Biological and Nanomechanical Systems: Contrasts in Evolutionary Capacity

Note:

Despite its title, the paper that follows is best read, not as discussion of nanomechanical systems, but as an exploration of broad and fundamental questions about the contrasts between biological organisms and machine-like systems of all kinds. It describes and analyzes the consequences of a pattern of profound differences between the products of design and the products of evolution, a pattern that is directly linked to their enormous and fundamental difference in evolvability. The reason for these differences explains why members of vast class of machine-like systems could never evolve, whether or not some of those systems would have potential functional advantages relative to the products of biological evolution.

The basic argument is as follows:

• Evolvable systems must be able, with some regularity, to tolerate (and occasionally benefit from) significant, incremental uncoordinated structural changes. This is a stringent constraint because, in an evolutionary context, "tolerate" means that they must function — and remain competitive — after each such change.

• Biological systems must satisfy this condition, and how they do so has pervasive and sometimes surprising consequences for how they are organized and how they develop.

• Designed systems need not (and generally do not) satisfy this condition, and this permits them to change more freely (evolving in a non-biological sense), through design. In a design process, structural changes can be widespread and coordinated, and intermediate designs can be fertile as concepts, even if they do not work well as physical systems.

In reading the paper, please keep in mind the obsolescence (since 1992) of my initial, 1986 suggestion of using small self-replicating systems as a basis for high-throughput atomically precise manufacturing. There are better ways to do the job, and it is perhaps unsurprising that factory-style systems are superior.

Thinking about machines in the context of self-replication did, however, draw my attention to deeper questions about the organization of biological systems, and why they are have so little resemblance to the products of intelligent designers. This paper is the result, and nanotechnology is almost beside the point.

The conclusions of this paper are relevant to current concepts of advanced nanotechnology chiefly because they explain what otherwise might seem mysterious: that systems entirely unlike living cells can, by several engineering metrics, implement better ways to perform atomically precise fabrication.

The field of nanotechnology includes the study of certain classes of artificial molecular machines and self-replicating systems. Its concern with molecular replicators relates it to the study of living systems, both natural and artificial. Consideration of proposed nanomachines and replicators shows that they lack certain *basic* characteristics that are essential to the evolutionary capacity of living things. This paper examines these characteristics and their evolutionary significance.

The first section below provides an overview of nanotechnology, comparing and contrasting it with biological systems. The next examines several distinctions in styles of development and function: diffusive vs. channeled transport, matching vs. positional assembly, topological vs. geometric structure, and adaptive vs. inert building blocks. These distinctions are used to define two overall styles of organization, organic and mechanical. The succeeding sections relate these styles to the evolutionary capacity (or incapacity) of biological and nanomechanical systems and then summarize conclusions regarding evolution, replicating systems, and proposals for artificial life.

OVERVIEW OF NANOTECHNOLOGY

Nanotechnology is a projected technology based on a general ability to build objects to complex atomic specifications.[6] It takes its name from the nanometer scale of the structures it can produce; a cubic nanometer of material typically contains over a hundred atoms. Not all processes that make nanometer-scale products (which include simple molecules, ultrathin films, and submicron lines) are examples of nanotechnology, just as cigarettes and bubble pipes (making micron-scale smoke particles and soap films) are not tools of microtechnology. Nanotechnology implies atom-by-atom control of complex structures; microtechnology implies the fabrication of complex, microscopic structures without this control. (Nanotechnology will not be limited to small structures, however.[8])

The molecular machinery of life demonstrates functions that will he important in nanotechnology. Some enzymes assemble small reactive molecules to build larger molecules. Ribosomes are genetically programmed machine tools that assemble small reactive molecules in complex patterns to form large molecular machines.

Nanotechnology will be based on programmable machine tools with more general abilities — devices termed *assemblers* — which will enable the construction of a wide range of molecular structures. Ribosomes can build machines only of protein, but a molecular-scale robot arm, able to work with a wide range of reactive molecules, should be able to build molecular machines with almost any chemically reasonable structure.[6,7] Synthetic organic chemists make a wide range of molecular structures by mixing reactive molecules in solution. Characteristically, they cannot make very complex structures (say, a billion-atom molecule with the complexity of an integrated circuit), owing to the difficulty of controlling the site of a reaction on the surface of a large

molecule. Diffusion bumps molecules together in all positions and orientations; reactions occur wherever they are chemically feasible. Assemblers will sidestep this limit by eliminating diffusion: they will position reactive molecules mechanically, making reactions occur only at the sites selected by the designer.

A well-established nanotechnology will likely make little use of biomolecules. The molecular machines envisioned for that era are surprisingly conventional, including gears and bearings,[10] electric motors (electrostatic, rather than electromagnetic[7]), and a full range of moving parts. Analysis indicates that digital logic systems based on molecular mechanical devices can be compact (fitting the capacity of a mainframe computer into a cubic micron) and can be reliable despite thermal noise.[9,11]

To visualize mechanical devices on this scale, it is important to recognize that molecules are objects, with size, shape, mass, strength, stiffness, and so forth. Large machines are made of parts with many atoms; nanomachines will be made of parts with few. Just as engineers prefer to work with light, rigid materials on a large scale, so nanoengineers will prefer such materials on a small scale. Thus, parts will typically contain patterns of atoms like those found in engineering plastics, ceramics, graphite, and diamond.

PATHS TO NANOTECHNOLOGY

The core idea of nanotechnology is to use molecular machines (assemblers) to build assemblers and other products. This appears reasonable, but circular. How might assemblers be built in the first place? Just as there was no principle preventing crude machine tools from building better machine tools during the development of macrotechnology, so there is no principle preventing crude molecular machines from building better molecular machines during the development of nanotechnology. Several paths lead toward this sort of spiraling advance of technology.

First-generation assemblers may be developed through protein engineering; biochemical analogies indicate that protein engineering (when sufficiently advanced) will enable the design and fabrication of complex, self-assembling molecular machines.[6,23,24,25,27] Likewise, first-generation assemblers may be developed through the synthesis of self-assembling sets of non-protein molecules.[14,16,17] Alternatively, advances in micromanipulation may enable the construction of first-generation assemblers through mechanically directed molecular assembly; reports of atomic rearrangement through field-induced evaporation from scanning tunneling microscope (STM) tips[1] and of highly localized chemical reactions induced by currents at an STM tip[13] are suggestive in this regard. In practice, development may well involve a combination of chemical, biochemical, and micromechanical techniques. However assemblers may first be built, later assemblers will be built using assemblers. The nature of nanotechnology and its capabilities will thenThe independent of the nature of proteins, conventional chemistry, and initial micromanipulation technologies.

Since multiple paths lead to molecular machines and nanotechnology, no one problem with development can block advance in this direction . With multiple paths, multiple research groups, and multiple chains of short-term rewards along each path, it is (in a

competitive world) hard to imagine that nanotechnology will not eventually be realized. This adds to its interest as an object of study.

NANOREPLICATORS AND BIOREPLICATORS

If assemblers, guided by nanocomputers, can build almost anything, then with proper programming they should be able to build copies of themselves (and of the nanocomputer, and its instructions and so forth). If assemblers are to process large quantities of material atom-by-atom, many will be needed; this makes pursuit of self-replicating systems a natural goal. The availability of atoms as prefabricated building blocks simplifies self-replication on this scale.

The following will compare and contrast living systems with systems (especially replicators) based on anticipated styles of nanomachinery. For present purposes, "living systems" are defined as systems based on cells, ranging from bacteria to blue whales (many observations will apply to viruses as well). For convenience, self-replicating systems of nanomachinery will here here termed nanoreplicators; living systems will occasionally be termed bioreplicators. (Note that this use of the term "replicator" is distinct from Dawkins' use,[2] in that it refers to the whole replicating system, genotype and phenotype, rather than to just its genetic material. In this context, a replicator in Dawkins' sense can be termed a "genetic replicator"; the distinction is vital, since only *genetic* replicators pass on mutations and evolve.)

The parallels between existing bioreplicators and proposed nanoreplicators are strong. Both rely on the use of molecular machines to position reactive molecules, thus directing the synthesis of complex systems, including more molecular machines. Assemblers are analogous to ribosomes; the systems that supply them with reactive molecules are analogous to metabolic enzyme systems. Both bioreplicators and nanoreplicators rely on digital control systems: the genetic system directs ribosomes; nanocomputers are expected to direct assemblers.[7,11] In a broad sense, each may be viewed as an instantiation of von Neumann's architecture for self-replicating systems.

Despite these parallels, the differences between existing bioreplicators and proposed nanoreplicators are great. Ribosomes get their parts, energy, and directions via diffusion, but assemblers in proposed nanoreplicators will get them via fixed channels. Ribosomes self-assemble via diffusion and matching of complementary parts, but assemblers will be made by operations analogous to manual construction. Cells and organisms have structures defined chiefly by patterns of containment and interconnection, but nanoreplicators will have structures defined by a specific geometry. Organisms grow, with their parts adapting to one another, but nanoreplicators will be constructed from parts of fixed structure. In summary, where bioreplicators have an "organic" style, proposed nanoreplicators will have a "mechanical" style, resembling factories more than they do living cells and organisms. The following sections will explore these differences in more detail, then argue that life has this "organic" style, not for reasons of technical efficiency, but because alternative "mechanical" systems could not arise through conventional evolution.

STYLES OF DEVELOPMENT, STRUCTURE, AND FUNCTION

Systems can differ in their means of transporting materials, energy, and information; in their means of assembling parts; in the definition of their structures; and in the adaptability of their parts. These differences distinguish different styles of development, structure, and function.

DIFFUSIVE VS. CHANNELED TRANSPORT

Bioreplicators make heavy use of diffusive transport for materials, energy and information. In living cells, metabolic substrates diffuse from enzyme to enzyme, as do energy-transmitting molecules, such as ATP. Small, diffusing molecules, such as cyclic AMP, serve as signals; diffusing RNA molecules carry whole blocks of organized, digital information.

Like factories, proposed nanoreplicators make heavy use of channeled transport systems. Examples of these include conveyor belts and pipes for moving materials, wires and drive shafts for moving energy, and cables for moving information. Compared to diffusive transport systems, channeled systems commonly have technical advantages in compactness, speed of transportation, and minimization of inventory.

Materials handling is important in manufacturing systems, including replica-tots. Typically, a part will go through several manufacturing operations, each per formed by a distinct machine. This pattern is familiar both in factories and in cell metabolism (where the parts are molecules and the machines are enzymes); it is to be expected in nanoreplicators as well. In a diffusive system, every machine is effectively linked to every other — it can accept inputs from anywhere, and its outputs are available everywhere. In a prototypical channeled system, in contrast, every machine must be specifically linked (by conveyor belts or the equivalent) to its input-suppliers and output-consumers. Thus, in a channeled system, a new machine can do useful work only if aided by corresponding additions to the transportation system; in a diffusive system, a new machine can do useful work without such additions. As will be seen, this difference is of basic evolutionary importance.

The general pattern of diffusive transport in living cells has limitations and exceptions. Eukaryotic cells contain numerous membrane compartments, placing regional controls on diffusion; active molecules pump some materials across membranes against concentration gradients. These modifications do not suffice to make the transportation system channeled, however, It has recently been argued[26] that some enzymes seldom release their products to diffuse freely, but instead transfer them directly to the active site of the next enzyme in the metabolic pathway (on encountering that enzyme through a diffusive process). Since these enzymes *can* transfer their products diffusively, however, they are not subject to the limits of a truly channeled system. More significant is the presence of systems such as the fatty-acid synthetase complex, which holds a partially completed molecule on the end of a swinging arm, cycling it through different active sites on the complex to add a series of two-carbon units, building up a fatty-acid chain.[18] This system is effectively channeled; it is significant that systems of this sort

appear rare in cells, despite their technical advantages in materials transport. These channeled islands are linked by a diffusive sea.

Diffusive systems have an advantage in reliability over simple one-path channeled systems. In a diffusive system, no vital channel can fail or be blocked by the failure of a processing-machine, since no such channel exists. A more complex channeled system can gain comparable reliability, however, by incorporating redundant paths and machines connected in a suitable network.

MATCHING VS. POSITIONAL ASSEMBLY

Bioreplicators make heavy use of spontaneous assembly based on diffusion amid matching. Molecular parts (such as the RNA and protein molecules that make up ribosomes) diffuse and bump together in all possible positions and orientations. Those that have corresponding surfaces (matching patterns of bumps and hollows, positive and negative charge, hydrophobicity, etc.) pull together and stick, forming a specific structure.

Automated factories and proposed nanoreplicators, in contrast, make heavy use of positional assembly. Here, the prototype is a blind robot thrusting a pin into the expected location of a hole. There is no finding and matching of parts — if the hole is elsewhere, the operation fails. Positional assembly has potential advantages in speed, and in the lesser constraints it places on the structure of device interfaces (no need to induce self-assembly, and no need to guide it by providing unique interfaces for differing parts).

In a system made by a matching assembly process, an increase in the number of matching parts A and B leads naturally to an increase in the number of assemblies AB. In a positional assembly process, in contrast, new parts A must be placed in new positions to which parts B must be brought; new assemblies AD thus require corresponding changes in the assembly process. This is directly analogous to the requirement, in a channeled transport system, for new channels corresponding to new machines.

In a matching assembly process, a change in the size or shape of a part, if it does not disturb its interface to another part, will seldom disturb assembly. In a positional assembly process, however, a change in position constitutes a significant disturbance to the interface: for example, the insertion position of a screw on top of a carburetor will change if the carburetor grows taller, and the screw will miss the hole. Worse, any change in the height of any part on which the carburetor is mounted will cause the same problem. If the screw were to diffuse to the hole, then match and stick, such problems would not arise.

TOPOLOGICAL VS. GEOMETRIC STRUCTURES

Geometric structures characterize conventional machines and proposed nanoreplicators. Parts have definite sizes, shapes, amid positions with respect to one another. The resulting fixed geometry lends itself to positional assembly amid channeled transport systems.

The structures of cells amid living organisms, however, are largely organized in a way that can be described as *topological* characterized not so much by specific positions as by patterns of connectivity. The shape of a membrane compartment in a cell matters less than its continuity and the contents of the volume it defines. Likewise, the length of a muscle matters less than its attachment points. Diffusive transport and matching assembly, with their lack of position dependence, lend themselves to use in the assembly and functioning of topological structures.

Adding a part inside a densely organized geometric structure typically requires changes in the relative positions of many other parts, and hence corresponding adjustments in design. Adding a part inside a densely organized topological structure, in contrast, typically leaves topologies unchanged — room can be made by stretching and shifting other parts, with no change in their essential design.

ADAPTIVE VS. INERT PARTS

Closely related to the notion of topological and geometric structures is the notion of adaptive and inert parts. The prototype of an inert part is a rigid object with a flange having a special shape and a special pattern of bolt holes — it fits a corresponding part, or it doesn't; a change in the interface of one demands a compensating change in the interface of the other. The prototype of an adaptive part is a coat of spray paint — it fits the part it coats, with no delicate dependence on that part's size or shape. Rubber hoses are relatively adaptive; rubber gaskets are less so. Typical metal and ceramic parts, like the rigid parts of proposed nanomachines, are essentially inert.

Bioreplicators make extensive use of adaptive parts. Skin grows to cover an organism; it need not be redesigned when genes or environment give rise to a giant. Likewise, skulls grow to cover brains, muscles grow to match bone-lengths, and vascular systems grow to permeate tissues. The inherent adaptiveness of tissues amid organs is demonstrated by healing in adults, and by the development of strangely connected but locally plausible organ systems in Siamese twins.

SUMMARY: O-STYLE VS. M-STYLE SYSTEMS

Living things are characterized by heavy use of diffusive transport, matching assembly, topological structures, and adaptive parts. Let us call systems that share these characteristics, whether living or not, *O-style systems* (O is mnemonic for organic).

Mechanical systems are characterized by heavy use of channeled transport, positional assembly, geometric structures, and inert parts. Let us call systems that share these characteristics *M-style systems* (M is mnemonic for mechanical).

The difference between O-style and M-style is not a hard distinction, but a matter of degree. The following will often speak of them as if they were distinct, but they form, at least in principle, a continuum. Molecular machines inside cells typically have M-style features; their parts are relatively geometric and inert. Automobiles contain hoses and coats of paint with a measure of O-style adaptiveness. Still, out the whole, cells (with their diffusive transport, matching assembly, topological structures, and adaptive parts)

are strongly O-style while automobiles (with their channeled transport lack of assembly operations, geometric structures, and inert parts) are strongly M-style. By this measure proposed nanoreplicators are far closer to cars than to cells and other living systems.

O-STYLE, M-STYLE, AND EVOLUTION

Each of the characteristics distinguishing O-style from M-style is of considerable importance to evolutionary capacity. In each case, the M-style characteristic introduces dependencies among parts such that typical changes in the structure of one are of no benefit (or do harm) without simultaneous, corresponding changes in the structure of others.

In a conventional evolutionary system, the genetic system does not somehow convert single mutations into properly corresponding changes in multiple parts. Further, selection pressures are applied after each mutation, with no favor extended to promising-but-harmful mutations while they await redemption in the form of a corresponding mutation elsewhere. In these circumstances, the characteristics of M-style systems effectively destroy their evolutionary capacity; the characteristics of O-style systems sustain it.

M-STYLE SYSTEMS AND EVOLUTION

Consider an integrated, strongly M-style box (perhaps a subsystem of a manufacturing system). It is built on a robotic assembly line and consists of a large number of rigid parts, some movable, mounted in a chassis. It includes motors, drive systems, and transport paths for workpieces and products. For the sake of familiarity, imagine that this box is a macromechanism made by machining and assembling metal parts, rather than a nanomechanism made by assembling reactive molecules; the issues raised are the same. For concreteness, imagine a system like the mechanism of a xerographic copier.

If this system is to serve as an example of (attempted) M-style evolution, we need some concept of a genetic system amid associated embryology. For an M-style system, engineering practice suggests the following assumption: digital programs form the genetic system. They control machines that shape amid assemble parts to make the box; this process constitutes the embryology. To make repeated structures, these digital programs might make repeated use of some segments of code. M-style positional assembly implies that these shaping and assembly operations involve moving tools to a series of specific three-dimensional coordinates (with respect to a local "workbench," say). Some mutation operations to programs will add or delete material to parts (changing their size and shape); others will change the coordinates at which a part is placed during assembly.

Certain trivial evolutionary changes are quite feasible in this system. Sections of individual parts that do not interact directly with other parts (or with assembly tools) can change shape with only local effects. A part might become thicker amid stronger, and

hence more reliable, or it might become thinner and lighter, and hence less costly. Plausible embryologies and selective pressures could lead to considerable optimization of part shapes.

Other trivial evolutionary changes run into difficulties. Some sections of individual parts form interfaces to neighboring parts. If a flange has a particular size and shape, its neighbor must correspond. A substantial "favorable" mutation in one (say, toward a larger, more robust configuration) would disrupt the interface in the absence of a (highly unlikely) simultaneous, corresponding mutation in the other. Only creeping changes in dimensions would be feasible, keeping each part's change within the other's tolerance at each step.

The trivial change of, say, lengthening a bracket similarly tends to disrupt positional assembly. Changing the size of a part on which other parts are mounted changes the position of the mounting-points. A substantial mutation of this sort, if it is to yield a functioning system, must be matched by a simultaneous mutation in the assembly coordinates of all the affected parts — and every added requirement for simultaneous mutation vastly lengthens the odds. This requirement is a direct consequence of positional assembly (and our choice of a simple genetic system and embryology). Again, only creeping change is possible — this time with each change falling within the tolerance of a potentially large number of other parts and assembly steps.

Non-trivial changes add parts or change system organization. Examples include inserting a gasket and connecting a tank to a pipe. The former forces a discrete change in the separation of two surfaces, precluding creeping change. The latter raises the specter of a useless section of pipe, running up to a tank with no opening, or (worse) a tank with a hole and no attached pipe. With positional assembly of non-adaptive parts, there seems no escape from the need for multiple, simultaneous, coordinated changes in such cases. In a viable system, however, mutations at any one site will be extremely rare; a simultaneous, matching mutation at another specific site will be astronomically rare. Several such mutations become effectively impossible.

A genuinely significant evolutionary change, for many purposes, would be one which lets our hypothetical box make a new product. This will, in general, require the addition of many parts, forming a new processing subsystem. In addition, parts will be required to form the channels linking the subsystem with sources of power, with preceding and following processing subsystems, amid so forth. Finally, given geometric structures, simply opening enough room for the new subsystem will require wide-spread restructuring of other parts and systems. In short, even small changes rapidly approach impossibility, amid the changes required to acquire new capabilities would be large.

It is easy to get some rough idea of the probabilities involved. In modern digital systems (which can't incorporate error-correcting codes), an error rate of one bit in a billion is commonly considered high; error rates in fact can be made arbitrarily low through redundancy.[21] DNA replication (with error-correcting enzymes) can achieve bit-error rates as low as one in one hundred billion.[5] In an M-style system (macro- or nano-) designed for reliability, transmission of genetic information should be at least this accurate.

Attaching a new pipe to a tank requires several coordinated changes: making a hole, making a fitting, attaching the fitting, making a pipe, and attaching the pipe. If each of these five changes took as few as eight bits to specify, then 40 changed bits would be needed. Given a 10^{-9} probability of changing a single, specific bit in a generation, the probability of independently changing sixteen specific bits is 10^{-360}. If every hydrogen atom in the observable universe were a genome and had undergone one generation every nanosecond for 10 billion years, the probability of having seen this 40-bit change anywhere, at any time, would be less than one in 10^{270}. A simultaneous, coordinated change of this sort is effectively impossible.

This resembles bogus arguments raised against the feasibility of biological evolution. How do living things escape its application?

O-STYLE SYSTEMS AND EVOLUTION

Bioreplicators have patterns of development, structure, and function that enable evolution to proceed without coordinated, simultaneous genetic changes. Each O-style characteristic contributes to this result.

Because cells and organisms make widespread use of diffusive transport for energy, information, and molecular parts, the evolution of new processing entities (enzymes, glands) is facilitated. A genetic change that introduces an enzyme with a new function can have immediate favorable effects because diffusion automatically links the enzyme to all other enzymes, energy sources, amid signal molecules in the same membrane compartment of the cell (and often beyond). No new channels need be built at the same time, because transport isn't channeled. What is more, no special space need be set aside for the enzyme, because device placement isn't geometric.

Changes in the number of parts — so difficult in a rigid M-style system — become easy. There are no strong geometric or transport constraints. This often allows the number of molecular parts in a cell to be a variable, statistical quantity. With many copies of a part, a mutation that changes the instructions for some copies is less likely to be fatal. Thus, diffusive transport facilitates quantitative redundancy, which facilitates qualitative evolutionary experimentation.

A matching assembly process (as in the formation of ribosomes, microtubules, and so forth) tolerates variations in system geometry and numbers of parts that would disrupt positional assembly. Further, the mechanical compliance of biomolecules, such as proteins, gives them a bit of adaptability, allowing small changes in the interface of one molecule toThe tolerated by the matching process, giving time for a corresponding change in the facing molecule to occur.

At the level of multicellular organisms, the striking adaptability of tissues and organs ensures that basic requirements for viability, such as continuity of skin, amid vascularization of tissues, continue to be met despite changes in size and structure. If skin and vascular systems were inert parts, they would require compensating adjustments for such changes. We have seen the problems involved in a minor change in

M-style plumbing, yet every individual has a different detailed vascular topology, without corresponding genetic gymnastics.

Thus, O-style systems are not described by calculations like the one above because they can undergo significant evolution *without requiring multiple, simultaneous, coordinated changes*. If one were to perform a similar calculation, allowing the 40 one-bit changes to occur separately, accumulating across generations, then the waiting time for the desired combination would fall from vastly longer than the age of the universe to a fraction of a second.

In short, the characteristics of O-style systems enable cumulative selection to operate; the nature and power of this mechanism have been well described by Richard Dawkins.[3] This mechanism, as Dawkins notes, does not enable all imaginable evolutionary steps, but only some. Among the prohibited steps are those that require multiple, simultaneous, coordinated changes. For example, vertebrate retinas have their neural wiring in front of their photosensors, reducing optical quality and necessitating a blind spot where the optic nerve passes through the sensor layer. Cephalopod retinas have the sensible structure, with the wiring behind. Why hasn't evolution flipped the vertebrate retina? Presumably because there is no small genetic change that would do the whole job, rather than just some damaging part of it; success would require multiple, simultaneous, coordinated changes. Likewise, all living things share essentially the same genetic code for translation between DNA sequences and amino acid sequences in proteins. Why hasn't the code changed in recent evolutionary time, perhaps to add a new amino acid? Presumably because to change the translation mechanism would (among other things) require the simultaneous recoding of the structures of many vital proteins — again, an evolutionarily prohibited step.

Today's O-style biological systems owe their existence to their ancestors' evolutionary flexibility. Since they have inherited that flexibility, they retain the capacity for further evolution; they can be said to have *evolved for evolvability*.[4] M-style systems with their radically different patterns of development, structure and function have not done so and hence lack O-style flexibility. As we have seen, even O-style systems suffer from substantial constraints on their available evolutionary moves; it seems that M-style systems suffer from constraints that effectively eliminate significant evolutionary moves.

M-STYLE SYSTEMS THROUGH O-STYLE DESIGN

If M-style systems cannot evolve, how can they exist? The answer lies in the relationship between design and evolution.

The above argument for the evolutionary incapacity of M-style systems depended on a certain kind of genetic system and embryology operating in a certain kind of evolutionary environment. It assumed that mutations produced isolated changes in the shapes and positions of parts (however broad the consequences of those changes might be), and that selective pressures went to work — in particular, that unworkable designs would fail and be lost, not kept and tinkered with. These assumptions regarding genetics are appropriate for an automatic manufacturing system patterned on present engineering

practice, with computer programs playing the role of the genome. They are likewise appropriate for a similarly programmed, self-replicating manufacturing system, like a nanoreplicator. The assumptions regarding selective pressures are appropriate for a system in a situation analogous to the natural environment, as opposed to a development laboratory.

Design is an evolutionary process that operates on different genetic replicators. It works, not by mutating computer programs in a factory, but by mutating ideas in the mind of a designer (or, eventually, high-level representations in an AI design system). If introspection is any guide, ideas are not limited to channeled transport and positional assembly within the mind. Rather, they "diffuse," encountering each other in various patterns and combinations. Some "match," and stick, forming larger systems. These systems seem more topological than geometric, in that their patterns of connectivity are important, and they seldom seem to have anything analogous to a detailed position or alignment that can be globally disturbed by introducing a new piece in the structure. Finally, ideas are typically adaptive, taking a form that depends on their relationships to other ideas, Design concepts — particularly in their formative stages — have the sort of O-style flexibility that implemented machines typically lack.

(Note, incidentally, that the development of software from machine languages to modern AI languages has moved away from M-style toward O-style characteristics, The former are strongly "channeled" amid "positional," but, rule-based AI systems use mechanisms analogous to diffusion and matching — this enables the introduction of rules without wiring them into an elaborate amid rigid control structure. A program based on these principles — EURISKO[20] — is one of the better examples of self-evolving software; Holland's classifier systems,[15] based on genetic algorithms, have these same abstract properties, as do proposed market-based software systems.[12,22] Attempts to make M-style programs evolve through mutation consistently failed.[19])

The evolutionary process of design has another fundamental advantage, independent of the O-style/M-style distinction: it operates under different selection pressures.[7] For a replicator in nature, selective pressures depend directly on function. If a system doesn't work, its failure has a direct, negative impact. In the design process, however, selection pressures differ. If a design doesn't work, it may still be retained because it is promising. A whole series of unworkable designs some containing errors, others simply too sketchy to be implemented can all be the genetic ancestors (in a mimetic sense[2]) of a later design that is novel and workable. The freedom of design processes from the constant-workability constraint of ordinary evolution is a powerful advantage: it enables the introduction of huge numbers of simultaneous, coordinated changes in a single "generation" between working systems. This breaks down otherwise insurmountable barriers.

GENETIC DIFFUSION AND MATCHING

Genes in a population of sexually reproducing organisms diffuse, encountering each other in various combinations. Those combinations that "match," in the sense of being advantageous, tend to "stick" via differential reproduction of that pattern. The genes

themselves typically code for parts of diffusive, matching-based systems, increasing the opportunities for recombination to produce useful results. Holland[15] notes some subtle advantages of genetic recombination in testing combinations of genes, but an elementary, quantitative effect is also of interest in efforts to model biological evolution.

A naive model of biological evolution treats it as similar to early experiments in machine learning[19] or to Dawkins' simple computer model of cumulative selection[3]: a generation is equated to a single trial in which one or a few mutant individuals are generated and compared to a single parent, selecting the best as parent of the next generation. The accumulation of mutations in an individual of any generation can, in this model, be seen as the sum of past favorable mutations to that same individual, with unfavorable mutations being discarded.

Consider a population with genetic diffusion. Consider a time span long enough to spread a favorable mutation (and eliminate an unfavorable mutation). The number of generations required to spread a favorable mutation is a selection-pressure-dependent multiple of the logarithm of the population size, in a well-mixed population. For times that are long compared to this time span, the accumulation of mutations in any individual of any generation is (roughly) *the sum of the favorable mutations to all past individuals in the population*, discarding all unfavorable mutations. If the population size is a million, then for the naive model above to have comparable quantitative results, the favorable-mutation rate would have to be multiplied by a factor of a million.

A million generations of a species with a million individuals can be expected to achieve a modest amount of evolution, in biological terms. Many species produce a generation a year, and it has been hundreds of millions of years since animals emerged onto land. A review of early machine-learning experiments describes millions of generations as "an immense number," but this equates to mere millions of trials. A million generations of a million-member species equates to a *trillion* trial-lifetimes. Even at one simulated lifetime per second, this many trials would take a computer roughly 30,000 years.

In considering the evolution of the protein machinery of modern cells, bacterial numbers and generation times are relevant, since bacteria dominated the biosphere for billions of years and eukaryotes are relatively recent. (Note that bacterial genes diffuse through a variety of viral and "sexual" recombinant mechanisms.) A planet-wide monolayer of bacteria managing one generation per day would, in a billion years, make some 10^{38} trials. For comparison, this is the number of machine cycles that a trillion computers, each with a gigahertz CPU, would execute in about three billion years.

Genetic diffusion typically multiplies the rate of evolution by many orders of magnitude. Thus, the evolution of mechanisms for genetic diffusion is a prime example of evolving for greater evolutionary capacity. There is, however, no engineering reason, to include such mechanisms in typical nanoreplicators.

STYLES AND GENOTYPE-PHENOTYPE MAPΠNGS

One can describe the difference between O-style and M-style systems in a picture analogous to Sewall Wright's "genetic landscape" — an n-dimensional space, in which

each point corresponds to a combination of genes (actually, a combination of gene-*frequencies*, in Wright's model) and the "height" at that point corresponds to the combination's fitness. In this picture, evolution tends to climb hills and may be blocked from a certain path by a deep enough valley (to sink too low is fatal). To move from Wright's model to one appropriate here, we define two points to be neighbors if they are separated by a single mutation. (This gives it the structure of Dawkins' "genetic space."[3])

O-style systems, because of their flexible organization, can vary in many ways without drastic, deleterious results. They have a relatively smooth mapping of genotypes to phenotypes. Accordingly, their genetic landscape is relatively smooth and continuous, enabling many long, uphill runs.

M-style systems, because of their brittle organizations, can survive few variations. Most significant genetic changes cause a mismatch in the patterns of inert parts and positional assembly procedures, and plunge the system into a deep valley. These systems have a relatively discontinuous mapping of genotypes to phenotypes, in functional terms. Accordingly, their genetic landscape is dissected into a host of tiny, isolated peaks; smooth changes in genetic space do not lead to smooth changes in results. This blocks cumulative selection by preventing small, beneficial steps.

One can imagine an even worse situation. A replicator (whether M-style or O-style) could have an encrypted genome. Each offspring replicator would receive a copy of the genome, then decrypt it in order to read its instructions. With a suitable choice of encryption algorithm, every bit in the genome would affect every bit in the decrypted result, and a single-bit mutation would lead to an effectively random output, flipping half the bits. To expect such an output to be a viable program for replication makes as much sense as programming a computer to perform as a text editor by loading its memory with bits from a random-number generator. (This is not a recommended software engineering practice.) For such a system, with an everywhere-discontinuous mapping of genotypes to phenotypes, the viable peaks in the genetic landscape would consist of isolated points, and cumulative selection would be utterly impossible.

NANOREPLICATORS AND EVOLUTION

Proposed nanoreplicators will be M-style systems, using channeled transport of materials, energy and information, positional assembly of parts, a geometric structure, and rigid, inert parts. Thus, for them to evolve would require multiple, simultaneous, coordinated changes that are impossibly unlikely to occur by accident. (Further, they will lack mechanisms for genetic diffusion.) While such nanoreplicators are amenable to design and modification by engineers (or presumably by AI systems), it seems they cannot undergo significant evolution.

One can imagine designing O-style nanoreplicators, perhaps patterned on bioreplicators, but there is reason to believe that their flexibility would be bought at the price of reduced technical efficiency and ease of design. One can imagine building M-style nanoreplicators controlled by O-style software, allowing evolution through what would amount to clever genotype-phenotype mappings, or through heuristically guided

mutation (like that in EURISKO[20]). Achieving this would, however, entail a substantial research task beyond and independent of the task of developing a functional replicator. Further, it would lower the performance of the final devices by imposing computational overhead.

It seems that building a self-replicating molecular system based on nanomachinery does not entail building a system capable of evolution. Indeed, it seems that the latter would be a distinct and challenging goal.

NANOREPLICATORS, BIOREPLICATORS, AND EVOLUTION

The differences between O-style bioreplicators and M-style nanoreplicators are of more than academic interest if nanotechnology in fact will be developed, then it is important to understand its relationship to familiar biological models and biological hazards.

Living systems are obvious models for nanoreplicators: they replicate, and they are based on molecular components. Indeed, the physical principles they demonstrate provide a firm basis for projecting the feasibility of many of the molecular operations needed in proposed nanoreplicators.

Beyond this, however, the models diverge: in structure and function, proposed nanoreplicators resemble factories more closely than they do cells. The difference between O-style and M-style systems is, in this case, the difference between evolving and nonevolving systems. What is more, living systems are evolved systems, while nanoreplicators will be designed: where the former are shaped to serve the goal of their own survival and replication in a natural environment, the hatter will be shaped (whether well or poorly) to serve human goals, perhaps in an artificial environment. These differences greatly limit the utility of analogies between living systems and nanoreplicators. They likewise make genetic engineering a poor prototype for nanoengineering.

Genetic engineering today involves not design of replicators from scratch, but tinkering with the molecular machinery of existing bioreplicators. Since bioreplicators were not designed, they are not necessarily structured in a way that lends itself to understanding processes based on diffusion and matching allow complex nonlocal interactions that can be hard to trace; not having designed them or completely analyzed them, we still lack complete system specifications. These replicators can be crippled, but having evolved in nature, they resemble systems that can survive in nature. Typically, they are able to exchange genetic information with wild organisms, raising the possibility of the introduction of new, unconstrained replicators in the natural environment. Finally, having evolved to evolve, they have a capacity for further evolution — to serve their own survival, not human goals.

These concerns have inspired great caution regarding genetic engineering. They are substantially mitigated, of course, by the observation that nature has been tinkering with

genes for a long time, and that engineered organisms are typically modified in ways that do nothing to help them survive in competition with their wild cousins.

Nanoengineering, in contrast, will involve building replicators from scratch. Because nanoreplicators will differ fundamentally from biological systems, there is reason to believe that novel and remarkably dangerous systems could be constructed — but are they likely to appear by accident?

Several facts make such accidents easy to avoid and difficult to cause. The most obvious and least fundamental of these is that, since these systems will be designed, their parts and structures will be known; moreover, with M-style organization, the relationships among their parts will be designed and fixed. More important, however, proposed nanoreplicators will be fundamentally alien to the biosphere, unrelated to anything that has evolved to survive in nature. For reasons of efficiency and technical simplicity, it will he natural to design nanoreplicators to function in environments found only in special chemical vats (providing, say, hydrogen peroxide as a source of energy and oxygen), placing a straightforward limit on their spread. Finally, engineering experience shows that, while the "capability" to fail (even explosively) can appear by accident, the capability to perform complex organized activities does not. Replicating in a natural environment, without the assumed special chemicals, would be such a complex activity.

The basic benefits of "free-living" replicators can be had without building such things. In a replicating system, a device A makes multiple copies of A, enabling exponential growth in a special environment. If A can in addition makes copies of B, which can make C (but not A or B) which can make D (but not A, B, or C), then many copies of D can be had by starting with a stream of copies of B. Devices B, C, and D can operate in any environment without raising the possibility of uncontrolled, exponential replication– though a copy of B can ultimately give rise to many copies of D, none of the devices in the chain can replicate. The ability to produce large amounts of D without replicators reduces the incentive to build nanoreplicators that operate in natural environments.

Genetic engineering operates on the design level in modifying replicators that are themselves evolved to evolve. Nanoengineering will operate on the design level in constructing replicators that need have no ability to evolve; with such replicators, there will be a clean separation of the evolutionary mechanism (designs and designers) from the individual replicator's genetic mechanism (its embedded program). In light of general engineering practice amid specific efficiency concerns, this seems the natural way to proceed. Developers would have to go out of their way to give M-style nanoreplicators an evolutionary capacity, at a substantial cost in design effort and software complexity.

In light of all this, it seems that by simply neglecting to solve some difficult problems, we need never come close to building nanoreplicators capable of runaway exponential growth, or capable of evolving into systems that pose that threat. Scenarios of massive destruction are a concern,[7] but accidents seem easy to avoid. The problem to focus on is not that of accidents, but of deliberate abuse.

CONCLUSION

When we speak of life, we speak of organic self-replicating systems with structures and behaviors that result from evolution and have a capacity for further evolution. M-style self-replicating systems will result from deliberate design, and (barring extraordinary efforts) will lack the capacity for further evolution. As a consequence, their behaviors can be expected to be stable and designed to serve human goals (however imperfectly) rather than being mutable and evolved to act as robust survival-and-replication systems. Although it is sometimes useful to consider them from the perspective of biological systems, M-style nanoreplicators will differ from organisms in such fundamental ways that it would be misleading to describe them as living things. It is entirely accurate to call them machines.

The pursuit of genuine artificial life will require special attention to O-style organization, or to conditions that can lead to its evolution. Work in artificial life will not automatically be furthered by the pursuit of useful nanomechanisms and self-replicating systems.

REFERENCES

Becker, R. S., J. A. Golovchenko, and B. S. Swartzentruber (1987), "Atomic-Scale Surface Modifications Using a Tunnelling Microscope," *Nature* **325**, 419-421.

Dawkins, R. (1976), *The Selfish Gene* (New York: Oxford Univ. Press).

Dawkins, R. (1987), *The Blind Watchmaker* (New York: Norton).

Dawkins, R. (1988), "The Evolution of Evolvability," these proceedings.

Drake, J. (1969), "Comparative Rates of Spontaneous Mutation," *Nature* **221**, 1132.

Drexler, K. E. (1981), "Molecular Engineering: An Approach to the Development of General Capabilities for Molecular Manipulation," *Proc. Natl. Acad. Sci.* **78**, 5275-5278.

Drexler, K. E. (1986), *Engines of Creation*, (New York: Doubleday).

Drexler, K. E. (1986), "Molecular Engineering: Assemblers and Future Space Hardware," *Proceedings of the 33rd* Annual *Meeting of the American Astronautical Society, Boulder, October, 1986.*