Inventions, Algorithms, and Biological Design

By John Bracht

Abstract—This paper outlines a new model for understanding biological invention, based upon an extensive study of human inventiveness originating in Russia shortly after the Second World War. This science, known as the Theory of Inventive Problem Solving and referred to by the Russian acronym TRIZ, relies upon the study of human patents to reveal general principles of invention. TRIZ recognizes two distinct mechanisms of invention, operating upon two distinct types of problems. The first, trial and error, is a remarkably accurate description of the neo-Darwinian mechanism, natural selection. Furthermore, the trial and error mechanism has been found to be severely limited in the sorts of problems it can solve; these limitations are also found to apply to the Darwinian mechanism. The second mechanism, lacking an explicit TRIZ name but referred to here as the intentional mechanism, is the source of true inventiveness. The science of evolutionary programming gives insight into precisely why the intentional mechanism is required; certain fundamental parameters must be given before the Darwinian mechanism can even operate and these parameters are themselves out of reach of the Darwinian mechanism. Certain key events in the history of life require alterations of this sort of fundamental parameter, and it is precisely these events that the neo-Darwinian model fails to explain. Consequently, it is the inescapable conclusion that there is a second mechanism, an intentional mechanism, which operates in nature and is responsible for the changes that are not accountable via the Darwinian mechanism.

Introduction

Invent: to produce or contrive (something previously unknown) by the use of ingenuity or imagination. –The American Heritage Dictionary, second college edition, 1985

A defining feature of intelligent agency is inventiveness—the production of something previously unknown and, by definition, new. The painter and poet, engineer, inventor, and scientist all rely upon some form of invention to ply their trade. Inventiveness has long been the object of study with the goal of enhancing the problem solving process. The ancient Greeks began the search for heuristics in hopes of discovering a universal problem-solving, inventive procedure. Indeed, the art of invention appears to be a fundamental human activity; early human remains are recognized, in part, by the presence of tools and other artifacts that were invented.

Yet for its importance, the art of invention remains largely mysterious. Inventor John Rabinov commented, "it would be nice if inventions were the result of a logical and systematic process. Unfortunately, this is usually not the case. They are the product of what psychologists call 'intuition'—a sudden burst of inspiration; a process hidden inside the human mind."[1]

Nonetheless, a detailed study of human inventions, recorded in patent records in the former Soviet Union, has provided key insights into the inventive process and a unifying framework that is broadly applicable to other fields—particularly, the study of biological organisms. Biological evolution is undoubtedly an inventive process, capable of generating a vast catalog of "inventions" –the biotic world. The insights gained from the study of human patents provide a model of how inventive processes work; thus, insight into how life itself has come to be.

Inventor and author Genrich Altschuller began the study of Soviet patents in 1946. In his lifetime, he studied 400,000 patents. Since that time, the founding members of the "Theory of Inventive Problem Solving"—known by its Russian acronym "TRIZ"—have

extended this study to over 2 million patents. However, only since the downfall of the Soviet Union has TRIZ been available to those in the West. In the last 10 years TRIZ has been making large inroads into the US through conferences and seminars aimed at teaching people how to use TRIZ in business and industry.

The Basics of TRIZ

TRIZ begins with a critique of existing problem-solving methods, especially the age-old trial and error technique. TRIZ theorists have characterized many deficiencies in this method, and indeed a primary reason to develop TRIZ was to provide an alternative to trial and error.

There are several different variations on the basic trial and error technique, but they all share a common essence: the sifting of variants. When faced with a problem, the inventor simply begins to generate possible solutions. As each candidate solution proves faulty, more are created and tested. As variants are eliminated, it becomes ever easier to find a solution.

However, there is a fatal flaw in this method, and Altschuller points it out: "The weakness of the trial and error method is clear…At first, it seems that the method is chaotic. This is not true. There is a definite system to it. Trials are made along the direction of least resistance. After all, it is much easier to pursue a familiar direction. The inventor follows this same route subconsciously—therefore, the odds are against him discovering anything new."[2] This characterizes the problem of *psychological inertia*, in which an inventor's previous experience causes him or her to produce variants of already existing systems rather than branching out into uncharted territory. Seymon Savransky, another TRIZ theorist and author, writes, "Psychological inertia focuses the mind on what is known, i.e., along the assumed search direction, thereby keeping the solver from the right solution. This inertia is useful only when the solution's direction is recognized correctly."[3] Psychological inertia is the key weakness of trial and error because it limits the sorts of solutions the problem solver can discover; he can only find solutions that may be reached by variation of concepts he is already familiar with. As we shall see later, this has far-reaching implications for the types of problems that are solvable via trial and error and for the inventive process itself.

In studying the process of invention, Altschuller focused a great deal of attention upon the sorts of problems that need solving. He found that there are two types: inventive problems and routine problems. Inventive problems always involve some sort of *technical contradiction* that must be the resolved. A contradiction, in logic, refers to a statement that cannot be true, written as "A & ~A." A technical contradiction embodies the same sort of logic.[4] In an inventive problem, making one aspect of an object better will cause another aspect to get worse, thus, there will be a need for A and ~A. Altschuller points to the icebreaker ship for an example of a technical contradiction in an inventive problem.[5] The normal way to improve the icebreaker's ability to penetrate the ice is to install larger engines, thereby increasing engine power —but there is a cost associated with that increase in engine power. The increased weight of the ship and the complexity of the support systems (cooling systems, massive gasoline tanks, etc) for massive engines effectively counteract any gains from increased engine power. Thus, there is a need for A (bigger engines) and not A (because of the weight and complexity

costs of those engines). This trade-off describes a classic technical contradiction. The TRIZ model proposes to overcome this contradiction, rather than simply finding the best compromise between engine power and ship weight. This requires a radical re-design of the entire system rather than simply modifying the existing system. For example, a solution proposed by the TRIZ method is to re-make the hull such that the upper portion of the ship (above the ice) is separated from the lower portion (below the ice) and thin walls along the left and right connect the upper and lower sections. This creates a hollow "tube" which can slip through the ice, cutting only at the left and right edges, and the leading edges can be made quite thin. The result is that friction with the ice is greatly reduced, allowing a reduced engine size, thereby increasing cargo space. This solution not only resolves the technical contradiction, but it also allows the cargo ship to be subsumed into the icebreaker. Notice how the effective resolution of the contradiction required a departure from psychological inertia, the tendency to assume that an icebreaker must have a hull shaped like most other ships. Because of this psychological inertia, the tendency in the evolution of icebreaker technology has been toward increasingly powerful engines—but this trend is doomed to forever wander in the wastelands of sub-optimality. Only the overcoming of a contradiction will propel the technology to new heights and new possibilities. And, in the study of problems and inventions it is the technical contradictions that hold the key to understanding the nature of inventiveness.

A technical contradiction forms the heart of an inventive problem—but not all problems are inventive. The non-inventive problems are known as *routine problems*; solving them only requiring modification of some characteristic of the existing system rather than the resolution of a contradiction. Savransky describes the different problem types this way: "In contrast to a routine design that leads to a smoothing of the contradiction (the trade-off dogma) or choosing one of the preferable combinations in the conflict (OR…OR), a design based on TRIZ aspires to permit and solve the contradiction, creating a system in which the improvement of one characteristic is not accompanied by deterioration of others (the AND…AND), so the so-called win-win principle can be achieved."[6] The routine problem may be solved by compromise, but the inventive problem requires resolution of a contradiction before it can be solved. In the example of the icebreaker, the ship may initially be improved by solving the *routine* problem of increasing engine power. But eventually no more improvement is possible by this route because the system reaches a local peak in functionality and any further increase in engine size will actually decrease the overall performance. Suddenly a contradiction arises and we cannot make progress. A routine solution would move the system back to the best compromise, and try to accommodate the contradiction. True progress, true invention, requires moving to an entirely new hull design, thereby removing the contradiction.

Therefore, TRIZ theory highlights the necessary connection between inventive problems, resolution of contradictions, and a move to a system that is in some way new. Routine problems, on the other hand, always produce compromise solutions, the "smoothing" of the contradiction rather than overcoming it, and the modification of the existing system without introducing anything new.

TRIZ theory also delineates two sharply distinct mechanisms for the two types of problem. The trial and error mechanism, or sifting of variants, is capable and quite useful

for the resolution of *routine* problems. However, solutions to *inventive* problems are the driving force behind qualitative change in technology, and as described earlier, the precise mechanism by which such invention occurs is shrouded in the psychological processes deep within the human mind. Studies indicate that the intuitive "flash" or inventive insight is tied closely with our more general creative and learning abilities. Joan Soloman, professor at the Centre for Science Education at The Open University, England, notes, "…[C]reativity includes two rather similar processes which have in common a gestalt shift within personal thinking:

- Moments of creativity when a theory or concept '*makes sense*' or '*clicks*';
- Moments of creativity in design when a '*new solution*' to a problem strikes us."[7]

It is precisely this creative gestalt shift within personal understanding which best characterizes the inventive mechanism and has inspired humans for centuries to try (without much success) to probe the process of inventiveness.

What is clear from this analysis is that inventiveness has two basic mechanisms for two basic types of problems. Trial and error techniques are useful for routine problems in which variants are sifted until the correct solution is found. However, this technique is fundamentally incapable of generating real novelty and there is another mental process, a form of creative insight or intentionality, which engenders qualitatively new systems.

One remarkable aspect of the trial and error or "sifting variants" concept is how closely it parallels classic definitions of the Darwinian mechanism of random variation and natural selection. In trial and error, the inventor generates the variants and acts as the selective mechanism by preserving those variants that solve the problem or seem to lead toward a solution. This is a form of artificial selection, a concept that Darwin himself used (in the example of dog and pigeon breeding) as a prototype for natural selection in developing his arguments in the *Origin of Species.*

Obviously, when nature generates variants and selects between them, there is no guiding intelligence as there is when humans utilize trial and error mechanism. However, the analogy still holds. For the fact that nature has no guiding intelligence directing the process of variation and selection means that it is more constrained than human-guided selection. The key problem of psychological inertia, which plagues trial and error processes, is amplified in natural selection because natural processes simply *cannot* get "outside the box" to produce entirely new solutions. Natural selection is limited to variants upon existing systems, whereas a human inventor could conceivably invent something new. Therefore, the problem of psychological inertia is endemic to natural variation and selection, but in a different form (since natural selection has no psychology) which I shall call *morphological inertia*. Morphological inertia simply states that nature is restricted to variants upon already existing systems. This is an extreme form of the psychological inertia that plagues trial and error systems, thereby rendering the Darwinian mechanism equally ineffective at creating inventive solutions. This insight from TRIZ theory is reinforced and clarified further by recent findings in the field of computer science and genetic algorithms.

Algorithmic Inventiveness

The field of evolutionary computation has been in existence since the 1960's, when John Holland began creating *genetic algorithms* to explicitly model how evolution

operates in nature. Since that time much work has been done in the field; computers offer an ideal situation in which to speed up, *in silico,* the long, slow process of evolution while (hopefully) retaining the essential characteristics of that process. This, in turn, opens up new possibilities for answering questions such as: why, exactly, can't trial and error produce inventive solutions? In other words, why can't Darwinian processes produce anything *new*? What, exactly, can trial and error produce? What sorts of changes require an inventive mechanism?

To begin answering these questions, I first turn to an example drawn from the writings of Richard Dawkins, professor of the Public Understanding of Science at Oxford University. In chapter 3 of his book *The Blind Watchmaker*, Dawkins describes a computer program he wrote which uses a recursive branching function to "grow" trees of varying degrees of bushiness, height, and width.[8] There were nine "genes" in each organism, and each gene consisted of an integer variable corresponding to one parameter of the resulting tree—such as angle of branching, depth of recursion, etc. The program produces and displays several trees at once, each tree only differing by one integer (gene) value from the ones around it. A human actively selects one of these trees to "reproduce" and leave offspring (with "mutation"—random altering of one gene up or down one integer value). The program then plots a number of new trees (parent plus several mutants), and the process repeats itself. Over many generations of mutation and selection, forms of striking beauty and complexity can be produced. The program's output quickly leaves any semblance of "trees" and produces images remarkably reminiscent of caterpillars, steel girders, airplanes, or even beetles and ants. Dawkins named his creatures "biomorphs" and speaks fondly of wandering through Biomorph Land in search of new discoveries.

Let us ignore the fact that Dawkins has successfully modeled artificial selection rather than natural selection since intelligent agency does the job of selecting organisms to reproduce. Rather, I use this example to bring out the nature of the total space of possibilities in which this program may operate. The nine genes of the biomorph program serve to lay out a space (Dawkins's "biomorph land") through which the program moves with the help of artificial selection. Dawkins notes, "I wanted to try to represent this genetic space in the form of a picture. The trouble is, pictures are two-dimensional. The genetic space in which the biomorphs sit is not two-dimensional space. It isn't even three-dimensional space. It is nine-dimensional space!"[9] Each gene in the program induces a corresponding dimension in the space of possibilities. Thus, if the biomorph program were modified to have only two genes, it would operate in a two-dimensional space in which every possible combination of values for those genes would be represented. The same holds for three genes and three dimensions, and for nine genes and nine dimensions. For each program, there is an n-dimensional hypervolume of possibilities in which that program operates, with n equal to the number of variable parameters. In the language of William Dembski's design inferential machinery, this n-dimensional hypervolume is equivalent to the reference class of possible outcomes, Ω.[10] The encoding of the program completely determines the reference class in which that program may operate. The fixed parameters of the genetic algorithm completely determine the possibilities that may be generated and tested by the program. In the case of the Biomorph program, the number of genes is fixed and cannot be adjusted by the evolutionary process the program simulates.

This observation suggests that we may consider any genetic algorithm to be operating within a certain n-dimensional hypervolume, and certain fixed parameters completely determine that hypervolume ahead of time. Furthermore, any particular n-dimensional hypervolume is completely isolated and separate from any other m-dimensional hypervolume ($m \neq n$). As an obvious example, consider how three-dimensional objects are in a completely separate class from two-dimensional objects. Altering the width, length, or other characteristics of a circle will never convert the circle into a sphere; circles and spheres are from completely separate hypervolumes.

The essential insight is that trial and error may only operate *within* a given hypervolume—but it may never jump to a new, higher-order hypervolume. The unbridgeable gaps between hypervolumes correspond to the technical contradictions in TRIZ theory. Recall that the overcoming of technical contradictions (or, synonymously, the jump to a new hypervolume) is the essence of inventiveness. This is precisely the reason that trial and error cannot invent new things; it is confined to operate within a pre-defined hypervolume of possibilities and therefore cannot overcome contradictions; therefore, it cannot invent. This hypervolume is fixed by certain non-varying parameters (In Dawkin's Biomorph example, the number of genes and the rules regarding how the integer values of each gene are interpreted) that an intelligent agent must set and which are not allowed to vary. These fixed parameters are found in the *encoding* of the program and provide the theoretical grounds for why routine problems and inventive problems are isolated from each other and cannot be solved by the same mechanism. A few more examples will help clarify.

The next example again comes from Dawkins's biomorph program. There exists an updated version of his program, available on the web, which has several new features not present in the original program.[11] The new program incorporates several new features, such as variable color, segmentation, and line shape. In all, six new genes have been added to the original program. This means that the later version of the program operates within a 15-dimensional hypervolume. Furthermore, there are many biomorphs within the 15-dimensional hypervolume that are completely inaccessible to the 9-dimensional hypervolume. The original program can never "evolve" a blue biomorph because to do so would require jumping the gap between hypervolumes. Indeed, it took the intentional actions of a computer programmer (possibly Dawkins himself) to add an extra gene to the original program in order to create the capability for a blue biomorph to be generated. The original encoding had to be modified such that there was a new gene, and a mapping function that could interpret that gene as a particular color on the screen. Notice that none of this can just "evolve" from the capabilities of the original program, because they affect the foundational characteristics of the program itself—characteristics that are fixed and form the boundaries within which the program operates.

While an undergraduate student at New Mexico Tech I attended a presentation that demonstrated another key feature of genetic algorithms: the crucial role of the fitness function. The public lecture was given on February 21, 2001, and the speaker was Dave Thomas, president of the local skeptics group New Mexicans for Science and Reason, and alumnus of New Mexico Tech in mathematics. In his lecture, Thomas presented a genetic algorithm that was designed to solve the Steiner problem. The problem entails finding the network that connects five pre-given points with minimal path-length. In true Darwinian fashion the program begins with a set of random networks, and with rounds of

mutation and selection it converges on a small set of minimal networks. Occasionally, the program even finds the universal optimum Steiner solution. Most of the time the program gets stuck in local optima with very short networks that are not quite as good as the Steiner solution. After the demonstration I had an email exchange with Thomas (personal communication), and I pointed out that the program created no real novelty and no information besides the information originally contained within the fitness function itself. My logic was as follows: the desired solution has (1) all five points connected, and (2) the shortest path-length. The program selected for networks that (1) connect all five points, and (2) have shortest path-lengths. It is no wonder that the program converges regularly upon short, optimum networks; it has been told precisely what to do by explicit instruction in the fitness function. Furthermore, the encoding of the program is also a key part of the problem solving process. The encoding of the program was carefully selected to fit the problem to be solved—the program was given five pre-existing fixed points, the possibility of adding floating points, and some way of interpreting its "genome" as line segments. All this encoding places the program in the hypervolume of networks and the Steiner problem. Furthermore, the fitness function explicitly targets the Steiner solution within that hypervolume—and the program simply follows the fitness function to find the answer. This holds with perfect generality; in any and every evolutionary algorithm it is possible to pinpoint precisely those parameters that have been set by an intelligent agent, parameters that must be carefully coordinated to allow the program can do the design work the programmer has in mind.

The SELEX procedure provides another example of the need for intelligent choices to guide the functioning of trial and error (evolutionary) processes.[12] SELEX, or Systematic Evolution of Ligands by EXponential enrichment, is a procedure in which an initial pool of randomized polynucleotides (RNA or DNA, single stranded) is created, containing on the order of 10^{15} molecules of a fixed length. The pool is then screened for some desired characteristic (for example, binding affinity for ATP). The molecules that are selected in this way are used as "parents" in the synthesis (with mutation) of a new pool of molecules, and the process repeats with more rounds of selection and amplification. The result is a set of highly functional molecules of DNA or RNA that perform their function while the scientists who use them are in complete ignorance of the actual sequence of nucleotides in the final product. Surely, proponents argue, this is an example of trial and error inventing something new?

Those who point to SELEX as an innovative process ignore the fact that humans have really done the innovation in setting up the experiment. First, humans created the initial pool of RNA or DNA molecules, and decided upon the length of the polynucleotide molecules. This pre-sets the number of dimensions in they hypervolume and the options that the evolutionary process may test. Secondly, the experimenters decided what to select for, and it is here that the real design work took place. For it is theoretically possible for a SELEX experiment to select for almost anything—binding to any molecule or catalyzing any reaction. The class of possible fitness functions (each of which would select for a different subset of RNA or DNA molecules) is enormous, and certainly larger than the class of molecules screened by the SELEX experiment. Therefore, it requires more information (rules out more possibilities) to set up the fitness function than it does for the chosen fitness function to select the optimum RNA/DNA molecules. In the words of William Dembski, it is "not a free lunch"—the information produced by the program

cannot be generated for free but requires intelligence to input an equal or greater amount of preexisting information.[13] In short, both the "encoding"(the length of polynucleotide and the characteristics of polynucleotides synthesis) and the fitness function are set by an intelligence seeking to invent a molecule with novel characteristics.

The take-home lesson is that selection and mutation processes can operate within pre-set hypervolumes to find solutions that we know *exist* but which may be intractable given our current knowledge. However, they cannot find the hypervolume or the fitness function apart from intelligence—we still have to do the design work (getting the program into the right hypervolume where a solution may be found, and then finding the right fitness function over that hypervolume) before the algorithm can take over and sift through the vast possibilities to find a workable solution.

Why, in theory, can't a genetic algorithm evolve the foundational parameters (the hypervolume and fitness function) before it optimizes within that hypervolume? The answer is that an evolutionary algorithm simply has certain requirements that must be in place *before it can operate*. It needs something that can be affected by chance (mutation), therefore it needs some variable characteristic that exists prior to the evolutionary process itself. It also needs a fitness function that allows selection to distinguish between competing possibilities. Once caveat is worth mentioning: some ingenious programmers have created evolutionary algorithms in which the fitness function itself is allowed to evolve (see, for example, David Fogel's checker-playing program[14]). However, such programs just push the problem back a step and install a higher-order fitness function that selects the desirable lower-level fitness function from the reference class of all possible lower-level fitness functions. In other words, the fixed fitness function may be pushed back a step or two, but it is still there and cannot be avoided. Furthermore, the selection of the super-fitness function from a super-reference class entails the generation of more information than does the selection of the sub-fitness function from the sub-reference class. The information problem does not go away; it just gets worse as it is pushed further back.[13]

Perhaps another question will be helpful: why is intelligence required to set up a hypervolume and fitness function? The answer can be summed up in one word: intentionality. Intentionality as I am using it refers to the ordering of matter and energy in such a way as to achieve some goal or purpose. In genetic algorithms, intentionality is required to set up the hypervolume and fitness function, coordinating them in such a way that the program can find the goal that the programmer had in mind. Chance and law simply don't have any means of ordering matter or energy in such a way that a distant goal will be achieved and they are thus in principle incapable of doing the work of an intelligent agent.

Another way to look at the problem comes from the Artificial Intelligence (AI) movement. One well-known problem for AI is the *frame problem*. It is known that humans, when faced with a particular problem to solve, somehow manage to consider only a relevant subset of all the information stored in memory—thereby weeding out a vast store of irrelevant information. All efforts to duplicate this process with artifical intelligence have, to date, failed. It seems that the frame problem is only solvable by an intelligence, and that the agent must somehow use intentionality to direct the selection of relevant information. Setting up a hypervolume (choosing the fixed parameters of a system) is directly analogous to picking out the relevant information to solve a problem,

since it directly determines which possibilities the program can generate and test—and those possibilities must include the solution to the problem (that is, the fixed parameters must be relevant to the problem). Therefore, if a program was created which could somehow "evolve" the hypervolume needed to solve a problem (without intelligent input), it would effectively be solving the frame problem—something which the AI community has thus far failed to do with a non-intelligent system and which, if this analysis is correct, is impossible without intelligence.

The implication is that genetic algorithms (and SELEX experiments) have not been found to do anything novel or new—they can sift through a huge space of possibilities and find the optimum (as defined by the fitness function) but setting up the parameters of the program practically solves the problem before any sifting takes place. Indeed, Melanie Mitchell notes

> Choosing a fixed encoding ahead of time presents a paradox to the potential GA user: for any problem that is hard enough that one would want to use a GA, one doesn't know enough about the problem ahead of time to come up with the best encoding for the GA. In fact, coming up with the best encoding is almost tantamount to solving the problem itself![15]

Geoffrey Miller of University College London, comments,

> But for hard problems and very large design spaces, designing a good genetic algorithm is very, very difficult. All the expertise that human engineers would use in confronting a design problem—their knowledge base, engineering principles, analysis tools, invention heuristics and common sense—must be built into the genetic algorithm. Just as there is no general-purpose engineer, there is no general-purpose genetic algorithm.[16]

By inputting large amounts of specific information into the system, the programmer/experimenter gives the program or system everything it needs to solve the problem—but in doing so he or she constrains the mutation and selection process such that it is highly specific to one particular problem and useless for other problems. To see this specificity, compare two evolutionary processes: a SELEX "evolution" experiment and the observed evolutionary increase and decrease in finch beak sizes on the Galapagos Islands. The SELEX reference class (hypervolume) is completely isolated from the finch reference class (hypervolume). No matter how much the RNA or DNA molecules of a SELEX experiment evolve within their own reference class, they will never be able to produce a finch, finch's beak, or even a single finch cell—they are simply not in the right reference class. What would it take to go from the RNA/DNA world of SELEX to the multi-cellular world of finches and their beaks? For starters, there is a cellular context that needs to be added. Furthermore, that cellular context must be embedded within a multicellular context. But not just any multicellular context will do; rather, the "finch" context, complete with the complex systems that make up the finch physiology, is required. It must have highly specific signaling and informational pathways in place before things like finch beaks can exist in the first place, much less be altered by mutation and selection. In order to move from the RNA/DNA world of SELEX to the

world of finch beaks, the entire multidimensional hypervolume (or reference class) in which the system resides must be re-engineered—and that requires making intelligent choices that exhibit intentionality. Put another way: once you are within one particular hypervolume of possibilities, it makes sense to speak of finding the fittest individual (assuming that you have defined some fitness criterion)—but that fitness function does not operate universally across hypervolumes. The fitness functions are reference-class specific; it makes no sense to speak of selecting for a finch beak's ability to catalyze a chemical reaction or a polynucleotide's ability to break open hard seeds. The logical conclusion is that something other than a mutation and selection mechanism must be responsible for generating new systems that reside within isolated hypervolumes and which evolve with very specific (local) fitness functions.

At first encounter the "RNA to finch" transition seems highly contrived, but if the naturalistic story of the origin and subsequent development of life is true, something very like this transition must have occurred. The first living molecules (often postulated to be from a primitive "RNA world") were capable of catalyzing certain reactions including their own replication. After around 3.5 billion years, the evolutionary process had produced complex things like finches and humans. However, a major question is this: is the evolutionary process in principle capable of producing the sort of change we observe in going from a self-replicating RNA/DNA molecule to a complex, multicellular organism? The model proposed by TRIZ and genetic algorithm theory suggests that it is not, and application of this model to a commonly cited example of evolutionary programming will demonstrate once more the reliance of the evolutionary process upon intentional agency.

Thomas Schneider, research biologist with the Laboratory of Experimental and Computational Biology at the National Cancer Institute, has published a paper in *Nucleic Acids Research* titled "Evolution of Biological Information."[17] In this paper he presents a genetic algorithm which generates a population of 64 "organisms" which each have a genome 256 "base pairs" long, taking any of four values (a,c,g, or t) at each "base pair." The genome is divided into two parts: a weight matrix and a set of binding sites. The weight matrix determines the numerical value associated with each "nucleotide" in the binding sites, and also contains a "threshold" value. If a binding site has a numerical value (the sum over nucleotides in the binding site) above the threshold (as determined by the weight matrix), a hypothetical protein is considered to have "bound" to that site. If the value is below the threshold, the protein did not bind and the program considers this to be a "mistake" which counts against the organism in terms of fitness and probability of reproducing. Also, if the a protein "binds" to an area outside of a binding site (in other words, the numerical value of a non-binding area is above the threshold) it is considered to have made a mistake. At each generation, the half of the population that makes the fewest mistakes is allowed to reproduce (with mutation) and replace the half making more mistakes. Schneider notes, "Remarkably, the cyclic mutation and selection process leads to an organism that makes no mistakes in only 704 generations."

But precisely how remarkable is this result? According to Schneider, "The ev model show explicitly how this information gain comes about from mutation and selection, without any other external influence, thereby completely answering the creationists." But has the information gain really been driven only by mutation and selection? On closer inspection it becomes clear that Schneider loaded a tremendous amount of information

into his program from the beginning, and that what the program accomplished was not as monumental as he would have us believe.

First, consider the external influence that Schneider has imposed upon the program by setting up the fixed parameters. There are many, but here is a partial list:

- genome size
- population size
- mutation rate
- number of binding sites
- length of binding sites
- position of binding sites (different from run to run but fixed within a run)
- overlap of binding sites not allowed
- interpretation of values in the binding sites
- position of the weight matrix
- position of the different nucleotide encodings within that weight matrix
- length of nucleotide encodings in the weight matrix
- position and length of the threshold value encoded in the weight matrix
- overlap of weight matrix encodings not allowed
- interpretation of the values in the weight matrix
- initial values given to a, c, g, t for the weight matrix to begin assigning values
- relationship between threshold value, binding site value, and successful binding
- criterion for success: fewer mistakes
- definition of what constitutes a mistake

In particular, the position and length of binding sites, the position and length of elements in the weight matrix, and the ways that these values and the threshold value are interpreted by the program are the essential parameters that map out the hypervolume in which the program functions.

Taken together, these parameters define the class of all possible "organisms" which the program can generate and test, and these parameters were all deliberately chosen by Schneider to produce the complex and interesting result. The fitness function counted as "mistakes" those instances in which a binding site's value is below the threshold or a non-binding site's value is above threshold. Therefore, the fitness function selected directly for organisms that had high binding site values (above threshold) and low non-binding site values (below threshold). However, Schneider's condition of success was precisely the same: to create a set of organisms that had binding site values above threshold and non-binding areas below threshold. He encoded the very conditions of success into his fitness function—and then claimed that his program had created this information from scratch!

Notice that the choice of the fitness function *could* have been any one of a near infinite collection of possibilities, and each fitness function would guide the program to a different target. (Just a few examples of different fitness functions: binding site value must be below the threshold, binding site value must be greater than ½ threshold, binding site value must be less than 2/3 threshold, binding site value times threshold < 100, etc). Therefore, by making the selection of one fitness function from the enormous array of possible fitness functions, Schneider created a great deal of information and injected it into the program. Likewise, the fixed parameters that define the hypervolume (like the

position and length of binding sites, position and length of weight matrix, and position and length of threshold value, and the logic for interpreting those encodings) were arbitrarily chosen from a near infinity of possible values and thus they also introduce a huge amount of information into the system. Contrary to Schneider's assertions, this program does not generate information "from mutation and selection, without external influence." The external influence was Schneider himself, who placed the program within the appropriate hypervolume (the general "ballpark") through his choice of fixed parameters, and told it precisely the characteristics of the solution he wanted through his choice of fitness function (the "map" of the ballpark, showing the program exactly where to go). The program did not innovate or invent anything new; it just followed the directions given by the programmer. The innovation, design, and creation of information are not properly attributed to the Darwinian process, but rather to the intelligence that set up the whole process in the first place.

The implication, then, is clear: Darwinian processes and trial and error are of necessity parasitic upon an underlying inventive, intentional process. They cannot innovate; they can only optimize (using a pre-set fitness function) within the pre-set range of possibilities. This is not to denigrate the capabilities of Darwinian processes, which can provide useful solutions to problems that are currently intractable to direct solution (as in the SELEX experiments), but it is rather to caution against attributing more to them than they are capable of delivering.

It is now clear why TRIZ maintains such a clear distinction between routine and inventive problems. Routine problems may be found by "smoothing" a contradiction—finding the best compromise solution within the pre-existing multidimensional hypervolume. Inventive problems, on the other hand, require resolving a contradiction—and that requires moving to a new hypervolume.

An example from TRIZ will help to clarify how this point plays out in the real world. This next example comes directly from the TRIZ literature;[18] it illustrates well the process of inventive thinking. A group of engineers were working on a machine that makes hardened steel balls (perhaps ball bearings). They had a problem with one part of the machine, however: a bent steel pipe through which the steel balls were carried at high velocity by air pressure. The problem was that the balls would wear out the pipe by repeatedly impacting the side when going around the bend. No matter how thick the engineers made the pipe, it still wore out. Many different materials were tried, but nothing could be made thick enough and strong enough. Finally, the inventive solution was found. A magnet was attached to the outside of the steel pipe, at the point where the steel balls impacted the wall. The magnet caused some of the balls to stick to the inside of the wall, and these balls provided a protective layer. Sometimes a few balls would be knocked loose, but more would be attracted to the magnet and take their place. In this way a self-maintaining protective layer was created just where it was needed.

Notice how the routine solutions (thickening the pipe wall) were unable to solve the problem because they never overcame the technical contradiction. The wall needed to be stronger but simply couldn't be made strong enough to withstand the constant wearing. This was the wrong approach—but it is a classic example of sifting through variants with psychological blinders in place. The sifting of variants within the existing hypervolume did not have a means of overcoming the technical contradiction and finding the inventive solution. Rather, a new piece had to be added to the system, thereby making a jump to a

higher-order hypervolume. With the magnet in place, there are several new variable parameters in the system. For example, one may increase or decrease the strength of magnetic field, or move the magnet to different locations on the pipe. These options add new dimensions to the hypervolume, and open up a new world of possible solutions that were inaccessible to the old system.

Notice how the sifting of variants (trial and error) could not move to a new hypervolume because it was restricted to varying the characteristics of the existing system. Adding a new piece to the system re-engineered the hypervolume in such a way that the solution was accessible. This re-engineering of the hypervolume is the essence of an inventive solution or invention. And it is precisely this re-engineering of the hypervolume that is out of reach of the Darwinian process because it requires the ability to step outside the hypervolume and make external changes. Endlessly sifting variations of some basic structure will never create a new structure because it moves from point to point within the hypervolume itself—it can never break out of the hypervolume to find something new. This is another way to re-state the problem of morphological inertia: trial and error is locked into variations of existing structures. Trial and error is shortsighted in the sense that it can only "see" to the edges of the hypervolume—and it can go no farther.

Biological Inventiveness

It has become evident that large changes and radical re-designs, which require re-working the hypervolume of possibilities, are off-limits to the Darwinian process, but we may expect organisms that differ only slightly, and co-exist within the same hypervolume, to evolve into each other. This accords well with the observed ability of evolution to change a finch's beak into a slightly larger finch's beak, or to alter the dominant color of moths in a population, or to produce bacteria that are resistant to penicillin. Each of these changes requires a variation of existing structure, and occurs within the hypervolume of possibilities in which the organism is situated. For these sorts of changes the Darwinian mechanism is ideally suited. However, there are other changes that require a fundamental re-engineering of the biological hypervolume, and which cannot, in any case, be considered mere variants upon previously existing systems. Interestingly, these are the very events that currently puzzle biologists most, and which have proven most recalcitrant to Darwinian explanation. A partial list of such puzzles include:

- the origin of life
- the origin of eukaryotes
- the origin of sexual reproduction
- the origin of multicellularity
- the origin of body plans (phyla)
- the origin of consciousness

Each of these requires a jump to a new, higher-order hypervolume in which new features appear and new variants are possible. The origin of life itself is a case in point. Just as in evolutionary programming the reference class of possibilities is determined by a pre-given and fixed set of rules, biological organisms also contain encoded information that relies upon a fixed set of rules and which, in turn, determines the reference class of possibilities to which the organisms belong. One such rule is the so-called "central

dogma" of biology, which states that information always flows from DNA to RNA to protein (with the exception of reverse transcription, in which information may flow from RNA to DNA). Another example is the fact that living things produce DNA, RNA, and proteins out of the surely huge number of possible macromolecules that could be produced from different arrangements of matter. These first few fixed rules originated with life itself, and it is the origin of life which remains as one of the great unsolved mysteries of modern science. Indeed, most scientists separate the origin of life entirely from the question of evolution, and freely acknowledge the difficulty of explaining this first mystery of biology. It is included here only for completeness, as one of the most stunning occurrences in life's history, and the most starkly mysterious. However, it is easy to point to other events after life originated which are very nearly as challenging as life's origin, and just as recalcitrant to a Darwinian explanation.

One of these Darwinian mysteries is the origin of eukaryotic cells from prokaryotic ancestors. The eukaryotic cell is highly complex and contains features that are not present, in any form, in the prokaryotic cell. For example, the eukaryotic cell has a nucleus (containing numerous, histone-containing chromosomes as opposed to the naked DNA in the single chromosome of prokaryotes), mitochondria, chloroplasts, endoplasmic reticulum, Golgi apparatus, cilia, cytoskeletal elements, and different gene structure (containing non-translated portions or introns), along with the different machinery to make sense of the new gene structure—to name a few of the differences. Clearly, the origin of eukaryotic cells is not a matter of altering the parameters of early prokaryotes—some major, qualitative changes had to occur. Once more, the current Darwinian model has no answer to the challenge of altering the very foundational elements of the cell in such a way—and indeed the model presented in this paper suggests that the Darwinian process simply does not have the ability to alter the hypervolume in which the organism exists.

Sexual reproduction again challenges any Darwinian account. For in producing a sexually reproducing race, many new structures and processes must be integrated. For instance, the male obviously must have the ability to produce sperm, while the female must be able to produce eggs. In most animals there are separate germ line cells which are set apart during development to become the gametes. Other cells perform supporting roles, such as the Sertoli cells in the testis of males that support and nourish the sperm cells while they develop and mature. Thus, in more complex organisms there are the male and female reproductive systems that must be present and be able to support the germ cells, produce gametes, and successfully join the gametes together in sexual reproduction. Another long-standing question relates to the "cost" of producing males. Male organisms represent the evolutionary "dead weight"—for they do not bear offspring. Their only function is to produce sperm, but to produce males incurs the cost of bearing 50% additional offspring. This cost, we are assured, is more than compensated for by the additional genetic variability which sexual reproduction allows. However, that genetic variability is another challenge to Darwinian mechanisms, since it requires encoded mechanisms for crossover and genetic exchange. The questions and mysteries of the origin of sexual reproduction are far-reaching and profound.

The origin of multicellularity presents another clear case of novelty and innovation that has not been adequately addressed by the Darwinian mechanism. Again it is clear that the mere sifting of single-celled variants simply cannot produce a multicellular

organism. Not only must multiple cells be arranged correctly in space, they must also have the requisite signaling pathways by which to communicate and coordinate their functions. In addition, there must be the ability for different cells to specialize in different functions—differentiate—and coordinate the functions of each tissue and organ to the survival and reproduction of the organism. It is here that we begin to encounter the mystery surrounding the origin of body plans, and the limitations upon morphology that are imposed by the developmental process.

In describing the importance of the developmental process in constraining morphological change, Richard Dawkins writes,

> Mutation is non-random in the sense that it can only make alterations to *existing* processes of embryonic development. It cannot conjure, out of thin air, any conceivable change that selection might favour. The variation that is available for selection is constrained by the processes of embryology, as they actually exist.[19]

The process of embryological development constrains the options that an organism may test. Therefore, the developmental process *pre-determines* the biological hypervolume by setting up the existing structure—the structure upon which mutation and selection may then operate. Just as the encoding of a computer program determines which parameters are allowed to change and how those parameters may change via an evolutionary process (and hence the dimensionality of the hypervolume), the developmental process determines precisely how each gene may be mutated to still produce a viable organism. Therefore, the embryological process entirely pre-determines what range of possibilities exist for each gene, and thence the number of dimensions in the biological hypervolume. Gene mutations act upon, and are channeled through, the developmental process to produce the adult organism. Some genes are highly constrained and even slight variation produces nonviable organisms. Some genes have little or no effect upon embryology, so mutations in these genes are of little or no consequence—they are very loosely constrained and highly variable. But the bottom line is that the hypervolume for any organism is determined by the range of possibilities allowed by the process of embryological development—not by the sheer number of possible mutations. Richard Dawkins again sheds light upon this idea, and due to the clarity of his thought it is worth quoting him at length:

> But any real-life Darwinian would acknowledge that, although any gene on any chromosome may mutate at any time, the consequences of mutation on *bodies* are severely limited by the processes of embryology. If I ever doubted this (I didn't), my doubts would have been dispelled by my biomorph computer simulations. You can't just postulate a mutation 'for' sprouting wings in the middle of the back. Wings, or anything else, can only evolve if the process of development allows them to. Nothing magically 'sprouts'. It has to be made by the process of embryonic development. Only a minority of the things that conceivably could evolve are actually permitted by the status quo of existing developmental processes. Because of the way arms develop, it is possible for mutations to increase the length of fingers and cause webs of skin to grow between them. But there may not be anything in the embryology of backs that lends itself to

> 'sprouting' angel wings. Genes can mutate till they are blue in the face, but no mammal will ever sprout wings like an angel unless mammalian embryological processes are susceptible to this kind of change.[20]

Notice what Dawkins is saying: small changes in existing developmental processes (the sorts that produce longer fingers or webbed fingers) are possible in Darwinian evolution. But larger changes, like angel-type wings sprouting from a human back, are not allowed. Why not? The answer, though Dawkins doesn't explicitly acknowledge it, is that the smaller changes are within the hypervolume defined by embryological development. However, causing wings to sprout from a person's back would require extensive re-engineering of the hypervolume through the reworking of the embryological process. Many new genes would be needed, and they would all have to be integrated into the rest of the developmental process before wings could be produced. These sorts of large-scale, coordinated changes require intentionality—which Dawkins rejects as absent from the Darwinian process.

The origin of these fundamental developmental processes presents a problem for the neo-Darwinian mechanism. D.T. Anderson points to the origin of body plans (or phyla) in the Cambrian explosion around 540 million years ago, where mystery looms large:

> In contemplating the problem presented by the evolution of the more basic organizational levels, such as the distinct body plans characteristic of phyla or classes, the gradualistic interpretation has often failed to satisfy. Special problems are posed by the sudden emergence of major new body forms in the Cambrian fossil record. The gaps are real; they can be filled if one accepts that small genetic changes can yield large morphological changes, and the opportunity is often presented to tender an imaginative proposal for a sudden evolutionary leap. Rapid evolutionary rates, in other words, are more often inferred for the big changes that for the small ones. This is an interesting paradox, because the proposed big changes actually involve far more fundamental functional redirection of developmental pathways than any of the small changes.[21]

The fundamental re-engineering of the hypervolume, the redirection of developmental pathways, is precisely the sort of change that is inexplicable and mysterious both within neo-Darwinian theory and in the fossil record. With regard to the Cambrian explosion, even Richard Dawkins noted, "It is as though [the animal phyla] were just planted there, without any evolutionary history."[22] Around 3.5 billion years ago we see the first simple photosynthetic (but non-oxygenic) prokaryotic life forms appear in the fossil record; one billion years ago we see single-celled eukaryotic life forms appear in the form of fungi and single-celled algae. This peaceful single-celled world was dramatically interrupted by the sudden invasion of multicellular organisms during the Cambrian explosion. Within a time period of only 5-10 million years, over 90% of phyla came into existence, suddenly, starkly, and fully formed. Several attempts have been made to account for the sudden appearance of the phyla, most suggesting that somehow the intermediate forms actually did exist but were not fossilized. However, the high quality of the Precambrian rock strata and the fact that we have well-preserved bacteria, algae, and even some very detailed sponge embryos from that era (for more information,

see Meyers, Nelson, and Chien, reference 23) suggest that at least some intermediate forms would certainly have been preserved had they existed.

Furthermore, not only are the intermediates lacking from the Cambrian explosion, they are also missing from the experimental data. Concerted attempts to show how one developmental process may evolve into another, by many small, accumulated mutations, have failed. No alterations in the deeply entrenched features of the developmental process have been found or created via mutation. The well-known hox gene mutants serve only to disrupt and alter existing structures or produce pre-existing structures in the wrong places (like fruit flies that develop eyes on their legs). The fact that no *new* structures or body plans have been produced by these experiments or by any other experiments to date suggests that the neo-Darwinian model simply lacks the resources to overcome this problem.

The lack of experimentally observed heritable changes in body plan coupled with the suddenness of the origin of phyla in the Cambrian explosion suggests that a new mechanism is needed to account for the data. Indeed, Campbell and Marshall, in attempting to explain the sudden origin of phyla in the fossil record, noted

> [Our] approach also has its difficulties. It provides no genetic mechanism for large-step changes, though these are said to be probably the most significant single factor regulating diversity at higher taxonomic levels. (We do not regard this as a serious criticism; rather it is a spur to explore the possible genetic mechanisms.)[24]

I propose that this missing genetic mechanism is an *intentional mechanism* that has operated at key points in the history of life and sets the boundary conditions in which the neo-Darwinian mechanism holds sway. Science has simply ignored this mechanism (often for purely metaphysical, hence nonscientific, reasons) to its own detriment; the limitations of the current paradigm are becoming painfully obvious.

Conclusion

In his book *Finding Darwin's God*, Brown University professor Ken Miller comments, "This, in essence, is how the [blind] watchmaker works, beginning with a set of basic structures and gradually modifying them to produce new functions from old parts."[25] In other words, evolution operates by making variants of existing systems, and over time this sifting of variants results in the biological wonders that fill our biosphere. This is the conventional view of neo-Darwinism: the idea that trial and error can do it all. However, the TRIZ model tells us that trial and error only operates within certain limits and cannot generate new inventions, in short, that it *cannot* do it all. The new science of genetic algorithms demonstrates precisely *why*: the very existence of trial and error processes requires certain more fundamental parameters that are themselves beyond the reach of the trial and error mechanism. These parameters must be set before trial and error can begin operate and they completely determine the area trial and error may explore and what characteristics it will select for. If new inventions are to be had, the old parameters must be superseded by new parameters. This is fundamentally an intentionality-laden process, for it requires that the new parameters be carefully

formulated and tuned to the desired end. Furthermore, it cannot be done gradually—the new parameters define the new system and must be present in their entirety before the new system even exists. Indeed, the invention comes in a sudden burst of inspiration; a gestalt shift in the mind of the inventor in which many ideas come together and a new technology is birthed. Biological history is replete with grand inventions and moments of sheer creativity unparalleled by any human invention. The origin of life, the origin of eukaryotes, the origin of sexual reproduction, and the origin of multicellular organisms and body plans all present cases in which major re-engineering of the parameters of the system occurred. In all of them a sudden jump was observed, and all of them lack any sort of detailed explanation or understanding in the current Darwinian model. The reason is simple: the Darwinist is trying to explain these events by invoking the *wrong mechanism*. The TRIZ model presents two separate mechanisms that operate upon two types of problems: the trial and error mechanism, and the inventive or intentional mechanism. The misapplication of the trial and error mechanism to explain inventive changes in biology has stunted scientific progress, and it is increasingly obvious that there are major shortcomings in our ability to explain the most important events in the history of life. The time has come for a reevaluation of the sorts of mechanisms permitted into biological science. The data is clearly showing that a new mechanism is required, a mechanism that is complementary to the Darwinian mechanism—and is characterized by the presence of intentionality.

The future of this new science, inclusive of intentional agency, is bright. It is hardly an exaggeration to say that the opportunities and insights to be gained are beyond compare and touch upon nearly every important question in biology, where finding real answers to some of these long-unanswered questions promises to absolutely revolutionize the science of the new millennia. New research questions, such as "how would an intentional agent have done this?" and "how do we separate the effects of intentionality from trial and error?" show that science will be far fuller and richer with this new framework. The concept of intentional agency is fertile new ground that even now is germinating the seeds of scientific progress.

As science moves forward, it evolves much as technologies do. New models and paradigms, corresponding to new hypervolumes of possibility, come into existence and go extinct. Science has currently reached a critical point, where one particular contradiction is becoming increasingly acute and its resolution requires a new model. The contradiction is between the neo-Darwinian model and the scientific data regarding key moments in the history of life; the new model includes the actions of intentional agency as a fundamental, non-reducible causal force. As we adopt this innovative new way of doing science, we will overcome the contradiction between the neo-Darwinian model and the scientific data—and awaken to a new dawn in our understanding of the universe and our place in it.

References

[1] Altschuller G. The Innovation Algorithm. Worcester: Technical Innovation Center, Inc.; 1999. p 37.

[2] Ibid., p 33.

[3] Savransky S. Engineering of Creativity. New York: CRC; 2000. p 8.

[4] Altschuller, p 89-99.

[5] Ibid., p 171-181.

[6] Savransky, p 60.

[7] Solomon J. Learning to be Inventive: Design, Evaluation and Selection in Primary School Technology. In: Ziman J, editor. Technological Innovation as an Evolutionary Process. Cambridge: Cambridge University Pr.; 2000. p 201.

[8] Dawkins R. The Blind Watchmaker. New York: Norton; 1987. p 43-74.

[9] Ibid., p 67.

[10] See chapter 2 of William Dembski's forthcoming book, *No Free Lunch* (published by Rowman & Littlefield).

[11] See http://www.phy.syr.edu/courses/mirror/biomorph/ (last accessed 3 Nov 2001)

[12] See http://www.lmb.uni-muenchen.de/groups/famulok/SELEX.html (last accessed 3 Nov 2001)

[13] See chapter 4 of William Dembski's forthcoming book, *No Free Lunch* (published by Rowman & Littlefield).

[14] See http://www.cognitivetherapy.com/neural_checkers.html (last accessed 22 Nov 2001)

[15] Mitchell M. An Introduction to Genetic Algorithms. Cambridge: MIT Pr.; 1998. p158.

[16] Miller G. Technological Evolution as Self-Fulfilling Prophecy. In: Ziman J, editor. p 208.

[17] Schneider T. Evolution of Biological Information. Nucleic Acids Research 2000;28(14):2794-2799

[18] Altschuller G. And Suddenly the Inventor Appeared. Worchester: Technical Innovation Center, Inc; 1996. p 55-6.

[19] Dawkins, p 312.

[20] Ibid., p 311.

[21] Anderson D. Developmental Pathways and Evolutionary Rates. In: Campbell K and Day M, editors. London: Allen & Unwin; 1987. p 146

[22] Dawkins, p 229.

[23] Meyer S, Nelson P, Chien P. The Cambrian Explosion: Biology's Big Bang. http://www.discovery.org/articleFiles/PDFs/Cambrian.pdf (last accessed 8/13/01).

[24] Campbell K, Marshall C. Rates of Evolution Among Palaeozoic Echinoderms. In: Campbell K and Day M, editors. p 97.

[25] Miller K. Finding Darwin's God. New York: HarperCollins; 1999. p 139.