29

Autonomous agents

Stuart Kauffman
Santa Fe Institute

Writing in Dublin in 1944, Erwin Schrödinger sought the source of order in biological systems. Given the recent radiation mutagenic evidence on the target size of a gene showing that a gene had at most a few thousand atoms, Schrödinger argued that the familiar order due to square root N fluctuations around an equilibrium was insufficient because the fluctuations were too large to account for the hereditary order seen in biology. He argued that quantum mechanics, via stable chemical bonds, was essential for that order. Then he made his brilliant leap. Noting that a periodic crystal could not "say" very much, he opted for genes as aperiodic crystals which, via the aperiodicity, would carry a microcode specifying the ontogeny of the organism. It was a mere two decades until the structure of the aperiodic double helix of DNA and much of the genetic code were known.

But did Schrödinger's book, *What Is Life?* actually answer his core question? I think not, and the aim of this chapter is to propose a different definition, one concerning what I call an "autonomous agent," that may have stumbled on an adequate definition of life. I will not insist that I have succeeded, but at a minimum the definition leads in many useful and unexpected directions with import for physics, chemistry, biology, and beyond.

Our questions drive much of our science. The material in this chapter derives from my third book, *Investigations* (Kauffman 2000). In it, I am driven by an initial question: consider a bacterium swimming upstream in a glucose gradient. We would, and do, all say that the bacterium is going to get food. That is to say, the bacterium is acting on its own behalf in an environment. I will call a system that can act on its own behalf in an environment an "autonomous agent." I do no mean that it is alone in its environment, but that it can act on its own behalf.

Science and Ultimate Reality, eds. J. D. Barrow, P. C. W. Davies and C. L. Harper, Jr. Published by Cambridge University Press.

But the bacterium is "just" a physical system. So my question became: what must a physical system be to constitute an autonomous agent?

I will jump to the definition I found my way to after much consideration: an autonomous agent is a self-replicating system that is able to perform at least one thermodynamic work cycle. Importantly, all free-living entities fit this description. Thus, I may have stumbled upon an adequate definition of life, but, as noted, will not insist upon it.

It is a stunning fact that the biosphere is filled with autonomous agents that continuously reach out and manipulate the universe on their own behalf. Further, it is deeply interesting that with autonomous agents, the concept of "doings" as opposed to mere happenings, takes its place in our conceptual system.

Note two features of my definition. First, it is a definition. Definitions are neither true nor false, but can be useful or useless. I hope mine is useful. Second, the definition subtly leaps Hume's naturalistic fallacy. Hume argued that we cannot deduce "ought" from "is." But my definition jumps this gap definitionally. For once we admit that the autonomous agent is acting on its own behalf, we have a locus of value in a world of fact. Is this legitimate? I suspect that what is going on is, roughly, the following. Wittgenstein taught us about language games that were not reducible one to another. I suspect that my definition gives the minimum physical conditions for a physical system about which the language game of doing, acting, and value becomes natural.

You may cavil, but the language of doing, acting, and value is the language with which we talk about autonomous agents. Thus, since the bacterium is just a physical system, physics talk, which has only happenings, must broaden to include doings.

We have no theory of organization

I wish to make a central point, that we have no theory of organization, by starting with the familiar Carnot cycle, to which I make a single addition: a handle. Figure 29.1 shows the apparatus of the Carnot cycle, a hot reservoir, and cold reservoir, and a cylinder and piston located between the two. I show a handle extending from the cylinder to emphasize the need for organization of the processes that constitute the work cycle. State 1 of the work cycle has the piston high in the cylinder, the working gas compressed and as hot as the hot reservoir. Now I pull on the handle and bring the cylinder into contact with the hot reservoir, then allow the working gas to expand in the isothermal part of the power stroke. As the gas expands and cools, heat flows from the hot reservoir into the working gas and tends to maintain it at the hot temperature of the reservoir. Half way down the power stroke, I push on the handle, and move the cylinder to a position between the two reservoirs, touching neither. The gas continues to expand in the adiabatic part of the power stroke and,

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

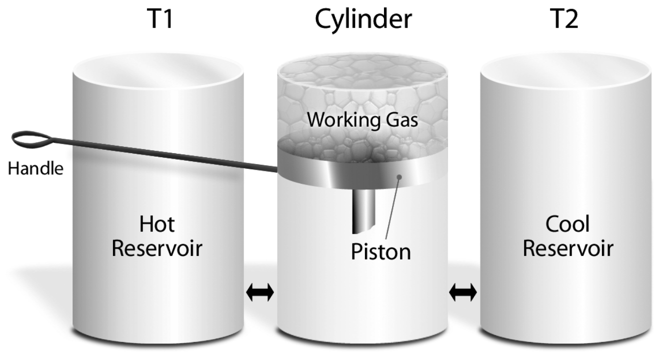

Figure 29.1. The Carnot cycle with a handle. See text for description.

because the cylinder is out of contact with the hot reservoir, the gas cools to the temperature of the cold reservoir at the end of the power stroke. Now I push the handle, pushing the cylinder into contact with the cold reservoir, then I push on the piston upwards, recompressing the working gas. As the gas compresses, it tends to heat up, but thanks to contact with the cold reservoir, the gas is held close to and somewhat above its cold temperature in the isothermal part of the compression stroke. Half way through the compression stroke, I pull on the handle, pulling the cylinder out of contact with the cold reservoir, then again push the piston up toward the cylinder head, thereby further compressing the working gas. Due to the compression, the working gas heats up to the temperature of the hot reservoir, completing the work cycle.

Two points require emphasis for our current purpose. First, the work cycle is a cyclic linkage of a spontaneous process, the power stroke, and a nonspontaneous process, my pushing on the piston to recompress the gas. Second, the handle and my pushing and pulling on it play a critical role, for they organize the flow of the linked processes. Now, in a real engine, gears, escapements, and chains, and so forth replace my pushing and pulling on the handle. Thus, the gears and so forth organize the processes of the work cycle. We understand, more or less, matter, energy, entropy, and information. But none of these is organization of process. As far as I can tell, we have no theory of such organization. Indeed, at present, it is unclear even what mathematical framework might be appropriate to a theory of organization. This will become all the more important in considering autonomous agents next.

A hypothetical autonomous agent

Our next task is to exhibit a conceptual example of a molecular autonomous agent. The example is not meant to be realistic in the detailed chemistry among the

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

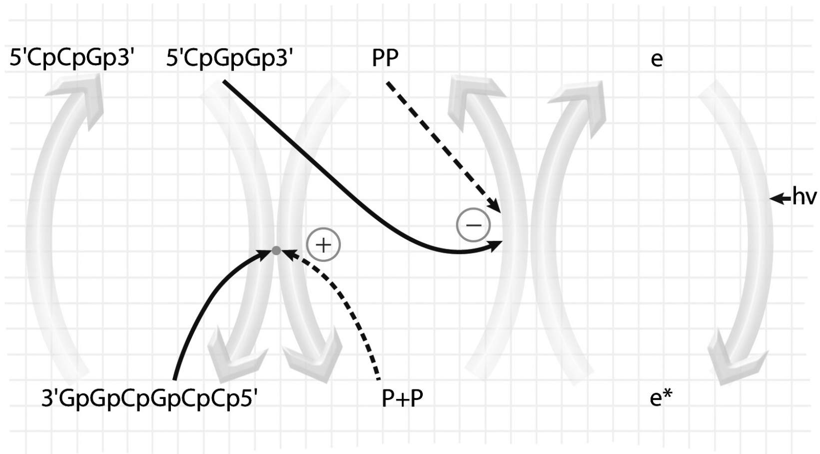

Figure 29.2. An hypothetical molecular autonomous agent. See text for description. For clarity, the reaction of the two DNA trimers to form the hexamer without coupling to pyrophosphate (PP) breakdown is not shown.

molecular components that it proposes. Rather it is meant to make concrete the concepts involved. Figure 29.2 shows a hypothetical molecular autonomous agent. It consists in a self-reproducing molecular system comprised of a single-stranded DNA hexamer and two trimers which are its Watson–Crick complements, coupled to a molecular motor that drives excess replication of the DNA hexamer. The entire system is a new class of open thermodynamic chemical reaction networks. This system takes in matter and energy in the forms of the two DNA trimers and a photon stream. As a preamble, I would note that self-reproducing molecular systems have been achieved experimentally, as have molecular motors. It remains to put the two together in a single system.

First, some familiar points should be made. The system I exhibit is a chemical reaction network some of whose components are enzymes. Enzymes do not change the equilibrium of a chemical reaction, they merely speed the approach to equilibrium. As will be familiar to most readers, a chemical reaction can approach equilibrium from either an initial excess of substrates or products. As a simple example, suppose A converts to B, and B converts to A. If one starts with pure A, A is converted to B. As B builds up, B is converted to A. Equilibrium is achieved when the net rate of conversion of A to B equals the net rate of conversion of B to A. Physical chemists draw this as an energy diagram, with the reaction coordinate on the *X*-axis and free energy on the *Y*-axis (Fig. 29.3). A spontaneous chemical reaction always proceeds "exergonically," in the direction of losing free energy until the minimum free energy is reached at equilibrium. If the reaction system is to be driven beyond equilibrium, say in order to synthesize more B that would be achievable by the undriven system, energy must be added to the system to drive it beyond equilibrium in an endergonic process.

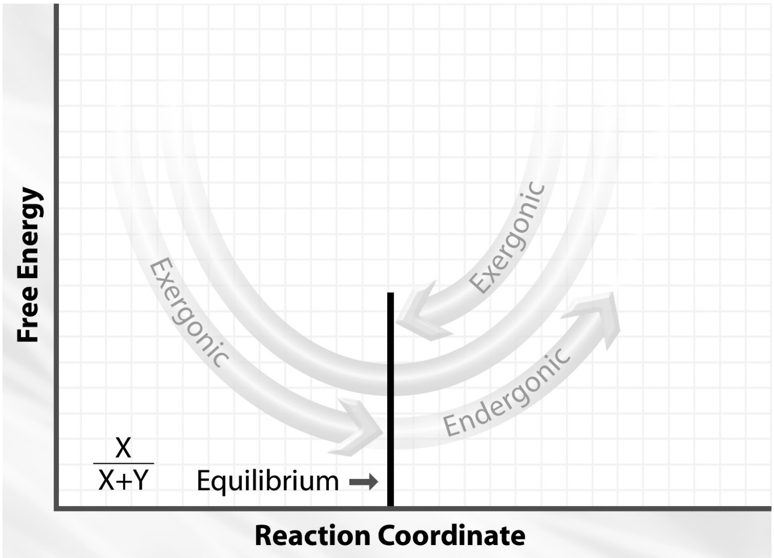

Figure 29.3. Exergonic and endergonic chemical reactions. See text for description.

Returning to the hypothetical autonomous agent, the DNA hexamer is capable of acting as an enzyme linking the two trimers into a hexamer which is a second copy of the initial hexamer, hence autocatalysis, or self-replication, is achieved. (This physical system, without coupling to pyrophosphate, PP, has actually been realized experimentally (von Kiedrowski 1986).) In the example, the synthesis of hexamer by linking the two trimers is coupled to the exergonic breakdown of PP to two monophosphates, P + P. This exergonic breakdown is utilized to drive endergonic excess synthesis of the hexamer compared to the equilibrium ratio of hexamer and trimers that would characterize the equilibrium of that subsystem. So, just as I pushed on the piston, the exergonic breakdown of PP is used to push excess replication of hexamer compared to equilibrium.

Once the pyrophosphate has been used to drive the endergonic synthesis of excess hexamer, it is necessary to restore the pyrophosphate concentration to its former level by driving the endergonic synthesis of PP from the two monophosphates. To supply the energy for this endergonic synthesis, I imagine an electron which absorbs a photon, is driven to an excited state, and, when it spontaneously and thus exergonically falls back to its ground state, uses that loss of free energy in a coupled reaction which drives the endergonic synthesis of pyrophosphate. I invoke one of the trimers as the catalyst that speeds up this coupled reaction.

Just as gears and escapements coordinate the flow of processes in the real Carnot engine, I invoke their analogs in the hypothetical autonomous agent. Specifically,

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

I want the forward reaction synthesizing excess hexamer to happen rapidly, then I want the reverse reaction resynthesizing pyrophosphate, PP, to occur. Accordingly, I assume that monophosphate, P, feeds back as an allosteric activator of the hexamer enzyme. Thus, the forward reaction proceeds slowly until the concentration of P rises sufficiently, then, thanks to the feedback activation of the hexamer enzyme, the forward reaction "flushes through," yielding excess synthesis of the new copies of the hexamer. Simultaneously, to stop the reaction resynthesizing PP from P + P, I invoke PP itself as an allosteric inhibitor of the enzyme catalyzing the resynthesis of PP from monophosphate. Thus, the forward reaction synthesizing hexamer occurs, then, after the concentration of PP falls, the resynthesis of PP occurs. My colleagues and I have written down the appropriate differential equations for this system and it behaves as described (Daley *et al*. 2002).

This system has a self-reproducing subsystem, the trimers and hexamer. And it has a chemical motor, the PP $\longleftrightarrow$ P + P reaction cycle. The engine's running can be seen by the fact that there is a net rotation of P counterclockwise around this reaction cycle.

Thus, this hypothetical system exhibits a molecular autonomous agent. Several points need to be made. First, this is a perfectly legitimate, if unstudied and new, class of open thermodynamic chemical reaction networks. Second, the system does not cheat the second law. The system eats trimers and photons, and, via the work cycle, pumps that energy into the excess synthesis of hexamer. Third, the system only works if displaced from chemical equilibrium in the "right" direction, an excess of trimers and photons. Agency only exists in systems displaced from equilibrium. Fourth, as pointed out to me by Phil Anderson, there is energy stored in the excess concentration of hexamer compared to its equilibrium concentration. That energy could later be used to correct errors, as happens in contemporary cells. Fifth, like the Carnot cycle, the autonomous agent contains a reciprocal and cyclic linking of spontaneous and nonspontaneous processes. Sixth, like the gears of a real Carnot engine, the allosteric couplings achieve the organization of the processes that is integral to the autonomous agent.

I would note that we are likely in the next decades to construct autonomous agents. Such systems actually do work and reproduce. They promise a technological revolution. More, if I have stumbled onto an adequate definition of life, they will constitute novel life forms. Sometime or another we will find or make novel life forms, and the way will be open to the creation of a general biology, freed from the constraints on our imagination of the only biology we know.

Considerations about the concept of work

I find work a puzzling concept. It is defined, of course, as force acting through a distance. But there are several unsettling aspects to the concept. Consider a concrete

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

case of work. I lift a pen, doing work on it lifting it in a gravitational potential. In any specific case of work, there is an organization to the process that is not captured in the definition of work. It is true that physicists invoke initial and boundary conditions in any specific case, but ultimately we want to consider the entire evolution of the universe, and it is precisely the coming into existence of those initial and boundary conditions that is in question in the evolving universe and biosphere.

Consider an isolated thermodynamic system, say gas in a large box. Now, an isolated thermodynamic system can perform no work. But if the box be divided by a membrane into two subcompartments, and the pressure is higher in one compartment, the membrane will push into the other compartment, doing work on it. Thus, it appears that for work to occur, the universe must be divided into two regions. Where did that come from?

The definition of work that I find most congenial is due to Atkins in his book on the Second Law. He points out that work is a "thing"; specifically, it is the constrained release of energy into relatively few degrees of freedom. So consider the cylinder and piston, with the working gas compressed into the head of the cylinder. As the gas expands, the chaotic thermal motion of the gas molecules is released into the translational motion of the piston.

But what are the constraints in the cylinder and piston system? Obviously, the cylinder and piston, and the location of the piston inside the cylinder are among the constraints. And now I ask a new question: where did the constraints come from? Well, it took work to build the cylinder and the piston, and work to place the piston inside the cylinder. So it appears to take work to make constraints and constraints to make work. I do not want to say that it always takes work to make constraints. One might start with a nonequilibrium system at high energy that falls to a lower energy state in which constraints are constructed. Nevertheless, it typically takes work to make constraints and constraints to make work. This is certainly true in real cells as I note below.

I said we have no theory of organization, but I have the deep suspicion that this reciprocal linking of work and constraints on the release of energy that constitutes work is part of that theory. If so, notice that this is not part of physics at present, nor of chemistry, nor of biology.

I want next to show that a cell can and does accomplish a kind of propagating work and constraint construction until the cell, astonishingly, builds a rough copy of itself. Figure 29.4 shows an example. A cell does thermodynamic work to synthesize lipid molecules from their building blocks. The lipid molecules then fall to a low energy state creating a bilipid membrane hollow sphere called a liposome. Inside the aqueous interior are two small organic molecular species, A and B. These can undergo three different reactions. A and B can convert to C and D; A and B can convert to E; A and B can convert to F and G. Each of these three reactions has its

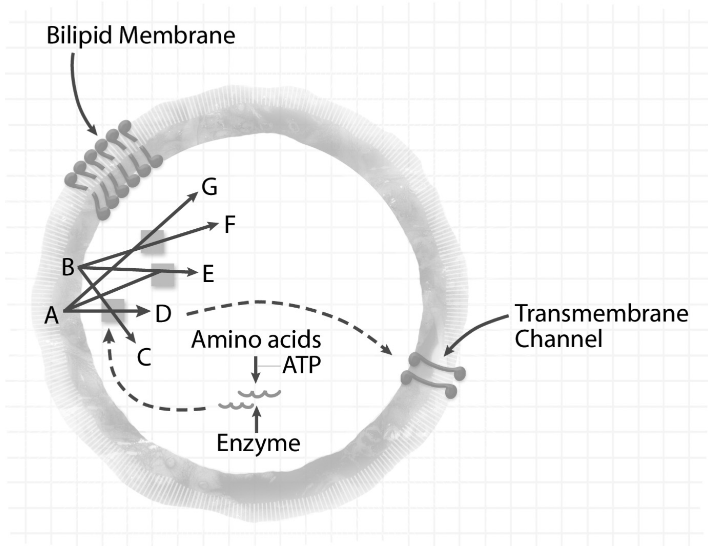

Figure 29.4. Propagating work and constraint construction in cells. See text for description.

own reaction coordinate and free energy diagram, with the substrates and products in potential wells separated by an energy barrier.

Now allow A and B to diffuse into the bilipid membrane. When that happens, the translational, vibrational, and rotational motions of A and B molecules alter. In turn, this modifies the reaction coordinate free energy profiles of the three reactions, perhaps raising some potential barriers and lowering others. But the raising and lowering of such barriers is precisely the manipulation of constraints. So cells do work to build constraints and manipulate them. Here the cell does work to construct a membrane which affords constraints and modifications of constraints. But more: the cell does thermodynamic work to link amino acids into an enzyme that happens to catalyze the conversion of A and B to C and D. Thus, free energy is released in constrained ways – the other two reactions are not catalyzed. But that released energy can propagate and do more work. For example, D may diffuse across the cell and bind to a transmembrane channel, giving up some of its vibrational energy to do work to open the channel to the ingress of an ion species. Hence "propagating" work has been done. In turn the ion species may cause further constraint construction and release of energy or other work. In short, a cell does in fact do work to construct constraints on the release of energy which, when released, does work to construct more constraints on the release of energy, which propagates until the cell completes a set of propagating work tasks and builds a copy of itself.

Note that we have, as yet, no developed language in physics, chemistry, or biology to discuss these matters. Consider also the miracle of the cell building a copy of

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

itself, then the two repeat the process to make four cells, then eight, then a bacterial colony. I can only stumble with ordinary English: the cell achieves a propagating organization of building, work, and constraint construction that completes itself by the formation of a second cell. Is this matter alone, energy alone, entropy alone, or information alone? No. Do we have a formulated concept for what I just described? No.

Yet just such propagating organization occurs. Kant told us the same thing long ago. We have no language, no mathematical framework that I know of, which captures this process. It appears to be a new state of matter – call it the living state.

Maxwell's demon flummoxed

I now make a detour to discuss Maxwell's demon in a nonequilililbrium setting. Good work shows that the demon cannot "win" in an equilibrium setting. It now appears that as the demon performs measurements on a system, for every one bit of entropy reduction there is a corresponding one bit increase in the most compressed description of the system that encodes what has been learned by the measurements. Since it ultimately costs energy to erase these bits, it does not pay the demon to measure in an equilibrium setting.

Now the way physicists always seem to phrase the issue about the demon is this: "And then, in principle, the demon can extract work."

I therefore want to consider the demon in a nonequilibrium setting and raise two important new questions. Consider the demon with the familiar box separated into two chambers with a flap valve between them. Let each chamber have the same number of particles, *N*. But let those in the right chamber be hotter, hence faster, than those in the left box. Thus, the system is nonequilibrium and work could be extracted. For example, let a small windmill be placed near the flap valve. Let the demon open the valve and a transient wind will flow from the right to the left box. The vane on the back of the windmill will "measure" the wind and orient the windmill perpendicular to the wind. So the system has detected and measured a source of energy, the wind. Then the wind will blow on the blades of the windmill and cause them to rotate. Thus the system detects a source of energy, responds to it "appropriately," and actually extracts energy to do work.

Now consider the demon confronting the hot and cold chambers, each with *N* particles in it. Let him perform an heroic experiment: he measures the instantaneous positions of all the *N* particles in both the right and left chambers. Note that, from this experiment, he cannot deduce that the particles in the right box are moving faster than those in the left box. Hence, he has carried out an ambitious measurement, but failed to detect a source of energy from which work can be extracted. Had he measured the positions of all the particles at two moments, he could have deduced that the right box was hotter, and would have detected a source of energy. Or he might have measured momentum transfer to the partition, or walls of the chambers and detected a source of energy.

How does the demon know what to measure?

The demon does not know what to measure. And that is the point. Next, how does the demon actually rig up a device to capture the energy and do work? Not so clear. And remember it takes work to rig the device. Does it pay in the sense that more work can be extracted than was used to construct and position the device? Now consider the biosphere. Let a mutation happen in some autonomous agent, say a protist, that allows it to detect, capture, and use a new source of energy to do some work of selective significance to the protist. Then natural selection will amplify this lucky mutation, and a new source of energy and work will become part of the biosphere. The combined system of autonomous agents, mutations, and selection does the job of picking out the useful measurements, detections of sources of energy, and getting work done that is useful. So the biosphere has and continues, presumably, to do just this. Think of the linked spontaneous and nonspontaneous chemical processes in a bit of biosphere on Darwin's tangled bank of an ecosystem. Sunlight falls, redwood trees grow.

I suspect that there is a deeper theory to be had here, but cannot prove it and, for reasons given shortly below, do not know how to construct it. The intuition, at least for a biosphere, and perhaps the universe, is that this process is part of the universe becoming complex and diverse. Consider the demon and the transient wind. Any flake of material, say mica, would flutter in the wind, hence extract work. Not much sophistication is required to detect the source of energy and extract work. Now consider an antiferromagnet not in its ground state. Thus, if the system could be triggered to fall towards its ground state, energy would be available to do work. But, intuitively, it takes a system as subtle in structure and behavior as the antiferromagnet to detect the displacement from equilibrium, trigger the fall toward equilibrium, and extract work. For example, a second antiferromagnet at the ground state might be in the vicinity of the nonequilibrium antiferromagnet, trigger the latter to fall toward the ground state, and use some of that energy to lift the ground state ferromagnet to a non-ground state. A source of free energy would have been detected and work would have been done.

Somehow, we need a theory of how such presumably increasingly subtle entities come into existence, couple spontaneous and nonspontaneous processes, and progressively build the complex structures of the universe and of a biosphere. I know of no such theory.

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

We cannot finitely prestate the future of the biosphere

I come now to the most troubling of the implications of autonomous agents. I suspect that we cannot prestate the future of the biosphere.

Consider the heart. Darwin would have told us that the function of the heart is to pump blood. That is, this is the causal consequence for which the heart was selected. But the heart also makes heart sounds. Heart sounds are not the function of the heart. Already this has important consequences. The function of a part of an organism is a subset of its causal consequences. To discover the function, we must study the whole organism in its environment. There is an unavoidable holism to biology.

Darwin also considered what he called preadaptations, and Stephen Jay Gould named exaptations. A part of an organism might have a causal consequence that is not of functional significance in the normal environment, but might prove useful in a different environment, hence be selected.

To make myself the butt of my example, I tell the following tale. The heart is a resonant chamber. Suppose I have a mutant heart, due to a single Mendelian dominant mutation, which renders my heart able to detect earthquake pre-tremors. I am in Los Angeles and feel something odd in my chest. "Uh oh," I think, "It must be an earthquake coming." I run into an open field. Millions die, but I am safe. Word leaks out that I detected the pre-tremor and I am invited onto "Good Morning America." I become famous. Women flock to my side. I mate with many. (This is necessary for my story.) Soon there are lots of little boy and girl progeny that have my mutant heart. If earthquakes happened often enough that this was of selective significance, soon the biosphere would sport earthquake detectors in humans.

Now my question is this: do you think that you can say, ahead of time, all the possible Darwinian preadaptations of, say, the 30 000 000 to 100 000 000 extant species? Can you say ahead of time all possible preadaptations for all the species that have lived? More formally, could you finitely prestate all the possible preadaptations of such species?

I think the answer is "No." Indeed, I have not found anyone who thinks the answer is "Yes." At least part of the problem is that I have no idea how to get started trying to state all possible environments for all members of all species. I want to say that we cannot finitely prestate the configuration space of the biosphere. Now you may tell me that, speaking classically, the configuration space is just some vast $6N$-dimensional phase space. Perhaps. But I will then respond by saying that you cannot prestate the relevant macroscopic collective variables of the biosphere that drive its further evolution, such as earthquake detecting hearts.

Next note that most major adaptations are Darwinian preadaptations. Most minor adaptations may also be preadaptations. Thus arose flight, hearing, lungs, etc.

Let us suppose the answer is, in fact "No." I want first to note that I do not know that the answer is really "No." Then I note that I do not understand whether this is a mathematical statement, perhaps akin to the halting problem or Gödel's theorem, or an empirical statement. Nor do I know if it is due to the finite computing power of the entire universe, should we be able to harness that power. It may be related to the fact that, for many algorithms, there is no shorter way to find out what they will do than to run the program. In any case, we seem to be precluded from knowing the future state of the biosphere, not due to quantum uncertainty, nor to chaotic dynamics, but because we do not have the concepts to pick out the relevant collective variables ahead of time.

Consider the frequency interpretation of probability. One begins by stating the space of possibilities. We cannot do this for the biosphere. The economists have a concept called Knightian uncertainty. Roughly it corresponds to the unknown that we do not yet know about. I suspect Knightian uncertainty is linked to what I am here describing. I am not a physicist, but my impression is that in Newtonian physics, quantum mechanics, statistical mechanics, and general relativity, one begins by prestating the relevant configuration space. If so, and if I am right, the biosphere seems to escape how we have been taught to do science.

But our incapacity to prestate the relevant collective variables is not slowing down the evolution of the biosphere. It therefore appears that we cannot prestate the future possibilities of the biosphere.

The Adjacent Possible

Finally, I want to touch on the concept of the Adjacent Possible. Consider a box filled with 1000 species of organic molecules. Call these the Actual. These molecules can undergo reactions creating, probably, molecules that are novel with respect to the Actual. Call the novel molecular species, reachable in a single reaction step from the Actual, the Adjacent Possible. The biosphere, 4.8 billion years ago, had only a few hundred organic molecular species. Now it has trillions. So the biosphere has been advancing into the Adjacent Possible over its history. The next point is that this advance is grossly nonergodic. Consider all possible proteins length 200. There are 20 raised to the 200th power, or 10 to the 260th power such sequences. We can make any single one we wish with today's technology. Now consider chemical reactions on a femtosecond timescale, and the estimated 10 to the 80th particle diversity of the known universe. Suppose the universe were busy building only proteins of length 200. Forget that distant particles cannot interact. The maximum number of pairwise interactions on a femtosecond timescale since the Big Bang is about 10 raised to the 193rd power – a vast number, but tiny compared to 10 raised to the 260th power. In short, it would take at least 10 raised to the 67th power repetitions of the history of the universe to make all possible proteins length 200. In short, we are on a unique trajectory, once one considers entities of complexity significantly greater than atomic species.

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press

Thus, we are, the universe is, advancing into an Adjacent Possible on some unique trajectory that is grossly nonrepeating, hence nonergodic. We are entitled to wonder whether there may be laws that govern how the expansion into the Adjacent Possible happens. No one knows, of course. But I want to close with a candidate law for biospheres: as a secular trend, (there are major extinction events that may interrupt this trend), it may be the case that a biosphere expands into the Adjacent Possible as fast as it can without destroying the order it has already assembled. Of course, this candidate law is still poorly stated, but the rough intuition is that biospheres anywhere in the cosmos, as a secular trend, maximize the diversity of what can happen next.

Conclusion

I have drawn attention to the issue of autonomous agents, systems that can act on their own behalf in an environment. All free-living organisms are autonomous agents. I have offered a definition of an autonomous agent as a system able to reproduce itself and carry out at least one work cycle. The issue is whether the defintion is useful or not. It appears to me that the concept leads us towards puzzles that we have not seen before, despite the fact that those puzzles are right in front of us. Central to this is a missing theory of organization. In discussing the flummoxed Maxwell's demon, I opined that there must be a theory for why the universe gets more complex. My difficulty in constructing such a theory is, at least, that I cannot see how to begin when I may not be able to finitely prestate what the entities are that will come into existence, detect sources of energy, link to them, and build yet new, unforeseen entities. Yet the biosphere, for 4.8 billion years, has been doing just this. But the biosphere is "just" a physical system. So, in the construction of a general biology, we must lift physics and chemistry, let alone biology, to some new level that can deal with biospheres anywhere in the cosmos.

References

Atkins, P W (1994) *The Second Law: Energy, Chaos, and Form*, revd edn. New York: W. H. Freeman.

Daley, A J, Grivin, A, Kauffman, S A, *et al*. (2002) Simulation of a chemical autonomous agent. *Zeits. Phys. Chem.* **216**, 41.

Darwin, C (1859) *On The Origin of Species*. Cambridge, MA: Harvard University Press. (Facsimile edn. London: John Murray).

Gould, S J (1977) *Ontogeny and Phylogeny*. Cambridge, MA: Harvard University Press.

Kauffman, S (2000) *Investigations*. New York: Oxford University Press.

von Kiedrowski,G (1986) A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed. Eng.* **25**, 982.

https://doi.org/10.1017/CBO9780511814990.032 Published online by Cambridge University Press