationships.

Many of Bateson's successors interpreted his last gropings as tipping toward a commitment to subjective idealism, an assumption that minds are independent homunculi, connected to the pleroma only by an article of faith. Others, Cashman and

ourselves included, believe that Bateson's "pattern which connects" offers a way to bridge this gap deemed by the idealists to be unbridgeable.

We argue that implicit in Bateson's distinction between *creatura* and *pleroma* is a necessary assumption that *creatura* emerged somehow from *pleroma*. To understand the pattern that connects the two realms, then, we must ultimately answer a very specific question: How could *creatura* get started in an exclusively *pleroma* universe? Or, to translate into the more familiar biological framework, how could life have emerged here or anyplace in a universe that doesn't already possess life's attributes of relationship, function, and evolvability—a universe governed solely by the known forces of physics and chemistry? Until this question is answered we cannot hope to fully grasp Bateson's "pattern which connects."

The problem with protocells

Conventional inquiry into this challenge often employs a reverse engineering approach. Researchers identify the minimal characteristics of life (nucleic acid replication, cell membranes, and so on), and the elements of living systems that embody those characteristics, (DNA, RNA, lipids, and so on). They then imagine scenarios in which by chance alone, these essential constituents of cells might come into proximity and "come alive." The current experimental work on the origins of life that is considered most promising, involves laboratory efforts to actively construct minimal cell prototypes, as if to provide what engineers would call "proof of concept."

If the goal were to engineer life, then building a successful prototype would be half the battle. But engineering is precisely what a pre-life universe lacks. Origin of life

researchers attempting to reconstruct new life from the scraps of living cells are not filling in the un-engineered missing link between physics and biology, but simply trying to put Humpty-Dumpty together again, with a few pieces left out. This may show that life is not some ineffable essence, but does not show how life's properties can emerge spontaneously from a lifeless world. So in this effort, while recognizing the circularity of arguments from intelligent design, origin of life researchers unwittingly employ reasoning similar to that of intelligent design's defenders. They act as foresightful watchmakers in their effort to prove that life is the product of a blind watchmaker.

Indeed, the reverse engineering approach to the origins of life parallels watch making. Watchmakers assemble parts that already have been shaped for a specific function. They kick start the resulting product and it becomes animated. Protocell research assembles components of life that have evolved with particular functions. and imagine how nature might bring them together, and then assume that once together, they would come alive.

In two respects, the accretion of life from non-life is more challenging than the reverse engineering approach would suggest. First, the elements to be brought together could not have been previously modified to work together, as are the molecular complexes of cells. They can only be those that can be spontaneously produced in a lifeless context irrespective of one another or any living function. Second, although protocells made of parts salvaged from once-living cells may come to function in an integrated fashion, any such synergy traces to their long evolution not to any spontaneously emergent complementarity. A bricolage constellation of once-living

components thus provides little insight into the transition from non-life to life. Protocells are in this way more like reanimated frankensteinian cells.

The true proof of concept requires demonstrating how critical properties of life could emerge by the spontaneous interactions among inorganically generated molecules. These properties include forming a complex that maintains its individuality despite perturbation and the ability to replicate itself indefinitely if conditions are right. This approach is a far more difficult alternative than reverse engineering from a known endpoint, and yet it inevitably involves fewer components in simpler combinations owing to the requirement that it must have arisen spontaneously. To meet this challenge requires vigilance to avoid what could be called the *Amnesiac Watchmaker Syndrome*—reverse-engineering from life, and then assuming that the engineering effort was irrelevant, as though we found this to be spontaneously produced.

In this paper we offer an exemplar proof-of-concept model that shows how properties of creatura could spontaneously emerge from mere pleroma. We propose a generic model of the transition to life that could apply anywhere in the universe; one that gets a cold start in relentless thermodynamics and molecular coincidence.

By cumulative standards, every step in the transition from non-life to life must be accounted for. Working forward to life instead of backward from life, it becomes obvious that anything approaching the complexity of even the simplest contemporary terrestrial life-forms could not possibly emerge de novo from inorganic chemistry and physics in a single transition. Our emphasis will therefore be on exploring the realm between non-life and life; forms that are not living but nonetheless possess properties that would bias

chance increasingly toward the self-maintenance and reproductive capacity that would be the necessary precursor attributes of anything evolvable.

The challenge for a non-engineering cumulative approach like ours is demonstrating not only how precursor elements might fall together, or even stay together but more importantly how they might become necessary to each other. In the abstract, we might describe this alternative to reverse engineering as searching for the possibility of spontaneous combinatorial synergy, but this is not quite sufficient. The concept of synergy is ambiguous. A weak interpretation of synergy is as a state in which each part is of some benefit to another. A strong interpretation of synergy is as a state in which each part is wholly dependent upon the others. The alternative must be held to the more stringent standard of strong synergy; with components that are both co-dependent and also co-productive.

By this standard, of all of life's features, the feature that is most central, and that must arise first is the reciprocally productive coupling of component products. As the 18th century philosopher Immanuel Kant argued, "The definition of an organic body is that it is a body, every part of which is there for the sake of the other (reciprocally as end and at the same time means)...An organic body is one in which each part, with its moving force, necessarily relates to the whole (to each part in its composition.)" (Kant, 1790)

In sum then, our goal in what follows is to see pure happenstance produce minimal evolvable products, a population of forms that would be differentially sustained over time depending upon environmental conditions.

The spontaneous co-productive approach to the origins of life also may provide important insights concerning the nature of other emergent transitions as well. Often emergence is treated as a problem for the philosophy of science; one of defining what we mean by sharp inexplicable transitions between the explanatory regimes of physics, chemistry, biology, and consciousness. Emergence occurs when the net effect of the interactions among parts of a system produces a significant shift in the characteristics and behavior of the whole system. If we can provide a complete step-by-step description of one such transition, we should be in a better position to make sense of others—especially as this is a paradigm case. So again, we may be able to improve upon mere descriptive or reverse-engineering analyses of emergent transitions, such as that leading to mind, by comparison to this constructive model. Beyond merely exemplifying emergence, then, we would hope that traversing this one transition, will pave the way to a more subtle and complete understanding of emergent processes in general.

Finally, one further benefit of providing a step-by-step account of the transition from non-life to life is that it can illuminate the concept of purpose or *telos*, and its origin from a previously purposeless universe. In the debate between science and fundamentalist religion that the world's citizens watch so attentively, the burden is on science to demonstrate how purposive processes can emerge in the absence of antecedent intelligence, carefully selected prior conditions, or intrinsically teleological components. Until it is possible to show at least in principle how teleological dynamics can emerge *de novo*, a shadow of doubt will remain about whether mechanistic causality and teleological causality are compatible, and whether consciousness, meaning, and value are ultimately ineffable. Bridging this threshold is the first step toward naturalizing

humankind's undeniably sighted watchmakers. Though we will not attempt to address any of the grand mysteries of subjective experience and value, we believe that by tracking the emergence of the fledgling precursor to these intentional relations—function—we will have provided firm evidence that the naturalization of purpose is not an impossibility.

The Thermodynamic Universe

A collection of interacting molecules (for example, a container of water) has no integration of parts, no for-the-whole contribution of one molecule to the rest, but, still, global properties emerge as these interactions produce a kind of vast regression toward the mean. Global features such as surface tension emerge in this way, and they are to some extent independent of many of the details of the molecules involved, since many substances can exhibit this phenomenon. Like the averaging of molecular movements producing temperature, surface tension is not a property of individual molecules or even of a few in interaction; it is a system-wide effect of a very large averaging process. Surface tension is a product of what are called van der Waals's forces: a weak energetic attraction between like molecules that keeps the space between them at a minimum (a little bit like the stickiness that can hold a statically charged balloon to the wall). The momentum of individual molecules is altered by the presence of nearby molecules in close interaction with each other, causing greater coherence between them than between them and molecules of air which are less tightly coupled.

Emergence in simple thermodynamic systems flows from the logic of the second law of thermodynamics, which, although based on the billiard ball-like interactions of

molecules, is itself quite different in character because it is intrinsically distributional and statistical. It is a rule about probable and improbable patterns and how they are likely to follow each other in time. Although this rule is not so specific or absolute as are Newtonian laws of motion, it is all but guaranteed, in the same way that shuffling a deck of cards hundreds of thousands of times is still hugely unlikely to ever result in an arrangement in which each of the suits is separated and numerically ordered. The organizing effect exemplified by system properties such as surface tension occurs because the statistics of the ramifying and recirculating exchanges of energy eventually pits every difference of movement against every other, eventually leading toward thorough mutual cancellation, and thus uniformity. So, before life and before purpose, we can apply the concept of emergence in a minimal sense to these highly robust system-level tendencies. It is important to note, however, that these are the product not of new forces, but, rather, of population effects of the same forces. And, yet, the system-level tendencies have a new kind of causal efficacy that while based on the system's individual elements is primarily the result of the distributional characteristics of their repeated interactions. This why both water and alcohol can behave in ways that are almost indistinguishable.

Thermodynamics plus Shape: Biased Molecular Interactions

In simple physical systems, molecules interact thermodynamically—colliding, rebounding, sometimes sticking. But it is not just their direction and momentum that matters. Shape matters. All of what makes chemistry different than billiard-ball interaction is a consequence of shape effects at some level and how this biases what is likely and unlikely to occur over and above the relentless evening out of the energetic

shuffling process. At collision velocities and angles inappropriate to create stronger, electron exchanging covalent and ionic bonds, molecules usually just rebound. But because molecules in a solution also exhibit van der Waals's stickiness, in some orientations this mutual stickiness aided by complementary shapes can overcome the momentum that would otherwise cause them to bounce apart. The strength of this attraction, called hydrogen bonding, is comparatively weak and does not often reach the level to keep molecules together for long amid the constant jostling with neighbors. The strength of this stickiness, and thus the duration of time molecules tend to stay attached, is a function of the quantity of conforming surface area between them. (Thus, very large molecules with lots of surface area in contact can actually be very tightly bonded this way, as in the case of DNA.) The closer the fit between the shapes in contact between two molecules, the more likely they will stick, and the longer they will stick in exactly that orientation.

On the one hand, this differential stickiness is purposeless, devoid of all teleological impetus. It is just a consequence of the chance distributions of shapes and bonding predispositions in a collection of molecules. Overall, the stickiness is random, and mutually canceling, moving molecular interactions toward equilibrium. On the other hand, the contribution of shape that makes some molecules more likely to stick to each other, and more likely to be oriented in certain ways when they do, makes the whole process nonrandom in other ways.

Thermodynamic emergence drives systems toward equilibrium, in all attributes. Still, there is a very low but non-negligible potential for modest local reactions that bias molecular types away from an all-things-being-equal equilibrium. The effects of

molecular shape are some of the most important of these biases. One major mechanism underlying this potential is catalysis, a result of shape-mediated molecular stickiness that affects rates of chemical reaction (that is, formation of ionic and covalent bonds, changing the structures of molecules). Catalysis occurs when one or more molecules mediate and potentiate specific reactions of other molecules but in which the biasing molecules are not permanently restructured in the process.

Most organic catalysts are proteins that increase the possibility that certain chemical reactions will occur by biasing the proximity and orientations of select molecules in ways favorable to a given chemical reaction between them. This is the result of the way the shape of the catalytic molecule temporarily captures other molecules with complementary shapes to it, briefly holding them in a specific orientation. This action may favor the captured molecule cleaving at a specific weakened point, or bringing two molecules into proximity with each other in orientations favorable to their forming a bond and fusing. The unaffected catalysts molecules retain their shape and therefore retain the ability to cause still other molecules to break apart or come together.

Catalysts have no purpose. Their biasing affect is merely a result of chance shape correspondences. But shape biases significantly contribute to skewing the thermodynamics of a molecular system away from billiard-ball randomness. A random distribution of molecules that happens to have strong catalysts present would spontaneously exhibit a system-wide behavior, a trend toward increased transformation of certain molecules to certain other molecular forms. This behavior is the net effect of many interactions between the elements. It is dependent on the energetic stickiness between molecules, and, hence, it is also a result of thermodynamic emergence—but not

just thermodynamics. This is one reason chemistry must take more into account than physics. The relationships between the shapes of the molecules, their relative stickiness, and their electron sharing/exchanging potentials combines with thermodynamics to determine the resulting behavior. Chemical solutions, though still subject to shuffling effects, can produce a biased sorting and transformation of molecular types. So chemistry combines an evening-out effect with shape-charge-bias effects.

From Thermodynamics to Morphodynamics

There are, however, special cases where the form effects don't tend toward an equilibrium, but, instead, compound to become even more prominent than the mutually canceling effect of thermodynamic shuffling. This happens in the special case when there is continuous thermodynamic instability, so that, within a part of a system, things never get a chance to even out. These are non-equilibrium conditions, and in these cases the biasing effects of shape can come to predominate. With new energy or materials continually entering and replenishing an original uneven distribution, interactions within this non-equilibrium context tend to generate compensating processes. Energy and new materials must be exported as quickly as they arrive, or else imbalances will build to the breaking point. Under these circumstances a different kind of evening-out process must occur: a balancing between input and output. Under these circumstances, mutually counteracting biases will still cancel, but complementary biasing effects will accumulate consequences and reinforce one another with respect to their combined contribution to regularizing this flow.

In non-equilibrium systems, as in the case of emergent features of thermodynamic

processes discussed above, regularities emerge at higher-scales because of mutually canceling interactions of components. It's just that some effects do not cancel; they amplify. The biasing effects of shapes interacting with shapes can reinforce one another, if they are complementary. This is most easily seen in cases where diverse shapes and sizes of solid objects, e.g. pebbles, are incessantly buffeted, e.g. by ocean waves hitting the shore, and end up distributing themselves on the shore according to these features. The constant through-put of disturbing materials or energy samples and re-samples these otherwise subtle biasing effects, and compounds these with each other while it damps the rest. This pattern of dynamics, which is a special case of thermodynamic tendencies amplifying specific form tendencies, can be called morphodynamics.

Autocatalysis: A Morphodynamic Chain Reaction

In a solution of many diverse kinds of molecules capable of catalytic effects, chances are fair that two or more catalysts will be mutually reinforcing in their biases. For example, one catalyst might transform raw ingredients into a molecule that can itself act as a catalyst. To risk a purpose-laden metaphor, this would be the equivalent of a production line that produces production line equipment. Though rare, it is also possible for a catalytic circle to form. To take an oversimplified example, imagine a catalyst, A, that catalyzes certain molecules, bonding them together in such a way that they become a second catalyst, B. Imagine, then, that catalyst B catalyzes certain other molecules, bonding them together in such a way that they become catalyst A. In this condition, known as autocatalysis (Prigogine, 1984), the reciprocally reinforcing relationship between catalysts produces a kind of runaway effect. Starting with one of two catalytic molecules and abundant raw materials for both catalysts, the amounts of A and B would

double with each catalytic cycle. This produces a compounding effect, because each cycle produces more catalysts, producing more cycles. Autocatalysis causes an accelerating chain reaction for as long as molecules are available to be catalyzed. Thus autocatalysis is a morphodynamic process determined by the way shape relationships self-amplify. But under normal circumstances substrates will be depleted very rapidly by runaway autocatalysis. So although autocatalytic sets diverge from initial conditions to rapidly bias concentrations of their components in some part of a solution, this is typically a very short-lived trend, eventually re-equilibrated.

Though it might be imagined that coincidentally reciprocal molecular shapes are highly improbable, work by Stuart Kauffman demonstrates that autocatalysis is more likely to arise spontaneously than one would generally suppose (Kauffman, 1993). In an arbitrarily chosen solution of diversely shaped molecules in which there is a modest probability of some acting as catalysts for the formation of others, above a certain level of diversity of molecules the chances for closed circles of catalytic reactions grows rapidly, and eventually is a near certainty.

Autocatalytic sets are coherent sets in theory only, however. Unlike the analogy to a metaphoric factory production line, here, there is no dedicated linkage between the producers. Just as any single catalyst drifts about, interacting with molecules by chance, so, too, each of the catalysts in a catalytic set drifts without affinity, except by chance encounter. Although autocatalytic chain reactions have a kind of causal efficacy as a set, shifting a chemical soup out of equilibrium, they are merely molecules in local stochastic interaction. And despite the fact that each molecule of an autocatalytic set is, to our teleological way of thinking, part of a whole and able to produce more of its own

components, this higher-order relation has no independent status. Individual catalysts encounter each other or diffuse randomly. Their interdependence is fleeting and serendipitous. The “set” itself is merely a descriptive attribute we use to explain the special dynamics that results, but which provides no independent constitutive contribution. Thus the set has neither individuality nor efficacy with respect to itself, and although its components can increase each-others’ concentrations the “set” is not reproduced and cannot be in competition with other sets.

Molecular Self-Assembly

Molecular self-assembly constitutes another class of molecular interactions that, like catalysts in an autocatalytic set, produce a runaway production of similar forms because of shape-fitting interactions. Self-assembly is a self-reinforcing pattern of molecular binding. In the same way that a catalyst and substrate molecule bind in a specific orientation, a single type of molecule can bind with other like molecules, forming an ordered structure, somewhat like a crystal. These structures could take various forms, depending on the symmetry of the molecule. Many molecules that bind into complexes produce clumplike or crystal-like structures, but some form regular polyhedral shells or tubes. The protein shells that encapsulate many viruses are well-known examples of self-assembled molecular polyhedrons and the microtubules that provide the virtual skeleton of eukaryotic cells are spirally assembled tubes. Self-assembly occurs spontaneously because molecules shaped so that they form regular arrays, when fit together, are at a lower energy state than when freely floating. And, as such structures grow, the number of facets in which new molecules can fit increases as well. If enough regular-fitting polygonal molecules are in the vicinity, self-assembly will result in

formation of a hollow container. As such structures form, they will inevitably enclose other molecules in the vicinity, as though in a molecular capsule. This can lead to an interesting phenomenon.

It is possible that one of the catalysts or by-product of an autocatalytic cycle could also be a self-assembling molecule. In this special circumstance, self-assembling molecules would spontaneously form shells in proximity to molecules of the autocatalytic set of which they were by-products. In so doing, the shell is likely to enclose some or all of the molecules comprising the autocatalytic set. We would then have autocatalytic elements drifting within a shell that keeps them in proximity.

Deacon (in press) calls such self-enclosed autocatalytic sets ‘autocells.’ We should be careful to point out that they are not cells, like bacterial and eukaryotic cells, nor are they alive by any usual definition, yet they have interesting lifelike properties we will now discuss.

Autocell Functions

Surprisingly, autocells are sufficiently complex molecular systems to illustrate how elementary functional organization can emerge spontaneously from self-organizing processes. The key feature is not any single type of process in autocells, but rather the synergistic relationship between the two self-organizing processes that reciprocally depend upon one another’s persistence. This feature does not depend on the continuous persistence of self-organizing processes, or of any component chemical reaction, only the *potential* of their persistence. Thus, paradoxically, one of the critical characteristics of autocells is that they are self-stopping. When a shell is complete, enclosing autocatalytic

molecules, it limits catalysts' access to substrate molecules. Enclosure causes its catalytic processes to run down more rapidly than it would if the catalysts were free floating, so that catalysis inside the shell ceases altogether shortly after closure. Nonetheless, enclosure also keeps catalytic molecules from diffusing; they maintain proximity to one another despite their chemical inactivity. So, self-assembly temporarily limits catalytic activity and halts further growth but also limits molecular dissipation that could *permanently* undermine autocatalytic capacity.

Enclosure is inevitably a temporary condition. Molecular shells are buffeted about, as are individual molecules. As a result, they will occasionally break, spilling their contents, allowing their previously sequestered catalysts to again come in contact with the external milieu. If an autocell shell contains a full complement of catalysts from the autocatalytic set, and breaks in the presence of catalyzable molecules, the autocatalytic cycle can begin again, producing more catalysts and more shell molecules and again reconstituting a new shell that closes around whatever molecules happen to be present. Moreover, new autocells formed from the breakup of a parent autocell will form in the same way as their parent, and so will maintain continuity of structural characteristics over an extended period of time. In this way, the addition of the shell produced as a by-product of the autocatalysis creates the minimal condition for sustainable autocatalysis. By alternating between an enclosed dormant form and an open catalytic form, the overall configuration will be capable of both self-repair and self-replication, even though components come and go. An autocell is effectively a two-stroke engine that alternates between two reciprocal states—active and passive—to achieve a best-of-both-worlds configuration. An autocell is not responsive to its environment—responsiveness would

have to evolve. Still the two-state nature of an autocell represents a necessary precursor condition to responsiveness, a life-like balancing act that natural selection could eventually shape. Thus, concomitant with co-dependency emerges another of life's primary requirements, an adaptable changeability that supports its continuity with consequences for the longevity of its form.

It is this systemic interdependence, or synergy, and not any component molecules or chemical reactions, that is the defining property of an autocell. In this way, autocellularity is not decomposable to any of its component molecules or reactions, even though they are necessary components. This special complementarity that exists between the two kinds of self-organizing processes in an autocell validates calling these processes *functions*; they are appropriately described as such precisely because of the way in which they indirectly aid their own persistence irrespective of specific material components. For example, we can now say that by virtue of self-assembly, an autocatalytic set functions to generate its own protection against diffusion; the shell protects the process that produces it. Similarly, we can say that container self-assembly functions to insure future container repair and replication by capturing the catalysts it needs. Autocatalytic sets by themselves do not endure, because they deplete resources and ultimately dissipate. For this reason, we would not use the term *function* to describe the reciprocal relationships between catalysts in the set. In contrast, within an autocell, these same relationships can be described as functional, precisely because they don't merely increase the probability of similar reactions (as do free-drifting catalysts of an autocatalytic set); they also play a role in ensuring that the *potential* to do so is more likely to persist as well. The functionality of component processes is defined with respect to this reflexivity whereby

each contributes to the contextual conditions that make them more likely to persist, even despite their periodic cessation. Autocellular synergy thus illuminates a necessary condition of *telos* that we describe as ‘function.’

There is nothing about an autocell that is greater than the sum of its parts, and, yet, there is nonetheless a new kind of causal efficacy exhibited that is irreducible to the precursor causal processes, treated in isolation. More precisely, although autocells are analyzable into component molecules and chemical reactions, autocellularity—and the special properties of self-reconstitution and self-replication it creates—cannot be reduced to these components alone. Although reaction energy, molecular interactions, and the thermodynamic and morphodynamic processes that result are necessary conditions for the production of autocells, once formed, autocells take on “a life of their own” as causal loci of a special type. Only where these conditions converge to produce the uniquely synergistic topology of causes that defines an autocell are these higher-order properties exhibited. This feature gives autocells a form of objective individuality that is quite concrete and distinctive, despite the fact that it is not identifiable with respect to any specific molecules or chemical reactions. It is individuality of a *potential* for a specific chemical dynamic, a particular structure, and most of all of specific systemic properties.

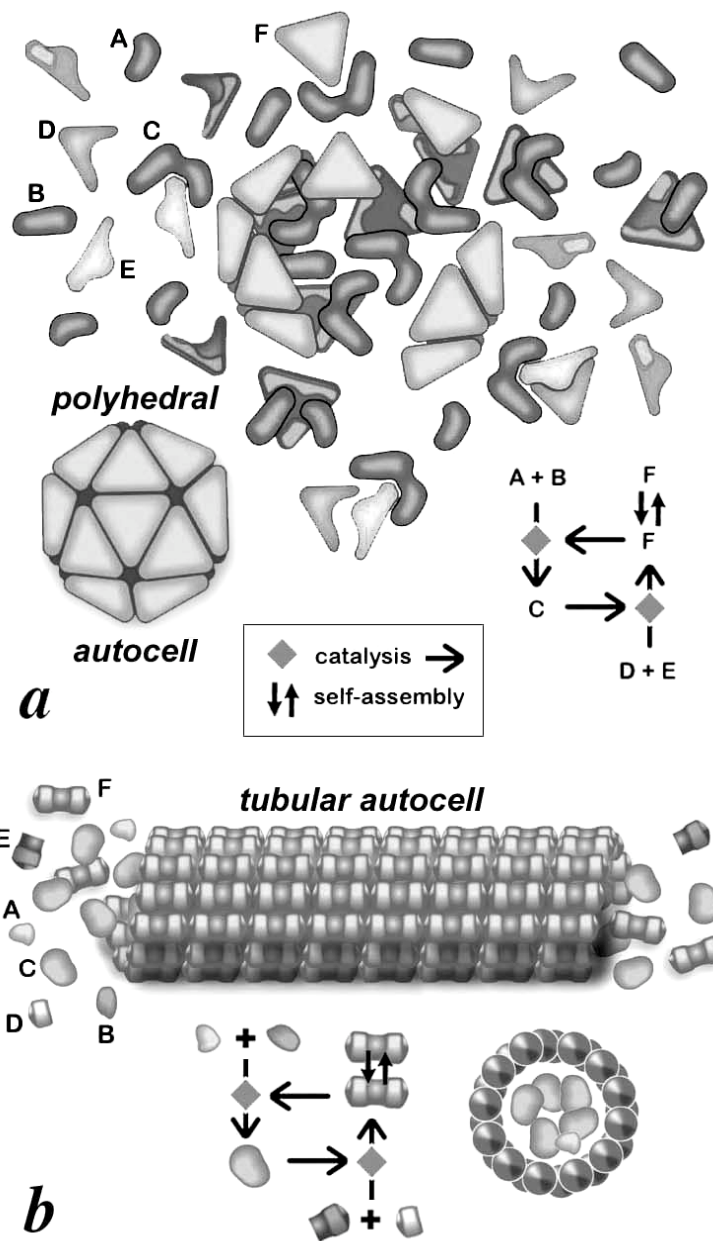


Figure 1. Two general classes of autocells are depicted as geometric constructions. An autocell produced by polyhedral containment is depicted in *a* and an autocell produced by spirally elongated tubular containment is depicted in *b*. Both are minimal autocells to the extent that each is constituted by only two catalysts (C and F in both). Catalysts are also depicted as synthesized from two substrate molecules in each case (A & B and D & E in both), though only in *a* is there any indication of the shape complementarities contributing to catalysis and self-assembly. Autocatalytic cycles are depicted with arrow diagrams for each (using letters in *a* and component shapes in *b*). From Deacon (2006 in press)

The Birth of Evolvable Teleodynamics

We have seen how an autocell makes possible a prototypical self-reconstituting configuration of molecules worth regarding as an individual. Though the autocell lacks many necessary features for life, it is an individual precisely because this configuration constitutes a locus of causal consequences that are irreducible to any precursors. Because of this individuality and self-directed, self-reconstituting, self-reproducing dynamic the locus of potential that defines an autocell exhibits a primitive form of end-directedness; in other words, a minimal *telos* or *entelechy*. For this reason, although the process is both a reflection of thermodynamic and morphodynamic processes, we can describe this higher-order dynamic as teleodynamic.

Autocells have one more critical capacity that makes them of central interest to our discussion of the origins of *telos*: they can evolve. Shell molecules breaking apart and coming together would form multiple shells as easily as forming one alone. Multiple shells forming in and around an autocatalytic set would capture slightly varied samples of autocatalytic molecules. Many shells would be empty, many would contain incomplete complements of the necessary set for reinitiating autocatalysis when opened, and many would open in solutions devoid of catalyzable molecules. In these circumstances, dissociated components would likely dissipate without reforming shells. But where a full set of catalysts is enclosed and then broken open in the context of catalyzable substrates there is the potential for the vague equivalent of multiple “offspring” being generated. Breaking open by happenstance in the presence of catalyzable molecules, the contents of

one autocell could reconstitute the original, or form several autocells.

Because of their periodic openness to the surroundings, given enough time and enough persistence, the autocatalytic set characterizing an autocell would likely come to interact with elements not in the original set. Some such interactions would undermine catalytic productivity. For example, a new catalyst entering the mixture might begin to transform other catalytic contributors, breaking them up even as their precursors produce them. In such cases, autocatalysis would significantly slow and decrease the effectiveness of the process of re-enclosure. Autocells thus handicapped would replicate more slowly and would fail to reconstitute more often than others.

It is also possible, however, that new catalytic relationships could accumulate without undermining the original autocatalytic set's productivity. Alternative catalysts or substrates might be present when an autocell opened that could be even more effective than the original precursors, or that provided an additional complementary catalytic route to autocatalysis, producing one of the autocatalytic set's component elements by other means. Having two ways to produce a catalytic element might accelerate the autocatalytic set's chain reaction. Various combinations of catalytic substitutions and partial duplications of cycles might also increase catalytic versatility, making the more complex autocatalytic set productive in environments that lack the raw ingredients for one or the other cycle. All such changes will result in alternative variant forms of an autocell lineage.

Biologists are coming to recognize how much of life's innovation comes about not just through random variation in genes but through homeosis, the chance duplication of components. Divergent modification of redundant components (typically, duplicated

genes) can increase both versatility and specialization over the course of subsequent evolution, by virtue of the way it predisposes the evolution of complementary function. Hemoglobin provides a classic example: Multiple duplications of the ancestral globin gene evolved divergent complementary functions. In the same way, despite their simplicity, the inclusion of two or more slight variant forms of component autocell processes, could serve as the basis for autocell complexification in evolution.

Any of these transformations would amount to the equivalent of speciation of autocell lineages. In effect, two species of autocell—the one with the original structure and the one with an alternative structure—would be in competition with each other for those resources they both utilize. There would be differential replication rates and differential stability, and, thus, differential lineage propagation, due to the difference between their alternative ways of carrying out the corresponding functions. Though it is a leap from autocell evolution to the evolvability in life, this simple model provides a plausible hypothesis for the minimal conditions for evolvability, without needing to postulate the existence of information-bearing molecules like DNA. Indeed, this gives us a way to understand how molecular encoding of information in specialized template molecules could be a later evolutionary acquisition—one that differentiates out of this more primitive functionality. There is no need to imagine that information-bearing molecules like RNA are a necessary primitive given for the origin of life.

Discussion

The origins of life have remained mysterious for so long that many assume it to be a recondite phenomenon. But we are aided in reconstructing this transition by virtue of

the fact that conditions in a pre-life world would have necessarily limited the origin's complexity. The autocell scenario neither relies upon engineering or radically improbable coincidence. Nor does it imagine that life with the features of contemporary forms could arise in a single step. Life originated, by necessity from extremely simple beginnings and the transition from non-living to living was thus likely a many-stage evolutionary process.

In our scenario, we started not with evolution (for how could evolution arise by evolution?) but rather self-organization or as we have termed it, morphodynamics, the production of short-lived dissipative structures.

Next in sequence was Kant's necessary condition of reciprocal co-production:

“In such a natural product as this, every part is thought as owing its presence to the agency of all the remaining parts, and also as existing for the sake of the others and of the whole, that is as an instrument, or organ...Its parts must in their collective unity reciprocally produce one another alike as to form and combination, and thus by their own causality produce a whole, the conception of which, conversely ... could in turn be the cause of the whole according to a principle, so that, consequently, the nexus of efficient causes might be no less estimated as an operation brought about by final causes.”

By Kant's standards, we could say that an autocell is an organism though not a living one. Autocells lack many features we associate with life: They are exergonic, relying on the bonding energy intrinsic to catalysts and substrates rather than generating their own energy through work cycles. They have no replicator template—no RNA or

DNA—and their forms do not differentially survive through replications so much as self-reconstitution. Indeed, they are not even responsive to their environment. Though they alter between two states—closed shell/inactive autocatalysis, and open shell/active autocatalysis, they open and close by chance alone, not in response to environmental conditions.

Still, as simple as autocytes are, they nonetheless possess some of life's essential features in primitive barely recognizable form:

Individuality

Autocytes are robust to perturbation, as are all morphodynamic (self-organizing, dissipative) structures. Unlike other morphodynamic products however, autocytes are also robust to disrupted energy flow. In their closed form autocytes persist without energy throughput. As such, autocytes maintain a systemic individuality in the face of both material turnover (as in all morphodynamic structures) and intermittent energy flow. Thus, with autocytes we have the emergence of a primitive form of "self." The ability to persist without energy throughput is a necessary but perhaps not a sufficient condition for defining self-hood. The attributes describe below also contribute to the reasonable application of the concept of self to autocytes.

Value, Purpose, End directedness and for-ness

In a lifeless universe there are no extant selves in the service of which things can be said to occur. With hindsight, one could argue that occurrences in a lifeless universe are in the service of the future living. Still, science is far from demonstrating that life was

inevitable, so strictly speaking before there were entities to benefit from occurrences in the universe nothing was of value. There were no persistent forms to derive value.

The line at which value emerges is crossed with autocells. Of reciprocal co-production one could say that the container is good for the present (not future) autocatalytic set's perpetuation and the autocatalytic set is good for the present (not future) container's perpetuation, and further that occurrences in the autocell's environment are good for, or bad for present autocell lineages. There arises therefore a primitive form of value—of good and bad for an entity.

Adaptation, function, and aboutness

Autocells oscillate between two states—closed and open. Possessing more than one state with regard to some contextual feature is a pre-requisite for evolvability. It provides behavior that can be modified through an evolutionary selective process. Variations between autocell lineages make such selection possible. It provides the requisite variation upon which selection can act.

We say that something that is good for an entity serves a function for that entity. The autocatalysis, the container and the relationship between them are good for the autocell's longevity. “Function” refers to a structure or process within a dynamical context that embodies the potential to promote the continued persistence of the dynamics that sustains this potential. As autocell lineages vary, stochastically acquired features could increase or decrease in functional value to the longevity of the autocells. Thus one could say that with autocells, primitive function and the evolution of function arises. An observer could describe a feature's function in two respects, the ways in which a feature

was selectively retained in autocells and the ways in which the feature prepared the autocell for probable conditions in a stable environment in which past is to some extent prologue. With autocells we thus have a primitive form of evolution by adaptation, one that has no independent germ, one that is more Lamarckian than Darwinian but is nonetheless a product of blind variation with selective retention from which adaptive function emerges. Adaptive functions are elements of an entity that respond to and thus reveal something about the nature of the entity's environment. They are by definition in reference to, or "about," something other than the entity itself, something absent from the entity with which the entity fits. With the autocell's reciprocal co-production, the autocatalytic set as an entity persists because of the container—it is to some extent about the container. And likewise the container as an entity persists because of the autocatalytic set—it is likewise about them autocatalytic set. The synergy produced by this mutual aboutness gives rise to the autocell, which through evolvability acquires traits that are "about," features of the autocell's environment. Through adaptations, evolving autocells could be said to represent their environment, as a shoeprint could be said to represent a shoe. With autocells there is representation without a separate medium for representation, without either DNA or RNA.

Conclusions

Bateson sought the connection between pleroma and creatura in two respects. He wondered how creatura interfaces with pleroma in cognition. That is, how we could ever know the *ding an sich*--a problem addressed in Cashman's paper. Bateson also wondered how creatura could emerge from pleroma; the ultimate problem of evolution, and the

implication of the necessary dependency of creatura on pleroma. This is the question we have addressed in this article.

Our most ambitious claim is that autocells fill the missing link between non-life and life. Though we do not yet have an experimental proof of concept, nevertheless, the autocell model is testable and requires no special physical or chemical assumptions. We also cannot provide conclusive evidence demonstrating that life as we know it begins where autocells leave off. Though the conceptual thrust of this model does not depend on its relevance to the origins of life, only on the fact that it embodies the critical principles demonstrating that it can emerge from self-organizing processes alone. Still, autocells, as described, fill a significant gap in the mystery of how life could emerge from non-life, and elsewhere one of us (Deacon, 2006) has provided a detailed description of how autocells might eventually evolve into life.

Even if we cannot yet say that life on earth emerged from a prior autocell stage, this model demonstrates that there is an unbroken continuity from thermodynamics to evolvability, and that it is possible to show how life could emerge from non-life following exacting standards of scientific credibility and feasibility. Were Bateson alive today and witness to these further developments of the evolutionary, cybernetic, semiotic, and complexity theories that underlie the autocell model, he would likely have found reason to abandon his tentative gestures toward idealism. The core questions that motivated his life's work focused on the possibility of identifying the pattern which connects life to mind and, ultimately, pleroma to creatura. His eventual—perhaps reluctant—conclusion that such a connection might only be achievable by an act of faith, appears to have been premature.

Bibliography

Kant, Immanuel (1790) *The Critique of Judgment: II. Teleological Judgment* (translation by James Creed Meredith). Reprinted in *The Great Books*, Vol. 42, (1952) Chicago, University of Chicago Press.

Prigogine, I. & Stengers, I. *Order out of Chaos* (Bantam Books, New York, 1984).

Kauffman, S.A. *The Origins of Order: Self-Organization and Selection in Evolution* (New York: Oxford University Press, 1993).

Kauffman, S.A. (1986) Autocatalytic sets of proteins, *J. Theor. Biol.* **119**, 1 (1986).

Deacon, T. (2006) Reciprocal linkage between self-organizing processes is sufficient for self-reproduction and evolvability (in press *Biological Theory*)