TELEOLOGY FOR THE PERPLEXED: HOW MATTER BEGAN TO MATTER

by Jeremy Sherman and Terrence W. Deacon

Abstract. Lacking a plausible model for the emergence of *telos* (purposive, representational, and evaluative relationships, as in life and consciousness) from simple material and energetic processes, the sciences operate as though all teleological relationships are physically epiphenomenal. Alternatively, in religion and the humanities it is assumed either that *telos* influences the material world from an outside or transcendental source or that it is a fundamental and ineffable property of things. We argue that a scientifically sound and intuitively plausible model for the physical emergence of teleological dynamics is now realizable. A methodology for formulating such a model and an exemplar case—the autocell—are presented. An autocell is an autocatalytic set of molecules that produce one another and also produce molecules that spontaneously accrete to form a hollow container, analogous to the way virus capsules form. The molecular capsules that result will spontaneously enclose some of the nearby molecules of the autocatalytic set, keeping them together so that when the autocell is broken open autocatalysis will resume. Autocells are thus self-reconstituting, self-reproducing, and minimally evolvable. They are not living and yet have necessary precursor attributes to *telos*, including individuality, functional interdependence of parts, end-directedness, a minimal form of representation, and a normative (evaluational) relationship to different environmental properties. The autocell thus serves as a missing link between inanimate (nonlife) and animate (living) phenomena. We conclude by discussing the challenges that a natural origin for *telos* poses for religious thought.

Keywords: apophatic; autocell; emergence; evolvability; meaning; morphodynamics; origins of life; purpose; self-organization; supernatural; teleodynamics; teleology; *telos*; theology; thermodynamics

---

Jeremy Sherman is a professor of social science at Expression College of the Digital Arts, 1830 Sonoma Ave., Berkeley, CA 94707; e-mail js@jeremysherman.com. Terrence W. Deacon is a professor of biological anthropology, Department of Anthropology and Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720; e-mail deacon@berkeley.edu.

[*Zygon*, vol. 42, no. 4 (December 2007)]
 ISSN 0591-2385

Among the issues of greatest and most pressing interest at the interface between scientific and spiritual understandings of reality are questions regarding the nature and origins of teleological phenomena: end-directed processes and properties such as functions, representations, intentions, purposes, meanings, values, and of course subjective consciousness. We refer to this diverse array of phenomena as all exhibiting a general property we will call *telos* (from the Greek: end, aim, goal, purpose, completion, fulfillment), referring to their common feature of being organized with respect to some end or intended content, and closely related to Aristotle's notion of a final cause—that for the sake of which something exists or is done.

Unfortunately, an immense logical chasm appears to exist between explanations of things given in the terms of *telos* and explanations given in terms of the familiar pushes and pulls of physics and chemistry. For the most part, the history of the natural sciences during the past two centuries has been characterized by a systematic effort to eliminate teleological explanations. This is because they are essentially truncated explanations, accounts of phenomena that point to black boxes and then stop. To say that an intention, belief, or desire is the cause of something does no more than point to some location, typically in a human agent, without saying anything about the specific details of the mechanism involved. There is very little doubt that physical-chemical processes taking place in a body are critical to the physical consequences that ensue, but such an account says nothing about the relationships that link these processes to the mental representations that human experience tells us were the origins of this process. So everyday human experience appears to result from an intractably contradictory combination of clockwork and purpose.

Contemporary science and philosophy have not yet found an acceptable way to deal with this dilemma that retains the precision and completeness of a scientific account yet also does justice to the distinctiveness of teleological processes, particularly the unique internal subjective experience of representation and agency. This is not necessarily a problem for those in the two extreme camps who are either satisfied with an account of teleology as ineffable mystery or convinced that it is purely epiphenomenal. Ultimately, however, lack of a constructive scientific account of *telos* is a dilemma for any more modest view that holds to both the rigor of science and the undeniable reality of teleological phenomena. This is a dilemma that must ultimately be resolved within science, by showing how *telos* can be consistent with natural science and not merely impotent and illusory.

Explaining and Explaining Away

Historically, we can discern four main categories of attempts to explain or explain away *telos*:

Preformationist answers posit that these special phenomena are already present fully formed in the very fabric of the universe. For example, some

argue that the mental realm with its implicit meaning and value is preformed in the mind of God, or that information is a basic component of all patterned physical phenomena so that it is implicit in quantum events and intrinsic to DNA molecules.

Eliminativist answers posit that these special phenomena not only aren't as special as they appear but in fact are conceptual mirages, the residue of prescientific thinking. For example, some argue that consciousness is nothing more than chemical processes in brains—that mental representation and the experience of agency are illusory and do not reflect the real underlying clockwork determinism of neurochemical mechanisms.

Mysterian answers posit that it is beyond the human intellect to understand the nature of teleological phenomena because their relationship to physical processes is in some way unknowable in principle. For example, some argue that, in the same way that a dog's cognitive limitations render it incapable of ever conceiving of how an internal combustion engine works, we may be limited by our evolutionary endowment to never being able to conceive of how consciousness works. A slightly more mysterian version of this is eloquently stated by Douglas Hofstadter: "Our very nature is such as to prevent us from fully understanding its very nature" (2007, 363). These express the mystery in terms of human or mental incapacity in general but do not go so far as to suggest some deeper ultimate mystery. But what if there is something about teleological explanations that makes them fundamentally incompatible with physical explanations? Or what if they actually *are* incompatible modes of causality? Could they actually arise from independent realms—for example, the material versus spiritual—that operate according to incompatible principles?

By an interesting convergence, preformationism and mysterianism often join forces. This is because if *telos* is ultimately mysterious and yet is a real influence in the world, it must be part of the very fabric of the world, unanalyzably necessary as a basic axiom, underived from any more fundamental principles or properties. After all, if it must be preformed, it cannot be derived. It is a black box that cannot be opened. So we might describe this common synthesis the mysterious preformation view.

Notice that none of the above alternatives is an attempt to explain the nature of *telos*. They are rather efforts to explain it away or to at least explain away the problem. They are halting moves that in one way or another keep their adherents from having to deal with the dilemmas *telos* presents. What, then, of serious attempts to explain how *telos* actually works in the world?

Proposals that attempt to offer scientific explanations of the nature and origins of these phenomena can be found clustered under the heading of *emergence*. They are efforts to have it both ways; that is, they consider *telos* as fully compatible with physical causality and yet attempt to show how such phenomena can exhibit causal properties that are unprecedented in

the physical and chemical sciences. In these theories, teleology is presumed to "emerge" spontaneously *under certain conditions* from physical antecedents lacking these properties. We believe that some version of an emergence explanation must be correct, and we are convinced that only a scientific effort to take teleological phenomena seriously, and understand it as emergent rather than illusory or inexplicably fundamental, can do justice to both scientific rigor and to its indubitable specialness.

The Burden of Proof

To our knowledge, no theory of emergence currently is adequate to this task. To demonstrate how *telos* could have emerged from merely physical chemical processes in a rigorous way that is compatible with the best theoretic and empirical science is not just a daunting challenge; some would argue that the idea is self-contradictory. In the face of this challenge, we offer only a first step: a proof of principle with respect to the most minimal conceptions of *telos* associated with the dawn of life, and the notions of function, self-determination, evaluation of circumstances, and end-directedness that this entails. However, in consideration of doubts that any such bridge is even conceivable—much less a scientifically meaningful and empirically promising avenue of research—even a minimal demonstration of emergent *telos* provides a fundamental counterexample to preformationist, eliminativist, and mysterian claims. This would be sufficient to warrant a radical reconsideration of these deepest rifts between the natural sciences and both humanistic and theological paradigms.

Life and consciousness were not around at the time of the Big Bang. These, and the many other higher-level teleological properties of human mentality (like meaning and value) emerged over evolutionary time. We argue that the first and simplest traces of these teleological phenomena to emerge were exhibited by the very basic processes that are associated with the origins of life. But, despite its necessary simplicity, we believe that this threshold between animate and inanimate matter is every bit as troublesome as that between mind and body.

Despite claims that the secret of life has been unmasked with the discovery of DNA and that life is therefore "just chemistry," this reductionistic optimism hides some unmentioned cards up its sleeves. It assumes, for example, that DNA is information, by analogy to human codes, and yet fails to say how it is that a molecule can come to carry information about other molecules' structures and interactions, and specifically uses our human interpretive capacity to see this correspondence *as* an information relationship but fails to explain how, without human interpretation, it is intrinsically informational. This plays on an ambiguity recently introduced into the use of the term *information*. In one sense it refers to merely a pattern in some medium, like the pattern of lighted dots on a computer

screen, but in another sense it refers to that content which is conveyed by those patterns. Most important, we refer to the pattern as information only because we anticipate that it is *interpretable* as being about something else. In other words, in this conception of life we have already smuggled in a fundamental teleological concept without explaining how it arose.

Living chemistry with teleological qualities arose from nonliving chemistry. This transition from nonteleological to teleological chemistry marks a key turning point anywhere in the cosmos where it has occurred. Physical and chemical processes will exhibit very distinctive features wherever life has appeared, and if it evolves for any appreciable period of time this will begin to have planetary-level effects, as for example occurred in the history of life on earth when life restructured the atmosphere by injecting vast amounts of that highly reactive molecule oxygen into it. Although an understanding of the transition from chemistry to life, mechanism to organism, non-*telos* to *telos* will not fully answer the many conceptual challenges posed by mental phenomena, until this first emergent step is thoroughly understood it will not be possible to approach the vastly more complex higher levels of teleological dynamics that have evolved subsequently from these humble beginnings.

Lacking a detailed understanding of the emergence of teleology, we cannot claim to have a complete scientific theory of either living or mental phenomena. But even just accounting for the emergence of the very simplest form of teleology is bound to have profound implications for the legitimization of teleology in the sciences and could also provide a first step toward a realignment of scientific interests with metaphysical and theological concerns.

A synthesis between science and spirit has long been sought. The progress that has been made is mostly rhetorical, however—scientists and theologians recasting their theories in ways that appear to accommodate each other without changing any fundamental assumptions. The wedge issue underlying this impasse is not science versus theology, it's mysterious preformationism versus eliminativism. Indeed, the impasse between these two irreconcilable dogmas is wider than the science-religion debate. It forms the great invisible and widening gulf on all college campuses today, the gulf between those humanistic academic fields that rely upon *telos* as the primary explanatory principle and those that insist that *telos* be ignored in the name of scientific rigor. No progress can be made on the synthesis of science and spirit without consensus first on the question of *telos*.

In this essay we sketch out both a research methodology and a specific testable and physically feasible model for how *telos* could emerge from non-*telos*. Our goal is to identify the critical steps by which simple phenomena with end-directed attributes could arise from nonliving and nonteleological physical-chemical processes. To whatever extent we succeed, we will have begun to erode the fundamental claim supporting both eliminativism and

mysterian preformationism, that such an account is unachievable in principle.

Indeed, because the physical origin of *telos* has never been easily imagined, for millennia conventional wisdom has gravitated toward the assumption that *telos* must be endowed from an outside transcendental source or that it must be an a priori principle implicit in all things. Otherwise, it would be rendered illusory. Both religion and philosophical traditions have been drawn to these alternatives. To the extent that any model system can demonstrate the feasibility of a physical origin for *telos*, it will open the door for alternatives to theories requiring transcendental or intrinsic *telos*.

Methodology and Model

Our approach to the physical origins of life and *telos* differs from other approaches to emergence in four respects.

First, in contrast to many other researchers' use of the term, we do not use *emergence* as a label for irreducibility, to indicate a failure of reductionistic explanation, to claim a causal disconnection between properties in hierarchically nested domains, or to suggest the introduction of cosmologically unprecedented kinds of causation. Rather we employ a dynamical and conditional conception of emergence. Emergence identifies a special class of physical transitions at which, under certain specifiable conditions, abrupt reorganization of global causal dynamics occurs. Consequently, we do not attempt to survey the metaphysical arguments for and against an emergent cosmology and instead assume that, for the phenomena we describe, an unproblematic use of the term will be acceptable to reductionists and emergentists alike.

Second, we assume, with all but the eliminativists, that teleological phenomena exist and have robust physical consequences, but we also assume (contra mysterian preformationists) that such phenomena did not always exist and that they came into existence out of nonteleological antecedent conditions. We are only interested in specifying how and under what conditions its emergence occurs.

Third, we are not working to provide an account of the origin of life on Earth, especially because research aimed at that goal is often myopically focused on specific classes of chemical reactions and planetary conditions considered consistent with the special and likely contingent features of Earth life and the primordial Earth environment. We are, instead, interested in the generic principles that are intrinsic to the emergence of any possible lifelike form, anywhere (although we will almost certainly underestimate the scope of possible forms). More important, all current approaches to the origins of life on Earth begin from tacit assumptions about the nature of teleological processes (typically assuming some version of eliminativism—for example, that it is "just" chemistry—plus cryptic

preformationism—for example, that DNA-like molecules embody intrinsic information) whereas our interest is in reexamining these assumptions.

Fourth, we do not resort to speculations about the possible role of mysterious forces beyond the fringe of current scientific understanding, such as paranormal forces, or intangible "morphogenetic fields." And we avoid passing the explanatory buck to strange aspects of physics, such as quantum effects. These are not considered in our approach to the emergence of *telos* for two reasons: first, because we believe we can show that the basic physical and chemical forces that dominate at our level of scale are sufficient to explain the emergence of life and the forms of *telos* that life exhibits, and more important, because this indulges in an implicitly mysterian maneuver to claim to have explained one mystery by invoking another.

Is a simplest life form (or protolife form) sufficiently different from "physics and chemistry as usual" to be useful as a model for the origins of teleological phenomena? We believe that we can demonstrate this constructively; however, to justify the effort in anticipation of this account, it is sufficient to consider what we already know and assume about even simple life forms. To the extent that living, evolving organisms exhibit functions, not merely chemical reactions, we cannot help but use explanatory tools that invoke end-directedness in some form. We describe even simple organisms as exhibiting adaptations or functions with respect to something like their "own good." They encounter favorable or unfavorable environments and have needs or appetites for some of what they find there. They compete to maintain themselves and their lineage. We recognize in organisms, then, the most basic analogues of what in our mental experience we describe as self, intention, significance, desire, and purpose. These attributes, even in attenuated minimal form, are significantly unlike anything found spontaneously in the nonliving world.

Spontaneous Generation. Although natural selection theory is often assumed to account for the emergence of novel forms of *telos* in biology, it necessarily assumes the prior existence of reproduction, function, and so forth. This is why it cannot be invoked to explain the spontaneous origin of life. This suggests a useful analogy. From early Roman times it was thought that some, if not all, life arose from inanimate matter by way of spontaneous generation. Evidence to support the theory of spontaneous generation was seen in the way maggots would emerge spontaneously from rotting beef. In 1668, in one of the world's first controlled biological experiments, Francisco Redi challenged the theory by demonstrating that maggots did not emerge from meat when put under glass. Belief that life emerged spontaneously from nonlife nevertheless died hard. Support for spontaneous generation persisted for more than two hundred years, until 1889, when Louis Pasteur, using sterilized rotting meat in a flask with an s-shaped neck, demonstrated that maggots do not grow spontaneously, so long as outside contamination is prevented.

Analogous to Pasteur, the present approach puts the test under glass in order to avoid contamination. We must imagine how, within a universe completely devoid of contamination by life or *telos*, life's *telos* could emerge. Ironically, we are not trying to disprove spontaneous generation; we are trying to demonstrate it rigorously, in a very restricted sense.[1] Evolution ultimately demands that, at least in a minimal sense, spontaneous generation is possible. Initially, life *was* generated spontaneously. The first life form was not reproduced, it had no parent, it did not evolve; it emerged. Evolution itself must have emerged, because it makes no sense to argue that evolution evolved by evolution.

Surrounded by teeming life so imperceptibly minuscule as to require extreme care to avoid contamination, spontaneous-generation theorists were fooled again and again. Trying to explain how *telos* emerges from within our current teleologically rich environment risks analogous contamination. Setting our experiments in a teleologically barren context is thus the equivalent of testing spontaneous generation in a sterile environment.

Amnesic Watchmakers. Scientific inquiry into the origins of life typically employs a reverse-engineering approach. That is, researchers examine current life forms and extrapolate backward, asking how the whole and its parts got put together. This often leads to a focus on one or another critical attribute of life as a plausible starting place: typically either its information molecules that make reproduction possible, its metabolic machinery that maintains an organism in a nonequilibrium state, or its lipid membrane container that selectively keeps critical components inside, troublesome molecules outside, and selectively influences which can pass in or out.

For example, biologists recognize that reproduction, supported by the replication of informational molecules (that is, DNA), is a fundamental feature of all life forms. Many origins-of-life researchers have therefore postulated that life began with something like a proto-replicator, a first self-copier (see Dawkins 1976; Maynard Smith and Szathmáry 1999; Woese 1967; 1998). So what about the accidental synthesis of a first "naked replicator" molecule (Dawkins 1976)? Given the truly cosmic improbability of such a complex accident, such an explanation is no more scientific than invoking divine miracle or intelligent design. Besides, it is not merely the replication of a molecule that matters but a complex chemical relationship that both defends against degradation and reproduces this capacity.

Likewise, since all life depends on mechanisms that operate to resist degradation and dissolution under the influence of the second law of thermodynamics, something like a metabolism is nearly ubiquitous (except for viruses which are parasitic on organisms with metabolisms). Many other origins-of-life researchers argue that life began with something like a proto–work cycle: a first, simple cyclic chemical process that makes more identical parts (molecules) faster than they break down (Kauffman 1986; Eigen

and Oswatitsch 1992; Shapiro 1988). And because all life is contained in cell membranes that define inside and outside, self and other, still other origins-of-life researchers have focused on the spontaneous appearance of the first lipid enclosures as the critical first step in the genesis of life (Deamer and Barchfeld 1982; Hanczyc, Fujikawa, and Szostak 2003). Some laboratories are attempting to combine all three features into structures called protocells (Szostak, Bartel, and Luisi 2001; Rasmussen et al. 2004).

If the goal of origins-of-life research were simply to engineer life, building a successful protocell would be success in itself. But engineered life is not tantamount to the spontaneous emergence of life, because engineering is precisely what a prelife universe lacks. Does the simplified combining of components found in living cells provide an adequate picture of an *unengineered* missing link between physics and biology, or is it more like a sort of Frankencell, reconstructed from components extracted from once-living cells?

In three respects, the emergence of life from nonlife is more challenging than the reverse-engineering approach would suggest. First, we cannot invoke prior teleological processes to explain either component fit or the means of their combination. Second, the molecular components cannot be the products of a prior evolutionary process, only spontaneous geochemistry. Third, components cannot merely be brought into proximity with each other, they must reciprocally produce one another and maintain these proximity conditions (for example, by generating containment). Merely collecting the critical molecular components of cells into a cell-like container is not enough, even if each performs chemical functions characteristic of those produced in life. This even goes for the self-replication of nucleic acids. In the absence of this synergistic coproduction and maintenance of reaction proximity there is nothing more than organic chemistry in a lipid reaction vessel.

To show how life can spontaneously emerge, unaided by intelligent intervention or astronomically unlikely lucky accident, we cannot employ component parts that have already been shaped by evolution for functional synergy, because that is what needs explaining. To meet the emergentist challenge therefore requires vigilance to avoid what could be called the Amnesic Watchmaker Syndrome. Imagine a watchmaker, with a serious case of Alzheimer's disease, absentmindedly fitting together parts from a previously disassembled watch. Forgetting where these parts came from and finding that they fit together only in certain ways, he combines them in many alternative configurations until he eventually fits them together so that collectively they function to tell time. Astounded by this, he muses that the ingredients for watches may be strewn about the world naturally and that time-telling machines might just fall together spontaneously by accident.

This is not to imply that protocell researchers and other origins-of-life theorists are unaware that they are often working with already evolved components. Nor does it suggest that nothing can be learned about the basic principles of time-telling mechanisms or living mechanisms by exploring the way the parts interact. Indeed, this is the probably the best way to understand firsthand how the basic processes work, *when the function is already presumed.* However, it may not be the best approach toward understanding how this functional logic itself came to exist in the first place, either for life or for watches. Exploring the interactive relationships among components that are already evolved for their functional contributions to a living cell can provide important insights about some of the basic processes of life, but it would be unwise to mistake these as modeling life's origin. But most important for investigating the origins of life's *telos*, this amounts to a spontaneous-generation experiment that is contaminated from the start by the *telos* of prior life.

Outline of a Theory of Emergent Dynamics

To set the stage to talk about the emergence of purposiveness from nonpurposiveness we must first have a clear operative definition of emergence. For many researchers, as we have seen, emergence connotes unexpected and novel shifts. These might be imagined as transitions that jump discontinuous gaps into new realms of causal properties. But this conception of emergence leaves cracks into which crypto-teleology can slip. So we must employ an operational definition that does not allow such gaps but will allow us to see how radical reorganizations nonetheless arise.

As we demonstrate below, what reliably characterizes and distinguishes emergent phenomena at all levels of organization is the transitional state between levels of dynamical organization. Emergence is a kind of phase shift between causal regimes whereby the net effect of the interactions between parts of a system introduces regularities not exhibited in the properties of the parts independently. Under certain conditions macro-regularities emerge, the properties of which, while dependent on individual microphenomena, cannot be decomposed into parts and isolated interactions without having the macro-properties disappear.

The Thermodynamic Universe. By our definition of emergence, the simplest kind of emergence is not that found in the behavior of living systems, nor even in the transition to life from nonlife, but rather in the behavior of physical thermodynamic systems. Hot and cold regions of liquid when combined become uniformly warm through the difference-cancelling interactions of the constituent molecules collisions.

This emergent effect is a consequence of the second law of thermodynamics. The discoverers of thermodynamics (Rudolph Clausius, James Maxwell, and Ludwig Boltzmann, among others) explained this phenomenon

by noting that molecules before they interact have uncorrelated movements and that after they interact they are even less likely to have correlated movements—that is, to be moving in nearly the same direction with the same force. In the aggregate, therefore, interactions will tend strongly to decrease correlations, and with every decrease in correlation a counter-increase in correlation becomes more unlikely.

For example, the correlations among molecular movements in the heated air of a cabin in winter are high within the hot region close to the heating stove and high within the cold region near an opened outside door. In other words, the velocity of any molecule in the hot region is highly correlated with the velocity of its neighboring molecules. With the door shut and the stove off, however, the second law of thermodynamics takes effect and the correlation decreases, so that eventually there is no higher or lower correlation of molecular movements in any part of the room. It has reached equilibrium.

A collection of interacting molecules (for example, a container of water) has no *telos*, no integration of parts, no "for the whole" contribution of one molecule to the rest, but still the global property—the tendency for the whole to equalize—does emerge. This tendency toward equalization is not imposed from the outside, and is not a property of the individual molecules in isolation. It is not novel, introduced into the universe from scratch to surprise an unsuspecting viewer. Nor is it an arbitrary eye-of-the-beholder epistemological property reflecting some personal definition of order versus disorder. It relates to order defined objectively as correlation. Before the separate regions mix, the correlation from one molecule to the next is higher than after the second law has taken effect. From the second law's system-wide tendency emerges a species of causal efficacy that is not reducible to the causal efficacy of the system's individual molecules and their interactions alone. The global distribution of these tendencies turns out to be the critical factor, and this changes even if the system is completely isolated and no energy enters or leaves.

Random Collisions and Geometric Biases. In our imagined universe without purpose, molecules interact thermodynamically, colliding, rebounding, sometimes sticking, sometimes bonding by entangling their component electrons and becoming larger molecules. Their likelihood of bonding is in part a function of the direction and momentum at which molecules come in contact with each other, but it also is a function of the molecules' particular shapes. All of what makes chemistry different from billiard-ball interaction is a consequence of shape effects, of molecules and their orientations, and how such shape effects bias what is likely and unlikely to occur over and above thermodynamics' relentless evening out of distributional asymmetries. 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 nevertheless exhibit Van der Waals attraction to one another (the molecular "stickiness" that provides the cohesion of the liquid state), in some orientations their mutual stickiness 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 length of time molecules tend to stay attached is a function of the quantity of conforming surface area between them. The closer the fit between the shapes of two molecules, the longer and more tightly they will stick in exactly that orientation.

This differential stickiness is purposeless, devoid of all teleological character. 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, with little effect on the march toward chemical equilibrium. And yet, given differences in shape, some molecules are more likely to stick to each other and more likely to be oriented in certain ways when they do.

One major mechanism underlying this potential is catalysis, a result of shape-dependent differential molecular stickiness that affects rates of chemical reaction (that is, formation of ionic and covalent bonds, which determine molecular structures). Catalysis occurs when one or more molecules mediate and potentiate specific reactions of other molecules, but where the biasing molecules do not get permanently altered in the process. Catalytic molecules temporarily capture other molecules with complementary shapes briefly holding them in a specific orientation. This 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 catalyst molecules retain their shape and therefore retain the ability to cause still other molecules to break apart or come together.

Catalysts are not teleological. Their biasing effect is merely a result of chance shape correspondences. But shape biases can 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 directionality, a trend toward the increased transformation of certain molecules to certain other molecular forms.

The statistical bias effect of catalysis is dependent upon the relentless shuffling of thermodynamic interactions. But because of this bias the probabilities of certain chemical reactions are significantly greater than others, and quite different than in a noncatalyzed system. So although a catalyzed chemical system will still reach thermodynamic and chemical equilibrium if not further perturbed, certain reaction rates are significantly augmented,

often by orders of magnitude. In this sense a catalyzed reaction is not *merely* subject to thermodynamics. This is particularly relevant for systems that are persistently maintained away from equilibrium (for example by being constantly replenished with certain substrate molecules or reaction energy), since the catalytically facilitated reactions will produce a rapidly increasing asymmetry of reaction products. Such a system will not only remain far from equilibrium but in this one respect will progressively diverge away from equilibrium. This spontaneous increase in asymmetry is a morphodynamic effect, in which one asymmetry generates an increase in another.

Autocatalysis and Self-assembly. The molecule-specific bias of catalysis not only allows nonequilibrium systems to exhibit increasingly deviant effects but also opens up the possibility for the specific form of the catalytic reaction to play a significant role in the dynamics. Molecule-specific interactions between specific catalytic pathways can produce runaway effects even in unperturbed systems, to the extent that they reinforce one another. Ordinarily, catalytic reactions are self-limiting. First, like any chemical reaction they tend to run to equilibrium, where global concentrations stop changing either because they have depleted raw materials (that is, catalyzable molecules) or because reactions in each direction become equiprobable. Second, the concentration of catalysts is itself a rate-limiting factor. The first limitation can be mitigated in open-system conditions where new substrate molecules are continually added or product molecules are continually removed. The second limitation can be overcome in the special case of reciprocal catalytic relationships, or autocatalysis.

Autocatalysis involves a circle of catalytic reactions producing catalysts. This condition is not particularly difficult to obtain spontaneously. In a solution containing many diverse kinds of molecules capable of catalytic effects, chances are fair that two or more catalysts will be mutually reinforcing in their biases (Kauffman 1986; Farmer, Kauffman, and Packard 1986), in which one catalyst contributes to the synthesis of 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. Autocatalysis is a special case of this relationship in which a circle of catalytic reactions occurs, such as when the product of a catalytic reaction is also a catalyst and catalyzes the synthesis of the catalyst that produced it. The molecules of such a catalytic circle constitute an autocatalytic set (Prigogine and Stengers 1984; Kauffman 1993); the reciprocally reinforcing relationship between catalysts produces a kind of runaway effect. This logic can be extended to any number of catalysts.

Consider an oversimplified example. Imagine a catalyst A that catalyzes the synthesis of a second catalyst B. Imagine then that catalyst B catalyzes the synthesis of catalyst C and finally that catalyst C catalyzes the synthesis of catalyst A. Starting with one of the catalytic molecules in an autocatalytic set and abundant raw materials for each catalytic reaction, the amounts

of all catalysts in the set double with each catalytic cycle. This produces an accelerating effect because each cycle produces more catalysts producing more cycles for as long as there are molecules available to be catalyzed. But molecules to catalyze will be depleted very rapidly, so although autocatalytic cycles diverge from initial proportions quite rapidly, this typically is a very short-lived trend in a closed system.

Autocatalytic sets of molecules are coherent sets in theory only. Unlike the metaphoric factory production line, there is no dedicated linkage between the autocatalytic 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. So although autocatalytic sets have a morphodynamic causal efficacy, shifting a chemical system out of equilibrium locally, they are merely ephemeral constellations of independent molecules. Their membership in a "set" is an extrinsic factor, identified by chemists observing this regularity and its consequences, but has no independent reality besides this. An autocatalytic set will spontaneously disperse and its synergy will be irrelevant as soon as substrate molecules are depleted to the point where replacement of catalysts in a region falls below the rate of their spontaneous diffusion out. We return to this problem below.

Catalysis and autocatalysis are not the only chemical mechanisms that can produce biases in a thermodynamic trend at the molecular level. Self-assembly is a self-reinforcing pattern of molecular binding such as also occurs in crystallization. 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, like building blocks that connect readily from one orientation to form a wall. Self-assembled structures take various forms, depending on the symmetry of the molecule. Many molecules that bind into complexes produce clump or crystal-like structures, but some form regular hollow shapes like tubes or polyhedrons. The shells that encapsulate many viruses are well-known examples of self-assembled molecular containers, but this is not specifically a biological phenomenon. Self-assembly occurs spontaneously because molecules that are shaped so that they hydrogen bond with neighbors to form regular arrays are at a lower energy state than when freely floating. As such structures grow, the number of facets in which new molecules can fit increases as well. If such structures form molecular sheets they can impede and constrain spontaneous diffusion processes, and if the sheets fold on themselves to form hollow structures they inevitably will form around and enclose other molecules as though in a molecular capsule. A container of this sort can completely block diffusion between inside and outside.

The processes described so far do not exhibit teleological properties. However, the not merely thermodynamic but also form-producing tendencies of these processes are the necessary stepping-stones from thermo-

dynamics to a simple form of teleological dynamics. In a specific combination these form-generating processes can cooperate to provide yet another emergent level of causal organization.

Autocatalytic sets by themselves dissipate, but it is possible that one or more of the catalysts or else molecules produced as by-products of an autocatalytic cycle could be of a variety susceptible to self-assembling into hollow structures. These self-assembling molecules would spontaneously form shells in proximity to molecules of the autocatalytic set that produced them. Shells so produced would therefore be likely to enclose a representative sample of the molecules constituting the autocatalytic set. In this case the otherwise independent autocatalytic molecules would remain in close proximity. We call such self-enclosing autocatalytic sets *autocells* (Deacon 2006). The name refers to the cell-like encasement of an autocatalytic set. We should be careful to point out that they are not cells in the usual sense, yet they have interesting lifelike properties that we now discuss.

Autocell Properties

Self-sustaining Synergy. Although simple, autocells are sufficiently complex molecular systems to illustrate how simple teleological processes can emerge spontaneously from self-organizing processes. The key feature is not any single type of molecule or process so much as the *synergistic relationship between processes that reciprocally support one another's persistence* (see also "autopoiesis," Varela et al. 1974). But this does not depend on the

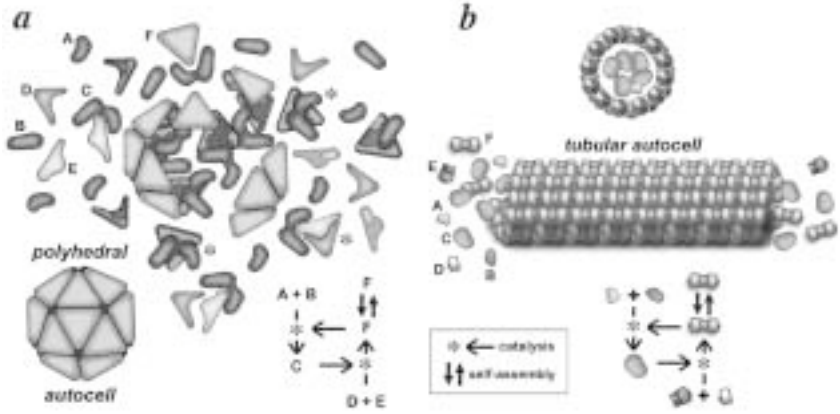

Figure 1. Two artist's conceptions of autocell architectures: polyhedral (a) and tubular (b) forms, with the linked autocatalysis and self-assembly logic depicted by arrow diagrams. To view an animation of the autocell model go to *www.teleodynamics.com*. For a more detailed description of the autocell mechanism, see Deacon 2006.

continuous persistence of self-organizing processes, or any chemical reaction, only the *potential* of their persistence. Thus, paradoxically, one of the 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 catalytic processes to run down more rapidly than if free-floating, ceasing altogether shortly after closure. Nonetheless, enclosure also keeps catalytic molecules from diffusing away from one another, maintaining close proximity to one another despite their chemical inactivity. So enclosure by self-assembly temporarily stops its chemical activity but also limits molecular dissipation that would permanently undermine autocatalytic capacity.

Enclosure inevitably is a temporary condition. Molecular shells are buffeted about, as are individual molecules. As a result, they occasionally will break, spilling their contents, allowing their previously sequestered catalysts to again come into 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 reconstituting a new shell closing 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 maintain continuity of structural characteristics over an extended number of "generations." 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.

It is this systemic interdependence, or synergy, and not any component molecules or chemical reactions that is the defining property of an autocell. Autocellularity is not decomposable to any of its component molecules or reactions, even though they are necessary components. This special complementary relationship that exists between the two kinds of self-organizing processes in an autocell licenses calling these processes *functions*. They are appropriately described as functions 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 promoting this self-assembly an autocatalytic set *functions* to generate its own protection against disruption; the shell protects the process that produces it.

Autocatalytic sets are merely abstractions that have no concrete individuality. They 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, autocells

are concrete individuals, not mere abstractions. They exhibit and maintain their distinctive properties and resist disruption or modification. For this reason, within an autocell an autocatalytic set is not merely an abstraction, either. These same reciprocal catalytic relationships can be described as functional, precisely because they now play a role in ensuring the persistence of this autocatalytic potential. The components of an autocatalytic set act as a distinct individual component of the larger individual of which they are a part. Autocellular synergy thus illuminates a critical defining feature of that form of *telos* we describe as function and that form of unity we describe as a self.

There is nothing about an autocell that is greater than the sum of its parts; nonetheless, a new kind of causal organization is exhibited by autocells that is irreducible to the precursor causal processes outside of this particular configuration. More precisely, although autocells are analyzable into component molecules and chemical reactions, autocellularity and the special properties of self-reconstitution and self-replication that it creates cannot be decomposed into simpler forms of these processes. The higher-order functional dynamics exhibited by autocells is in this regard an emergent consequence of this recursive and codependent relationship between self-organizing processes. It exists only because of this synergy of dynamical processes and their synergy with conducive environmental conditions (that is, one containing the appropriate substrate molecules). So, once formed, autocells take on a "life of their own" as causal loci of an unprecedented 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 gives autocells a form of objective individuality that is quite concrete and distinctive.

Evolvability. The autocell example shows us that a configuration of complementary self-organizing processes—one metabolic in the sense of transforming molecules from one form to another, and the other maintaining collective integrity through the formation of a virus-like shell—makes possible a prototypical self-reconstituting system that has distinctive individuality and simple selfhood. Although the autocell lacks many necessary features for life, it is an individual precisely because it is the source of causal efficacy for maintaining this identity and integrity. Autocells may even be capable of a primitive form of evolution. Let us imagine how autocell evolution might arise.

Shell-molecules breaking apart and coming together will tend to form multiple shells as easily as forming one alone. Multiple shells forming in and around molecules of an autocatalytic set will also inevitably capture other molecules in this vicinity, some of which may be slightly varied versions of autocatalytic molecules. This will result in variation of autocells. Many shells will contain incomplete complements of the necessary set for

reinitiating autocatalysis when opened. Many shells will open in solutions devoid of catalyzable molecules, and their components will merely dissipate. But where a full set of catalysts is enclosed there is the potential for the vague equivalent of multiple "offspring." Breaking open by happenstance in the presence of catalyzable molecules, the contents of one autocell could initiate the formation of several autocells, each with different random combinations of contained molecules.

Because of their periodic openness to the surroundings, given enough time and enough persistence the autocatalytic set characterizing an autocell could come to interact with elements not in the original set. Some such interactions would undermine catalytic productivity. For example, new molecules entering the mixture might degrade other catalytic contributors, or merely interfere with catalysis or self-assembly. In such cases autocatalysis would significantly slow and decrease the effectiveness of the process of reenclosure. Autocells thus handicapped would replicate more slowly and fail to reconstitute more often than others, thus leaving fewer "progeny."

It is possible, however, that new catalytic relationships could accumulate without undermining the original autocatalytic set's productivity. For example, alternative catalysts might be encapsulated that are even more effective than an original catalyst, or that provide an additional complementary catalytic route to autocatalysis, increasing the probability of cycle maintenance. This might increase catalytic versatility, making the more complex and flexibly survivable autocatalytic set productive even in environments that lacked the raw ingredients transformable by the original catalytic elements. We can even imagine a handing off of function from the original to an alternative autocatalytic element or subset acquired in this way, thus producing an alternative variant form of autocell.

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 and the one with an alternative autocatalytic member—would be in competition with each other for resources. There would be differential replication rates, differential stability, and thus differential lineage propagation due to the difference between their two alternative ways of carrying out the corresponding catalytic function. Though it is a leap from autocell evolution to the evolvability of life, this simple model provides a plausible hypothesis for the minimal conditions for evolvability (Darwin 1859).

End-directedness and Evaluation. Are such processes as we see in autocell self-reconstitution and replication really anything like purposive activities? Most biologists tend to assume an eliminativist position and would therefore resist this assessment. Instead, biological function and goal-directedness are treated as purely mechanistic, only giving the appearance of purposive design. They are teleological in description only—that is, merely teleonomic (to use a term invented by Colin Pittendrigh [1958] to

describe the presumably nonteleological but teleology-like processes in organisms and other cybernetic mechanisms, such as thermostats).

In current biological theory, the natural-selection algorithm has become the explanation for all functional organization and good design in living organisms. From an eliminativist perspective, "good design" is merely "remembered accident" generated by the nonteleological algorithm of natural selection.

The autocell example does not contradict this abstract logic. An autocell could be described as a mechanism arising by accident. Still, although the natural-selection algorithm is itself nonteleological, the accidents it remembers are not arbitrarily selected. Selective retention is possible only for self-reconstituting forms. Such forms would not be exclusively thermodynamic, because thermodynamics cannot produce entities that will reconstruct themselves against the currents of thermodynamics. Nor can they be exclusively the product of the combination of thermodynamics and shape dynamics, because any self-organizing form such as an autocatalytic set or self-assembling capsule survives only as long as thermodynamic conditions are maintained and they lack the capacity to reconstitute after thermodynamic conditions fail.

An autocell is an accidental configuration that under certain conditions would be preserved precisely because its two component self-organizing processes fall into a relationship of *reciprocal cofacilitation*, a sort of meta-autocatalysis that is a function of their independent and mutual shape-biased self-organizational dynamics. Natural selection's memory of this accidental configuration is wholly dependent upon the two component self-organizing systems' mutual cofacilitation.

An autocell system is not in fact alive. Still, as simple as autocells are, they nonetheless possess some of life's essential features in primitive barely recognizable form. For example, they exhibit a minimalistic form of individuality and self. Autocells are robust to perturbation, as are all morphodynamic (self-organizing, dissipative) structures. Unlike other morphodynamic products, however, autocells also are robust to disrupted energy flow. In their closed form autocells persist without energy throughput. As such, autocells maintain a systemic individuality in the face of both material turnover (as in all morphodynamic structures) and intermittent energy flow. Thus, with autocells we have the emergence of a primitive form of "self." The ability to persist without energy throughput is a necessary though probably not a sufficient condition for defining selfhood, for the other attributes described below also contribute to the reasonable application of the concept of self to autocells.

Even more important for our purposes, one can identify a minimal form of value as well. In a lifeless universe there are no entities that have the individuational properties that would justify describing them as experiencing benefit or harm from occurrences in their environment. We contend

that the line at which value emerges is crossed even with these simple molecular systems. For example, it would be legitimate to describe the self-assembling container of an autocell as functioning for the maintenance and perpetuation of the autocatalytic cycle's, and the autocatalytic set could likewise be described as functioning for the maintenance and perpetuation of the self-assembly process. So it would not be a stretch to describe autocatalysis as an adaptation evolved for the "good" of the whole, and thus also itself, and that certain features of the molecular environment are "good" and others "bad" for autocells. This is a primitive form of value—the notions of good and bad *for* an entity.

Adaptation and Function. Autocells oscillate between two states—closed and open. Possessing more than one state with regard to some contextual feature is a prerequisite for evolvability. It provides behavior that can be modified through an evolutionary selective process. Variations between autocell lineages make such selection possible. They provide 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, incidentally 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 the past is to some extent a prologue to the present.

With autocells we therefore have a primitive form of evolution by adaptation: one that has no independent genetic code, one that is simpler than allows for the distinction between Lamarckian and Darwinian evolution, and yet one that nonetheless is 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. In this sense they embody in their form and dynamic potential—as if a photo negative—certain features of their environment that if present will be conducive to their persistence. But the presence of these conditions may or may not obtain. In other words, we may be justified in describing these autocell components as appropriate or inappropriate to their context, and in different contexts their functional organization will succeed or fail to be adaptive. In a crude sense, then, we can describe the organizational features of an autocell as a re-presentation—

with respect to autocell preservation—of these extrinsic conditions. And like other forms of representation this can be in error; there may be nothing in the immediate environment to which it corresponds. Thus, in this very basic sense, autocells could be said to represent their environment, in the same sense as a shoeprint could be said to represent a shoe, but without either DNA or RNA.

From Autocell to Life. We have developed a conceptual model for a process by which protoqualities of purpose might arise. If we can generate autocells with real molecules, we would have a proof of concept for the emergence of these essential protoqualities of purposive systems.

This should not, however, be presumed to be a full explanation for teleological processes experienced at the level of human consciousness, and not even as they are found in the simplest living forms on Earth. This "proof of principle" is in this regard quite minimalistic, and yet we believe it is still a definitive exemplar of this fundamental emergent transition that separates the mechanical world from the normative world.

Autocells lack many of the fundamental features we associate with life. They are exergonic, relying on the bonding energy intrinsic to catalysts and substrates rather than acquiring energy that can intrinsically drive the catalytic reactions. They have no replicator template molecules—no RNA or DNA—and their forms do not differentially survive through replications so much as self-reconstitute. Indeed, they are not even responsive to their environment. Although 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. They do not have a metabolism that continually maintains them in a far-from-equilibrium thermodynamic state, although once they generate the deviation from local equilibrium that results in autocell closure they have essentially captured in ratchetlike fashion this new deviation.

So we do not claim that autocells provide an account of the origin of life. Life as we know it is vastly more complex. Nevertheless, we believe that life was itself emergent from autocell beginnings, so one might describe this as a demonstration of the origins of what may better be described as protolife (that is, the most basic autonomous, end-directed system capable of self-reproduction and evolvability). Consequently, we see this work as the beginning stages of a new kind of research approach to the origins of life. The experimental enterprise of exploring the transition to life would be a subsequent, quite demanding, and extended enterprise involving many decades of future molecular research. But the autocell model, by starting vastly simpler than life, provides a useful platform for exploring stages that may have intervened between protolife units lacking most of the familiar components of living processes (such as genetic inheritance, cell membranes, and metabolism) and living organisms that rely on information transmission and incessant far-from-equilibrium dynamics.

The use of the autocell concept and eventually of exemplar molecular autocells as model systems promises a powerful new approach to questions hardly even conceived in contemporary biology. Specifically, we believe this can allow us and future researchers to investigate the origins of many of the most enigmatic features of Earth life that hitherto have been taken for granted and accepted as inevitable givens in the study of life. These include the informational character of the genetic code, the maintenance of constant far-from-equilibrium molecular dynamics via metabolism, and the critical role played by lipid-based cell membranes.

In particular, understanding the origin of the informational and semiotic aspects of life is critical for addressing the mystery of life's teleological properties, which are the precursors to the higher-level teleological processes of mind. What makes the autocell approach to these mysteries special and innovative is the possibility of exploring both theoretically and biochemically how these functions arose from a precursor protolife process (autocellularity) that lacked these specializations. To put this in basic terms, the autocell paradigm suggests a methodology for discovering how DNA became information.

Rather than beginning with the currently popular but (we contend) unconvincing assumption that remarkably complicated DNA or RNA molecules originally appeared de novo, already capable of somehow replicating autonomously (RNA molecules acting both as templates and as catalysts for forming replica templates), the autocell approach opens up the field to consider evolutionary mechanisms that could have led to this complicated class of molecules and informational relationships via a long se-

A. Autocell = **A**

P. Poly-autocatalytic (multiple linked alternative autocatalytic cycles) = A ⇨ **P**A

E. Energy capture-transfer cycle (dedicated autocatalytic cycle for energy capture-transfer) = PA ⇨ **E**PA

L. protein matrix + **L**ipid membrane (matrix enclosed by lipid bilayer, after transition to aqueous environment & polyamidine conversion to peptide) = EPA ⇨ **L**EPA

D. selectively **D**iffusing – continuous catalysis (protein matrix elements penetrating lipid membrane to create "pores" to allow selective diffusion) = LEPA ⇨ **D**LEPA

S. **S**elf-initiated division (internal control of container fission initiated via catalytic mechanism) = DLEPA ⇨ **S**DLEPA

N. polymerization of energetic molecules (nucleotides/sides) for storage and or phosphate inactivation = SDLEPA ⇨ **N**SDLEPA

T. coöption of nucleotide polymer to **T**emplate (RNA) = NSDLEPA ⇨ **T**NSDLEPA

I. **I**solation of duplicate non-reactive template (DNA) = TNSDLEPA ⇨ **I**TNSDLEPA

quence of evolutionary developments. The autocell approach suggests that the current complex molecular information functions are the result of an extended evolutionary ancestry and not the immensely improbable accidental starting point of life. Indeed, it is possible to sketch a rather detailed series of evolutionary transitions that could have led from the minimalistically simple autocell architecture to the complexity of a nucleic-acid-information-based form of life, such as now characterizes all Earth life. Such a tentative sequence is outlined in the table below.

Although the above scenario and this particular series of evolutionary stages are highly speculative and probably could be reordered in a number of ways and have many stages substituted in various ways by others, it must be noted that most contemporary approaches are restricted to assuming the de novo accidental achievement of this final stage in one miraculous leap from unspecified inorganic precursors. In comparison, at present and despite its highly speculative nature, there is no alternative paradigm that offers such a wealth of ways of conceiving of evolutionary mechanisms antecedent to and capable of constructing nucleic acid information functions. Simply formulating such a model system opens a vast range of research questions about the origins of life that had never before been considered. Being able to deconstruct the logic of life well below and before the complexity of existing forms cannot help but deepen our understanding of the very nature of life as it exists on Earth and as it may exist in myriad diverse forms elsewhere in the universe.

Implications

It is easy to imagine that this emergent worldview could be troubling for the approaches to life's ultimate questions offered by the world's religious and spiritual traditions. An account of the origins of life that can trace an unbroken logic from thermodynamics to the chemistry of catalysis to the simplest self-reproducing evolvable proto-organism addresses many concerns of those who find life's "irreducible complexity" to be an invitation to invoke divine intervention. However, this account does legitimate one element of this critique, which biologists often have disregarded as mere rhetoric: the indecomposable synergy that makes life fundamentally different from a mere agglomeration of chemical reactions inside a lipid membrane. The autocell example suggests that there are no crucial classes of molecules or sources of energy required for this property to arise, and even the classes of molecules that are now ubiquitous to life on Earth may be incidental side effects of our particular planetary chemistry. Autocellularity is effectively a functional property, not a chemical or energetic property. The cofacilitation between two morphodynamic (or self-organizing) systems could arise in functionally equivalent forms from a variety of substrates. This suggests that the same is likely to be true of life as well. In

simple terms, then, life is not just chemistry, but neither is it magic. It is the product of relationships and the forms they fall into.

Probably the most profound lesson to draw from this exploration of the logic underlying the nonlife-life transition is that it illuminates the source of the paradox of teleology—autonomy produced by codependence. The simplest possible material system that could exhibit end-directed behaviors is an interdependent reciprocity of self-organizing processes that collectively and synergistically constituted an autonomous, self-maintaining, self-reconstituting unit: a self that benefits its own persistence, but does so because of a larger compatibility with conditions intrinsic to its environment. Without this compatibility, none of these properties exists.

Indeed, it could be argued that the fundamental condition for life is not a template molecule like DNA, a work cycle, or a cell membrane but rather a relationship of codependence. The philosopher Immanuel Kant suggested as much in his *Critique of Teleological Judgment* ([1790] 1952), arguing that "an organized natural product is one in which every part is reciprocally both ends and means," because each is "reciprocally producing the others," such that "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" (pp. 557–58).

Perhaps of greatest philosophical significance, the autocell model provides a constructive proof-of-principle that the teleological phenomena of the world do not require antecedent *telos* to account for their existence. Teleological processes from function to representation to consciousness can be understood as emergent phenomena, capable of arising spontaneously from a universe devoid of any such property. In other words, the domain of meaning and purpose is not alien to the domain of physics and chemistry. There is no gulf of incompatibility separating them. The logical fabric of the universe is the willing midwife to the spontaneous birth of *telos*. So, as Stuart Kauffman (1996) announces in the title of his book, we should feel "at home in the universe," not alien to it.

So what does such a metaphysical turnaround mean for the relationship between science and the world's spiritual traditions? Spiritual traditions have for millennia intuited and argued from personal and collective incredulity that there is no way for mattering to emerge from mere matter and that therefore by default *telos* must be a special sort of substance conditionally injected into matter by and from an infinite outside source. Not only were we unable to describe how *telos* might emerge from an otherwise purely mechanistic universe, it seemed impossible that it could.

I believe we have shown that this can no longer be taken for granted. To be sure, an autocell is no more than a theoretical entity—an empirically inspired and constrained thought experiment that may or may not translate into chemical reality in a laboratory, or be discovered thriving on some other planet—but it makes no special assumptions and imagines no fanci-

ful type of chemical reaction. The power of this thought experiment is that it could be tested empirically, and without any yet-to-be-invented laboratory methods. And it is simple. Anyone can visualize an autocell as easily as one can visualize the behavior of a two-cycle lawn mower engine. Even if the specific chemical details of autocellularity as presented here were to prove infeasible, there can be little doubt that it is now possible to imagine at least one way that *telos* could emerge from non-*telos*. And ever since Darwin we have been able to imagine a feasible model for the transition from such simple beginnings to more complex forms.

Why is it possible to imagine what heretofore has been so elusive? It is not that people haven't tried to imagine matter emerging from matter before now. For millennia the question emphasized in the treatment of the origins of *telos* was an extreme one: How could a merely mechanical system ever give rise to consciousness? The autocell approach dissects the problem quite near to the base, and even there it does not suggest a simple mapping between chemical mechanism and end-directed chemical systems but rather suggests that this occurs in hierarchical stages, the central stage of which has been all but ignored until recently. Teleodynamic processes do not emerge fully formed from thermodynamic processes. The process is mediated by morphodynamics—spontaneous self-organizing, form-producing, dissipative processes. Morphodynamic processes supply dynamical structures that, when reciprocally and synergistically coupled, can become self-reconstituting, self-benefiting, adaptively responsive, and capable of evolving. Teleology is indeed not mappable in any direct way to mere mechanical processes, but it can emerge from them.

Clearly, once a culture resigns itself to the infeasibility of mattering emerging from matter, it finds virtue in the alternative and accumulates reasons to endow *telos*, ultimate purpose, moral value, and personal encouragement in some transcendental source of *Telos*. The alternative model we are proposing may be seen as an affront to such vested sensibilities, and, as with all such effrontery, the premium is on reasons not to subscribe to the alternatives rather than reasons to embrace them.

But science has provided many reasons to doubt the existence of ultimate transcendental *Telos*. The ratio of nonpurpose- to purpose-driven behavior in the universe's history is astronomically high. The amount of geological time before there is any evidence of purposive behavior in the known universe is likewise enormous. The fraction of time that purpose as we know and value it has been evident is minuscule. The ratio of purpose promoted by nature to purpose thwarted by nature is likewise tiny. The evidence points against there being some overarching *Telos* that moves nature ineluctably toward an end, let alone a happy one. Unfortunately, it has led many to extrapolate from these facts to the claim that all *telos* is illusory and that even the subjective experience of *telos* is a mirage. This is

not warranted. The improbable, tenuous, and delicate status of *telos* is precisely what one should expect if it is a high-level emergent phenomenon. The choice is not between transcendental *Telos* and the pure relativism of blind valueless clockwork. There is a vast emergent middle ground.

While the autocell model has not proven that *telos* emerges from certain reciprocal relationships between morphodynamic processes (like autocatalysis and self-assembly) or that this is the only possible way it could happen, it has shown that it is plausible and, indeed, empirically testable. Chemical experiments attempting to demonstrate the autocell mechanism will be required to prove its feasibility, but the logic of such an emergent transition is no longer inconceivable. And if it is imaginable, the burden of proof has already shifted. Conventional wisdom should feel put on the defensive. We thought that there was no way to even imagine how meaning, purpose, and value might emerge from mere matter, and now we realize that it is not so difficult to imagine after all.

This is not necessarily welcome news to many religious believers. A good amount of religious doctrine is dependent on the notion that this world is only a bleak alien stopover en route to a realm that is exclusively of the stuff that *telos* is made of, a world of mental existence independent of material existence in which an equally separable mental essence—the soul—can find a more congenial environment after separation from the organism machine. But even were it not for this personal motivation to maintain an insoluble dualism, the demonstration of the possibility of teleological phenomena spontaneously arising out of an otherwise blindly mechanistic universe has threatening implications. If we conceive of the world as inanimate stuff intrinsically incapable of representation, feeling, or purpose, it seems to require some outside creative influence to introduce the spark of teleology. Indeed, all teleological phenomena would have to trace their lineage to this essential origin. God the creator, in all the ways this concept is formulated, becomes a necessary external essence out of which (or within which) all end-directed phenomena flow and to which all owe their special otherworldly nature. If, however, *telos* not only is able to arise spontaneously but also is a complex emergent phenomenon that is dependent on specific classes of lower-order formative and energetic dynamics, the notion of disembodied antecedent *telos* becomes problematic. The essence of purposefulness is no longer an unanalyzable given but a property that can be analyzed and even possibly created "artificially" by some future science. And if it is a property that is necessarily defined with respect to certain less-than-purposeful dynamics, disembodied *telos* is nonsensical. Supernatural transcendent purpose may be not only unnecessary to explain life, mind, and value but an intrinsically incoherent concept.

These are potentially troubling implications for religious traditions that require teleology to be a fundamental unanalyzable property of the universe. There is another side to this emergent consequence, however. In a

mechanical universe where teleology is an intrinsically alien feature, we are aliens as well. Worse, in a universe in which all teleological phenomena are parasitic on a single divine source, all our experiences, desires, purposefulness, agency, and identity are parasitic as well. This is an extreme variant of the classic dilemma concerning free will and an omniscient all-powerful God, but in this case it is not just autonomous agency and moral responsibility that are at stake but human experience itself. If, however, teleological phenomena are emergent from the causal fabric of the universe, there is nothing illusory or impotent about the experience of subjective consciousness or agency. We are what we seem, a unique individual locus from which experience, meaning, end-directed activity, and value emerge spontaneously, as if out of nowhere.

Does a view of emergent *telos* and the abandonment of faith in preformed transcendental values imply moral relativism? This complex question cannot be adequately addressed in the closing paragraphs of this essay, but it is one of the more important issues that this metaphysical turn brings up for theology. We can only hint at the approach that this may suggest. To the extent that the emergence of *telos* depends on rare, precisely reciprocal, and contextually fitted relationships, it seems unlikely that emergent values should be arbitrarily flexible. Indeed, we may expect the "solution space" for the emergence of these most complex teleological relationships to be quite small.

Does an acceptance of this necessarily physical conception of *telos* force an abandonment of deistic claims? It does suggest that teleological properties are not essential to conception of God or notions of ultimacy. If *telos* can emerge from nonteleological beginnings, investing God with this property is redundant. If *telos* is necessarily defined with respect to a material base, transcendent immaterial *telos* would have to be something unrecognizably different. What would it mean to conceive of the ultimate without invoking ultimate teleology, or in fact any teleological notion? This may not be a challenging perspective for certain Buddhist and Taoist-inspired traditions, but the Abrahamic traditions are deeply wedded to a teleological conception of God, and one may worry that there is no reconciling these two perspectives. However, even within this tradition there have been views that are compatible with the denial of ultimate *Telos*. We might usefully compare this view with apophatic traditions that deny that any of God's traits can be described in ultimately idealized human terms. For example, Maimonides argued that conceptions of God based on idealizations of human traits must be rejected as infinitely far from the truth, leaving only negative attributes. The position suggested here is likewise the epitome of negative attribution. If open-ended *telos* is spontaneously emergent from physical processes, and therefore a physical property defined in physical terms, how can the same concept be the defining feature of a

nonphysical God? If all forms of teleological phenomena derive their character from a specific emergent dynamic, disembodied *telos* may ultimately be a contradiction in terms. Our point is not, however, to explore definitions of God but merely to reclaim *telos* for the natural, not supernatural, world. The world does not rely on a nonworldly origin for its purposes.

In contrast, open-ended *telos is* an essential feature of human experience. At minimum, we have provided a plausible explanation for the miracle of the *telos* that we have inherited from the dawn of life and that we now exemplify in one of its more elaborated forms. To show how it could have arisen from material origins does not render it less precious or less astonishing. It merely bursts the myth that the *telos* that we find in the world and that we experience in our lives is entirely dependent on some ultimate transcendental *Telos*. This alternative view may not provide the comfort of unquestioned certainty, but it definitely shows that we belong here, that we are the legitimate offspring of this world, and that our experience of self-creation is what it seems.

In conclusion, demonstrating that teleological phenomena are natural emergent features of physical processes and not dependent on some ultimate transcendental *Telos* threatens only comic-book versions of theology. Abandoning the notion of ineffable *ur-Telos*, which exists independent of and prior to its material embodiment, does not necessarily entail denying *Theos*; but it does force us to reconsider many theological assumptions. The ultimate questions won't go away with the abandonment of ultimate transcendental *Telos*, but they will change and undoubtedly become more difficult, more challenging, and more interesting.

NOTES

A version of this paper was originally delivered at the Star Island conference, "Emergence: Nature's Mode of Creativity," organized by the Institute on Religion in an Age of Science, 29 July–5 August 2006.

1. Actually, as his own writings divulge, Pasteur was also trying to discover the conditions that lead to the emergence of life from nonlife, and he was convinced that the single "handedness" or 3D twist of biological molecules was its secret. He was just a meticulously self-critical experimentalist, thus allowing him to easily spot the errors in others' reputed demonstrations.

REFERENCES

Darwin, Charles R. 1859. *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life*. London: John Murray.

Dawkins, Richard. 1976. *The Selfish Gene*. New York: Oxford Univ. Press.

Deacon, Terrence. 2006. "Reciprocal Linkage between Self-organizing Processes is Sufficient for Self-reproduction and Evolvability." *Biological Theory* 1(2): 136–49.

Deamer, D. W., and G. L. Barchfeld. 1982. "Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions." *Journal of Molecular Evolution* 18:203–6.

Eigen, M., and R. W. Oswatitsch. 1992. *Steps Towards Life: A Perspective on Evolution*. New York: Oxford Univ. Press.

Farmer, J. D., S. A. Kauffman, and N. H. Packard. 1986. "Autocatalytic replication of polymers." *Physica D* 22:50–67.

Hanczyc, M. M., S. M. Fujikawa, and J. W. Szostak. 2003. "Experimental models of primitive cellular components: Encapsulation, growth, and division." *Science* 302:618–22.
Hofstadter, Douglas. 2007. *I Am a Strange Loop*. Cambridge, Mass.: Basic Books.
Kant, Immanuel. [1790] 1952. *The Critique of Judgement: II. Teleological Judgement*. Trans. James Creed Meredith. The Great Books 42:550–613. Chicago: Univ. of Chicago Press.
Kauffman, Stuart A. 1986. "Autocatalytic sets of proteins." *Journal of Theoretical Biology* 119:1–24.
———. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford Univ. Press.
———. 1996. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. New York: Oxford Univ. Press.
Maynard Smith, J., and E. Szathmáry. 1999. *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford: Oxford Univ. Press.
Pittendrigh, Colin S. 1958. "Adaptation, Natural Selection, and Behavior." In *Behavior and Evolution*, ed. A. Roe and G. G. Simpson, 390–416. New Haven: Yale Univ. Press.
Prigogine, I., and I. Stengers. 1984. *Order out of Chaos*. New York: Bantam.
Rasmussen, S., L. Chen, D. Deamer, D. Krakauer, N. Packard, P. Stadler, and M. Bedau. 2004. "Evolution: Transitions from nonliving to living matter." *Science* 303:963–65.
Shapiro, R. 1988. "Prebiotic ribose synthesis: A critical analysis." *Origins of Life and Evolution of the Biosphere* 18:71–85.
Szostak, J., D. Bartel, and P. Luisi. 2001. "Synthesizing life." *Nature* 409:387–90.
Varela, Francisco J. 1979. *Principles of Biological Autonomy*. New York: North-Holland/Elsevier.
Varela, Francisco J., Humberto R. Maturana, and R. Uribe. 1974. "Autopoiesis: The organization of living systems, its characterization and a model." *Biosystems* 5:187–96.
Woese, C. R. 1967. *The Genetic Code*. New York: Harper and Row.
———. 1998. "The universal ancestor." *Proceedings of the National Academy of Science USA* 95:6854–59.