

# Intelligent Design or Evolution?

Why the Origin of Life and the Evolution  
of Molecular Knowledge Imply Design



Copyright Information:

Library of Congress Control Number 2005903070

© 2005

Intelligent Design Books

[www.lifeforigin.com](http://www.lifeforigin.com)

## Table of Contents:

<b>Introduction: Evolution vs. Design.....</b>	<b>1</b>
The Philosophy of Science.....	1
The Molecular Theory of Evolution.....	2
Is Evolution Possible?.....	4
The Evolution of Molecular Knowledge.....	5
Chemical Evolution.....	8
The Origin of Life.....	10

## Part 1: The Evolution of Knowledge and Information

<b>Chapter 1: Information vs. Knowledge.....</b>	<b>17</b>
The Nature of Evolution.....	19
Trapped Scientist with Three Coins.....	20
Information Is Closely Related to Probability.....	22
Mathematical Definition of Information.....	24
Trapped Scientist Who Cannot See the Results.....	25
<b>Chapter 2: The Evolution of Molecular Knowledge.....</b>	<b>28</b>
The Trapped Scientist.....	28
Information, Knowledge, Complexity and Order.....	31
A Scientist Locked in a Room with Multiple Doors.....	32
Important Definitions.....	36
Probability and Information.....	37
Information in Biological Systems.....	38
<b>Chapter 3: Information Storage and Transfer in Life.....</b>	<b>40</b>
Transcription and Translation.....	40
Information Using Coins.....	41
A Four Sided Coin.....	42
The Four Sided Coin Code.....	44
The Information in DNA is Similar to a Four Sided Coin.....	46
DNA Replication.....	47
The Genetic Code.....	49
Transcription.....	51
Each Codon Specifies an Amino Acid.....	52
Translation.....	53
Proteins.....	57
Protein Folding.....	58
Eukaryotes and Prokaryotes.....	60
The Accurate Trapped Scientist.....	61

## **Chapter 4: Information & Knowledge in the Protein Insulin...63**

Determining Allowed Amino Acids.....	64
How Much Information Opens the Door.....	65
Why Does this Work.....	68
Total Information in Insulin A and B Chains.....	70
Molecular Knowledge of Insulin.....	74
Cartoon and Space Fill Models of Insulin.....	78
The Probability of Insulin Evolving.....	79
Insulin May Not Imply Design.....	80
How Accurate is the Technique?.....	82

## **Part 2: Chemical Evolution**

### **Chapter 5: Information & Knowledge before the Genetic Code..88**

Meteorites.....	90
The Evolution of Primordial Knowledge.....	93
How Many Solutions?.....	93
Molecular Knowledge Before Life.....	95
The Composition of the Primordial Soup.....	95
The Evolution of a Functional Protein in the Primordial Soup.....	98
Molecular Knowledge in the Primordial Soup.....	101

### **Chapter 6: Introduction to Chemistry and Entropy.....104**

Chemicals and Atoms.....	104
The Hydrogen Atom.....	105
Representing Chemicals with Symbols.....	106
Chemical Bonds.....	108
Multiple Bonds.....	109
Chemical Symbols.....	109
Matter, Energy, Heat and Temperature.....	111
Quantum Mechanics.....	112
Micro-states and Entropy.....	113
The Second Law of Thermodynamics.....	116
Heat Flows from Hot Objects to Cold Ones.....	118
Entropy and Chemical Reactions.....	121
Chemical Kinetics.....	122
Chemical Equilibria.....	124
Closed vs. Open Systems.....	127

<b>Chapter 7: Implications of the Second Law.....</b>	<b>129</b>
How Does Life Exist So Far from Equilibrium?.....	129
Is Life Really Different from Non-Life?.....	135
Do Energy Sources really Help?.....	136
Non-Equilibrium Thermodynamics.....	137
Chemical Oscillators.....	139
Entropy and Biological Evolution.....	140
<b>Chapter 8: The Structure of DNA, RNA, and Proteins.....</b>	<b>142</b>
DNA Structure.....	142
RNA Structure.....	144
Proteins.....	147
Organic Chemistry Functional Groups.....	147
Structure of Amino Acids.....	148
<b>Chapter 9: Prebiotic Synthesis of RNA, DNA and Peptides.....</b>	<b>156</b>
Zero Tries.....	159
Investigator Interference.....	161
Protein Synthesis.....	162
Miller's Experiment.....	162
Thermal Proteins or Proteinoids.....	164
Short Peptide Chains in Water.....	165
Long Peptide Chains in Water.....	166
RNA Synthesis.....	168
RNA Building Block Synthesis.....	169
Adenine and Cytosine.....	169
Ribose.....	170
A Pre-RNA World?.....	171
Assembling the Building Blocks.....	171
Activated Monomers.....	173
Review of Investigator Interference.....	174
Interference Strategy #1: Eliminate the Undesirable Chemicals.....	174
Interference Strategy #2: Concentrating Volatile Chemicals.....	175
Interference Strategy #3: The Use of Condensation Agents or Activated Monomers.....	175
Interference Strategy #4: Controlling the Energy Sources.....	176
Interference Strategy #5: Substituting Human Knowledge.....	176
Conclusion.....	177

<b>Chapter 10: Self Replicating Molecules and Systems.....</b>	<b>180</b>
A Self Replicating Peptide.....	181
RNA Self Replication.....	184
How Much Knowledge is Required to Create a Ribozyme.....	187
Molecular Knowledge in the Primordial Soup.....	188
Self Replication and Perpetual Motion.....	191
<b>Chapter 11: The Myth of the Primordial Soup.....</b>	<b>196</b>

### **Part 3: The Evolution of the First Genes**

<b>Chapter 12: Irreducible Complexity.....</b>	<b>200</b>
<b>Chapter 13: Nucleic Acid Synthesis: Adenine.....</b>	<b>203</b>
Nucleic Acids are Critical to Life.....	203
Adenine Synthesis in Life.....	207
Preliminary Calculation of the Information and Knowledge.....	208
3-D Structure of Several Key Enzymes.....	209
Calculation of the Knowledge and Information.....	212
Did Life Arise All at Once?.....	214
<b>Chapter 14: ATP Synthesis.....</b>	<b>217</b>
Oxidation Releases Energy.....	220
ATP Created by Proton Gradients.....	222
ATP Synthesis with a Single Enzyme.....	223
G3PD Knowledge.....	224
Molecular Knowledge in G3PD.....	226
Correction for Primordial Information.....	230
Odds of ATP and Adenine Synthesis Evolving Concurrently.....	231
Does this Approach Generate False Knowledge.....	233
ATP Synthase.....	234
Conclusion.....	237

<b>Chapter 15: How Many Tries.....</b>	<b>238</b>
How Does Time Factor Into the Equation.....	238
How Fast Does the Tree Grow?.....	241
How many Stars?.....	244
How Many Primordial Soups?.....	244
RNA Self-Replication.....	245
Protein Evolution.....	245
Upper Limit on the Number of Tries.....	246
Constraints on the First Self Replicating Molecule.....	246
The Trees Help, but . . . ..	247
Natural Selection Limits the Number of Tries.....	248
This Simple Example Shows that Evolution Does not Work Quite Like Darwin Imagined.....	250
Natural Selection Reduces the Number of Tries.....	251
Implications for RNA Self-Replicating Molecules.....	252
 <b>Chapter 16: Evolution Since the Cambrian Explosion.....</b>	 <b>254</b>
 <b>Chapter 17: Alternatives to Intelligent Design.....</b>	 <b>256</b>
Science Without the Observable Axiom.....	257
Consequences of the Observable Axiom.....	258
 <b>Appendixes:</b>	
Appendix 1: Shannon Entropy and Information.....	259
Appendix 2: Relative Entropy and Information.....	261
Appendix 3: Math Review.....	265
Appendix 4: Review of Yockey’s Approach.....	268
Appendix 5: Review of Schneider’s Approach.....	272
Appendix 6: Experimental Support for Molecular Knowledge .....	275
Glossary.....	280
World Wide Web Resources.....	285
Index.....	286

## **Preface**

---

When Darwin proposed the theory of evolution, he envisioned that the process was continuous. He believed that the complexity and diversity of life could be explained by numerous slight modifications to existing species. His whole theory hinges on the idea that nature will preserve beneficial variations (natural selection) and thereby guide evolution creating new and more complex animals and plants.

“If it could ever be demonstrated that any complex organ existed which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down.” - Charles Darwin

Over a century later, science has yet to show that complex organs can be formed by numerous, successive, slight modifications, and many scientists do not think that this is important. To understand why, it is first necessary to examine the assumptions on which science is based. Science is based on two assumptions called axioms.

- Observable axiom: scientists can accurately observe reality, and then propose theories and laws to explain their observations.
- Naturalistic axiom: everything can be explained by the laws of physics, chemistry, biology and mathematics.

Axioms are self-evident assumptions. Without the observable axiom, science cannot function properly. If scientists cannot observe reality or if they observe it incorrectly, then all scientific theories are suspect. Fortunately, this axiom is self-evident to most.

The naturalistic axiom is different in two respects: 1) it is not self-evident to 90% of the population 2) science can function properly without it because unlike the observable axiom the naturalistic axiom can be tested.

Today science does not understand how life originated. The better science understands the nature of the problems associated with the origin of life the more mysterious life's origin becomes. Today, every theory concerning the origin of life suffers from the same problem. The probability of life originating is so small that the extreme age and size of the universe cannot offset the poor odds. The most popular theory concerning the origin of life proposes that life arose as a perpetual motion machine (see chapter 10). Yet physics clearly states that perpetual motion cannot exist.

Joyce and Orgel describe the situation best with the following quote: “ It must be said that the details of this process remain obscure and are not likely to be known in the near future.”

- The RNA World, p72-73.

This dilemma places science in an awkward position. The naturalistic axiom forces science to only consider origin of life theories that do not require a supernatural explanation. Yet these theories continue to fall short and do not explain how life originated. Because evolutionists are unwilling to abandon the naturalistic axiom, they must make one final assumption. They assume that one day science will unravel the mystery of life's origin. Unfortunately, many evolutionists are unaware that they have made this assumption. Thus, they continue to claim that evolution is a proven fact when in reality they have only assumed that it is true.

This third assumption obviously has far reaching consequences because with this assumption, evolution cannot be disproved. Any observation that evolution does not explain is simply ignored. This places the theory of evolution on a pedestal. The theory cannot be disproved because science has already made the assumption that it is true. Furthermore, this assumption also explains why scientists are in no hurry to apply Darwin's proposed test to evolution. Most scientists simply do not see the need to prove that a complex organ can form through a series of numerous, slight, continuous modifications. They have already assumed that evolution explains the origin of complex organs. Why test something that must be true?

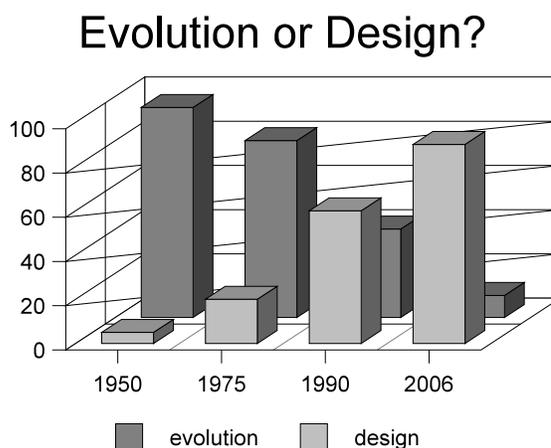
## Intelligent Design

Intelligent design is based on the observable axiom. The theory does not assume that the naturalistic axiom is true, but rather attempts to verify the axiom, and it accomplishes this with indirect logic.

Consider the problems with life's origin. The probability of a living system evolving on the earth or anywhere in the universe can be calculated, and if this probability is so small that the vast size and age of the universe cannot overcome the poor odds, then design should be inferred. The goal of this book is to perform the above calculations and to show that the design inference is justified.

Unlike evolution, intelligent design can be disproved. Suppose tomorrow that one scientist unravels the mystery of life's origin, and that another demonstrates that a complex organ can be explained by the laws of physics, chemistry, biology and mathematics. When and if this happens, the theory of intelligent design will be disproved.

Intelligent design like science requires one final assumption. The theory assumes that science will not solve the mystery of life's origin. In the 1950s, this would have been a very dangerous assumption. Most scientists would have argued that they were just beginning to understand the nature of the problem and that they had not had time to solve it. Fifty years later, such arguments are weak. Not only has science had ample time, but the solutions have become more elusive with time. As the years pass, the probability of science finding a solution will continue to fall, and the design inference will continue to become stronger as indicated by the height of the bars in the figure to the right.



## **Intelligent Design and Creationism**

Creationism is a broad term that can mean many different things. Many creationists believe that the earth is only 5000 years old and that it was created in six 24 hour days. Others believe that the earth is 5 billion years old and that life was created slowly as time passed or in spurts separated by large expanses of time. Creationism is different from intelligent design because it is based on the existence axiom.

- Existence axiom: God exists.

This axiom separates intelligent design from creationism in an important way. If science explains the mystery of life's origin, most people will still believe in God. So unlike intelligent design, creationism cannot be disproved.

## **How is this Book Different**

The design inference in this book is only drawn from the origin of life and the evolution of the very first genes and proteins. Others have attempted to infer design from the evolution of new classes of animals. This book stays away from this practice because the inference is weak. While evolutionists assume that naturalistic laws can explain how a bird evolved from a reptile, many intelligent design advocates seem to just assume the opposite. Neither side has introduced accurate mathematical models or computer simulations to describe such an evolutionary transition. Life is definitely very good at evolving. It has developed techniques that create new information by re-arranging and shuffling existing information. Furthermore, the DNA of a bird is not that different from that of a reptile. These factors weaken the design inference.

In my opinion, it is illogical to invoke design for the origin of life and then assume that the designer played no part in the rest. Thus, this book will focus on the origin of life and the evolution of the first genes and proteins. These two events clearly imply design.

## **Introduction: Evolution vs. Design**

---

### **The Philosophy of Science**

By definition, science must explain everything in terms of naturalistic laws. So before analyzing any data, science rules out the possibility that life was created. While this philosophy has served science well, it is somewhat problematic if a creator exists. Intelligent design differs from science because it does not use assumptions to eliminate possibilities. Instead intelligent design allows the evidence to lead where it may.

Scientific experiments test how evolution happens and sometimes experiments are designed to test if evolution can happen. Science never asks why it happens. The naturalistic axiom does not allow science to ask why. Science assumes that naturalistic laws are responsible.

Scientists often abuse the naturalistic axiom by assuming that events that happen in the distant past are possible. For example, the evolutionary transitions between a simple self-replicating molecule and a simple cell undoubtedly involved the emergence of several hundred new genes and proteins. Rather than analyze the probabilities associated with the evolution of these new genes, science simply assumes chance and natural selection were responsible, and it invokes the naturalistic axiom for justification. The logic seems to be as follows - *well it happened, so it must have been possible*. This philosophy does not allow science much freedom. No matter what the evidence shows, the assumptions on which science is based ensure that it will always support the theory of evolution. Science is trapped. It cannot disprove the theory of evolution without disproving one of its most fundamental axioms.

The failure of science to ask if evolution is possible has led to the premature acceptance of many ideas concerning evolution. For example, when Darwin proposed the theory of evolution in 1859, the chemistry of living organisms was a complete mystery, so the testing of Darwin's theory was limited. For nearly 100 years following Darwin's first publication, science only had three ways to test his theory: 1) search for the fossils that link existing animals and plants, 2) design experiments to observe how animals and plants change ever so slightly from one generation to the next and 3) accumulate evidence that the earth is very old. While such experiments were critical for evolution's acceptance, the chemical processes behind evolution remained a mystery. Science just assumed that naturalistic laws were responsible; thus, science was able to embrace the theory of evolution as a proven fact without really understanding it.

### **The Molecular Theory of Evolution**

In 1953, scientists began to unravel the chemistry of life when Watson and Crick proposed a model for DNA. Soon thereafter the genetic code was broken, and the chemical mechanism behind evolution became clear. The hypothesis put forth is outlined below:

Sections of DNA called genes store the information needed to make proteins, and this information is passed from one generation to the next when genes are replicated during reproduction. The replication process is not perfect, and as such it may by chance introduce errors. Errors during replication, mutations, have the potential to create new genes. Mutations may create new information, or they may simply alter existing information. In either case, nature preserves beneficial mutations through the process of natural selection and other mutations survive by chance. Over many millions of years, changes in existing genes yield new genes; therefore, animals continually evolve and adapt.

Soon after its proposal, this hypothesis became the framework for the theory of molecular evolution. While scientists have modified it over the years, the basic framework of the theory remains intact with one important exception.

If an existing gene evolves into a new gene with a new function, then the original function will be lost, and natural selection will not allow this to happen. So Ohno suggested that existing genes do not evolve into new genes unless they are first duplicated.<sup>10</sup> The duplicate copy is free to evolve a new function while the original maintains its current function. Others have refined the theory further by suggesting that pieces of existing genes may be duplicated and then rearranged to create new genes with new functions. With these modifications, the molecular theory certainly explains the origin of many genes.

But even with these improvements, the concern raised earlier remains the same - why not ask is evolution possible? Science describes how it happens, but why not take the next step and investigate the probabilities associated with the required events. That is rather than assume that naturalistic laws are responsible, prove that these laws are responsible. This avoids the trap. Thus, more experiments are needed to test whether or not evolution is possible.

A ten-year experiment can hardly hope to model a billion years of evolution, but today there is a solution to this problem. Scientists around the world are actively sequencing the DNA of many animals, plants and bacteria, and after more than three decades of characterization, this information is freely available in online databases. These databases allow science to ask for the very first time two important questions. Can mutations operating over billions of years and guided by natural selection create new genes? And perhaps more importantly, are naturalistic laws responsible?

## Is Evolution Possible?

To test if evolution is possible, consider a gene that is common to all living things. That is the gene is found in algae, bacteria, oak trees, carrots, mice, fish, people and all other living things. This gene will not be identical in all living things, but it will be similar. The form of the gene found in bacteria has had at least 2 and probably 3 billion years to change independently from the same gene found in animals and plants. The form of the gene in an oak tree has had at least 1 billion years to evolve independently from the gene in man. This gene and many others like it allow science to conduct experiments that look back in time - almost to the origin of life itself.

A comparison of this ancient gene in many different species will reveal how much information the gene contains, and from such an analysis, one can calculate the probability associated with the gene's evolution. If the probability so calculated is for all practical purposes zero, then design may be inferred. Unfortunately, this inference is not compelling because it fails to consider the effect of natural selection. In order to make the argument for design stronger, the concept of information must be replaced with another familiar concept, knowledge. The two definitions that follow are important.

*Molecular information* is the information found in a gene today. It is calculated by comparing the differences found in the same gene in many different animals, plants and bacteria. Information has a precise mathematical definition as defined by information theory.

*Molecular Knowledge* is the minimum amount of useful information required by a gene to have any function. Any region of DNA that does not contain molecular knowledge has no function at all. Thus, such a region cannot be preserved or optimized by natural selection, and it cannot be classified as a gene.

Molecular knowledge is always less than molecular information. Molecular knowledge is more difficult to calculate. Nevertheless, it is possible to bound molecular knowledge and perform experiments to determine its exact value (see appendix 6).

## **The Evolution of Molecular Knowledge**

Because of natural selection, information cannot be used to calculate the probability that a gene will evolve. Information is useful because it has a precise mathematical definition not because it can answer questions concerning whether or not the evolution of a new gene is possible. Chance is not in control if natural selection is guiding which mutations survive. Therefore, relating the amount of information in a gene to a probability that it can evolve is not a valid mathematical analysis.

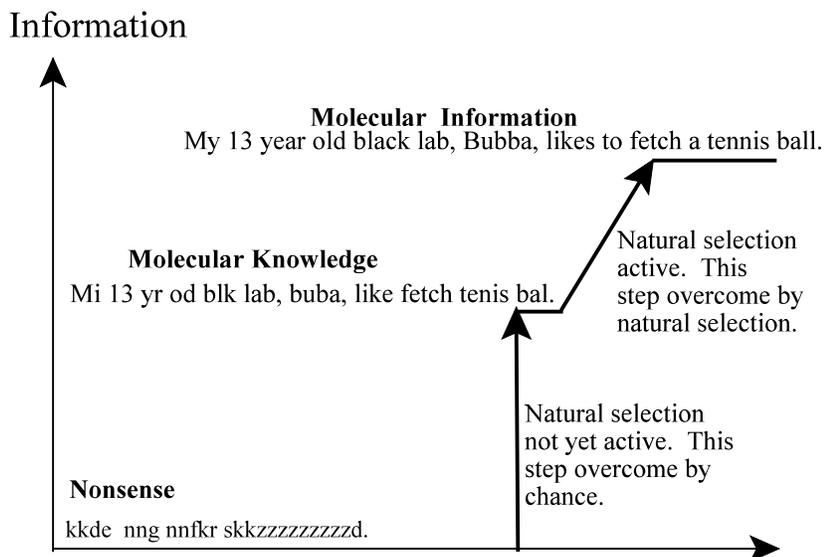
Molecular knowledge is the minimum amount of useful information required for a gene to have any function. If a gene does not contain molecular knowledge, then it has no function, it confers no selective advantage, and it is not a gene. Thus, before a region of DNA contains the requisite molecular knowledge, natural selection plays no role in guiding its evolution. Chance controls which mutations survive. Thus, molecular knowledge can be related to a probability of evolution. Figure 1 helps illustrate these important concepts.

Notice that the first step that creates the required molecular knowledge is vertical, and the subsequent step that creates molecular information is sloped. This difference is important in that it is meant to show that natural selection can help guide the second transition, but it plays no role in the first. Thus, it is the size of the first step that determines whether or not a gene can evolve.

A simple example will now be introduced to help clarify this concept. Consider the following sentences:

- My 13 year old black lab, Bubba, likes to fetch a tennis ball.
- mi 13 yr od blk lab, buba, like fetch tenis bal.
- kkde nng nnfkr skkzzzzzzzzzd.

Figure 1: Information in Life



Each sentence represents a gene. The first sentence uses the most letters. The second sentence uses fewer letters, communicates the same points and is not grammatically correct. The information in the first sentence is greater than the second (it has more letters), but the useful information or knowledge is the same (the same points are communicated). The last sentence is nonsense. It actually contains information, but this information is not useful. If a person is asked to perform the job of natural selection by optimizing each of the above sentences, then it is easy to see that they might leave the first sentence alone and change the second sentence to read more like the first. They will not change the last sentence into anything remotely similar to the first. In fact, they will not be able to optimize this sentence because they do not know what it should say. If they do happen to change it into a sentence similar to the first, then they do so by chance. By analogy, natural selection cannot guide the evolution of genes that contain no useful information.

If these sentences are composed by randomly selecting letters from the alphabet, then the probability of spelling the second sentence is

much better than the first. So only this sentence and all sentences similar to it are useful for assigning a probability to the evolution of this particular sentence. In this example, the first sentence represents molecular information and the second one represents molecular knowledge. If some concepts are not required, the second sentence may be simplified further. For example, the sentence fragment, mi black lab, still communicates some knowledge.

At least one biologist, Richard Dawkins, has suggested that sentences like: “kkde fwegzzzzzzzzd.” still confer a selective advantage; therefore, given time these sentences will evolve into something useful under the guidance of natural selection.<sup>1,14</sup> Dawkins ran many computer simulations with sentences like the one above, and they all evolved into the desired result. But his programming and logic are both flawed because natural selection cannot preserve or optimize a gene that offers no selective advantage (the nonsense sentence above represents a gene that confers no selective advantage). A gene must contain some useful information before natural selection activates. Thus, chance and chance alone must create the initial knowledge. Because the probability that chance can accomplish this goal is proportional to the height of the first step in figure 1, the size of this step completely determines whether or not a new gene will evolve. It is the goal of this book to characterize the size of this initial step. If it is small then naturalistic laws explain the evolution of knowledge. On the other hand, as the step size increases, the probability that chance will create the required molecular knowledge approaches zero, and at some critical threshold, the design inference becomes valid.

Large steps are associated with the evolution of genes that are completely different from all other existing genes. For these genes, the probability of chance finding an appropriate solution (even given 50,000 billion years) is very close to zero. The origin of these genes imply design.

## **Chemical Evolution**

This book will also evaluate another hypothesis put forth in the 19<sup>th</sup> century. This hypothesis attempts to explain the origin of life and its basic premise is as follows:

The early earth's atmosphere was different from today in that no free oxygen was available. Under these circumstances, energy sources like sunlight and electrostatic discharges might create the chemicals necessary for life (chemical evolution). As these chemicals were concentrated in a small pond or puddle, the primordial soup, they organized themselves in such a way to form the first living organism. Because life is very complex, the first living thing is usually assumed to be a self-replicating chemical rather than a living cell. Because the first living thing was able to replicate itself, it evolved into life as it exists today.

This hypothesis or some form of it is found in almost all biology books where it is put forth as the generally accepted theory. Yet in the scientific journals, scientists routinely dismiss many aspects of the hypothesis as highly improbable (Shapiro 1995 and 1999; Miller 1995 and 1998; Joyce 1984 and 1989; Nissenbaum 1975; Ferris 1987; Joyce and Orgel 1999; Thaxton:1984). When it comes to chemical evolution and the origin of life, science just does not have the answer.

One of the first experiments concerning the origin of life was conducted in 1953 by Stanley Miller. Miller created several amino acids (the building blocks that life uses to make proteins) in an electrostatic discharge chamber. The experiments conducted since Miller have demonstrated how difficult it is to create the biological precursors required for life. While several amino acids can be created under plausible conditions, proteins cannot be. Furthermore, DNA is much more problematic because its building blocks are difficult to create. Many of these building blocks are unstable and decay rapidly. Science has yet to offer a plausible explanation for how these hard to make and easy to destroy chemicals accumulated in the primordial soup (see references 3,4,5, 6, 7, 8 and 13 on page 15, and chapter 9).

The most prevalent myth concerning chemical evolution suggests that a continuous flow of energy through a complex system of nonliving chemicals will promote the formation of biologically relevant molecules. The researchers who hold to these views suggest that life arose spontaneously when these biological precursors combined in a small pond or puddle several billion years ago. While such energy flows are critical to the survival of life today, it is not clear how they solve the mystery of life's origin. Life knows how to use these energy flows to do work. Such knowledge is completely lacking from a system that only contains nonliving chemicals. Plentiful energy sources if anything do more harm than good. Sunlight bombarding a small pond on the earth 4 billions years ago is much more likely to destroy any useful biological molecules than create one (see for example Fox, Molecular Evolution and the Origin of Life, p37).

Surprisingly, such difficulties are often overlooked; as a result, many biologists mistakenly believe that it is quite easy to synthesize all of the required biological molecules. Nevertheless, a quick review of the relevant literature reveals that this is not true. For example, to synthesize adenine (one of the most important chemicals found in DNA and RNA), chemists start with a concentrated solution of hydrogen cyanide and ammonia. Concentrating ammonia is not an easy task since it is a gas that boils at sub-freezing temperatures, and it also decays rapidly in the presence of sunlight. Furthermore, concentrating hydrogen cyanide in the presence of water is impossible because it reacts with water quite readily yielding formic acid. Scientists tend to focus on the fact that adenine can be synthesized in a laboratory and ignore the fact that the conditions required for its synthesis did not exist on the primitive earth.<sup>3,4,6</sup>

After 50 years of investigation no plausible prebiotic path exists to synthesize cytosine, ribose or deoxyribose (three critical subunits of DNA and RNA). The problems with ribose and cytosine synthesis are so severe that Miller and several others have suggested that the first self replicating molecule probably contained neither.<sup>7,8</sup>

Biological molecules may contain thousands of subunits all linked together by chemical bonds. Coercing the subunits to form a large biological molecule like DNA or RNA is not easy. These problems are often discussed in scientific journals like Nature, Science, PNAS, and the Journal of Molecular Evolution. For example, even today, investigators have yet to identify a plausible prebiotic method to link cytosine, thymine or uracil to ribose (a step necessary for DNA and RNA synthesis).<sup>11</sup> Nevertheless, not finding the answer is not news. So only the scientists who read these journals are aware of the difficulties involved.

Finally, the greatest challenge to the origin of life lies not with creating the chemical precursors, but instead with creating the required knowledge. The chemicals that make up life contain useful information, and it is this knowledge that allows life to propagate. The implication is that even if a few of the biological precursors required for life existed in the primordial soup, such precursors would not contain the knowledge necessary to live and evolve.

Joyce and Orgel sum up the situation best “After dreaming of self-replicating ribozymes emerging from pools of random polynucleotides, and having nightmares about the difficulties that must have been overcome for RNA replication to occur in a realistic prebiotic soup, we awaken to the cold light of day . . . It must be said that the details of this process remain obscure and are not likely to be known in the near future.” - The RNA World, p72-73.

## **The Origin of Life**

Before trying to understand the hurdles associated with the origin of life, it is useful to define life. In its simplest terms, life is a group of chemicals that possess molecular knowledge. The word *knowledge* implies that the information possessed by the chemicals is useful unlike information which may or may not be useful. The word *molecular* indicates that the knowledge resides in a chemical molecule instead of in a book or some other source.

It is this molecular knowledge that allows the chemicals in life to maintain a state that is very different from nonliving chemicals like vinegar, ammonia, and water. The molecular knowledge that life possesses is both procedural and conditional. Procedural knowledge is knowledge about how to do something. For example, how to extract energy from a sugar molecule and use it to build something else. Conditional knowledge is knowledge about why and when something needs to be done. For example, when there is no sugar present certain metabolic pathways should be turned off. Conditional knowledge in molecules is similar to that found in computer programs. A computer program may execute one command if a certain condition is true and another command if the condition is false. Computers do not think. The decisions are predetermined by the logic used in the computer's code.

It is now possible to develop a concise and accurate definition for life: *Life is a system of chemicals possessing molecular knowledge and a mechanism to implement this knowledge in such a way that the system can survive long enough to replicate itself.*

Today, life requires several chemicals to survive, grow, and reproduce. Two chemicals, DNA and RNA, store the required knowledge. Proteins and to lesser extent RNA implement this knowledge, and a third chemical, ATP, provides the energy to power the implementation. At a minimum, the simplest living system must be able to perform four critical functions:

- Store molecular knowledge.
- Implement this knowledge.
- Tap a plentiful energy source to power the implementation.
- Synthesize any biological molecules required for replication that are not plentiful in the primordial soup.

Herein, lies the mystery behind life's origin. The origin of life is a classic example of the chicken or the egg paradox because none of the critical functions listed above can exist without the others.

Many investigators have tried to overcome the paradox by suggesting that the first living thing was a single chemical that contained both the knowledge and the ability to implement the knowledge. RNA is a natural choice for the first living chemical because it can both store and implement knowledge. Nevertheless, after 25 years of experiments, the RNA hypothesis has yet to live up to its expectations. RNA has quite a bit of trouble with self replication (see Joyce:1989 and chapter 10).

Investigators have for the most part over looked the third critical function required for life, the need to tap an energy source to drive replication. Without this function, self replicating molecules become a special type of perpetual motion machine. A perpetual motion machine is a machine that runs forever with no energy input. Perpetual motion machines do not exist. They may run for a short time, but without a continuous input of energy, they eventually stop. Furthermore, all machines must know how to tap an energy source. A car with an empty gas tank cannot be driven to the gas station just because the sun is out. The sun provides an almost unlimited source of energy, but a gas engine does not know how to convert this energy into work. The same constraints apply to a self replicating RNA molecule. Unless such a molecule knows how to tap a plentiful energy source to drive its own replication, it can only exist in text books and in the imagination of researchers.

To summarize, life requires some minimal molecular knowledge to replicate. This knowledge can be possessed by a single chemical, or it can be spread out among many. In either case, the system must possess the knowledge to replicate, a way to implement this knowledge, and a way to power the implementation. A system that does not possess all three is not a living system. Furthermore, any system that is unable to synthesize the chemicals that are required for replication is not robust. These systems cannot evolve because they cannot self replicate.

Figure 2 depicts the focus of this book. The following chapters will concentrate on the genes and proteins that were required for the origin of life and on the chemicals that gave rise to these first genes and proteins. These are events that happened more than 2 billion years ago.

If a new gene evolves early in life's history, and it is completely different from any other existing gene, then the possibility that it arose by gene duplication can be eliminated. This makes the analysis much more manageable. Furthermore, the techniques used in the book simply will not work to prove that man did not evolve from apes. The DNA in a chimpanzee is almost identical to that of man's DNA. This similarity makes it difficult to infer design.

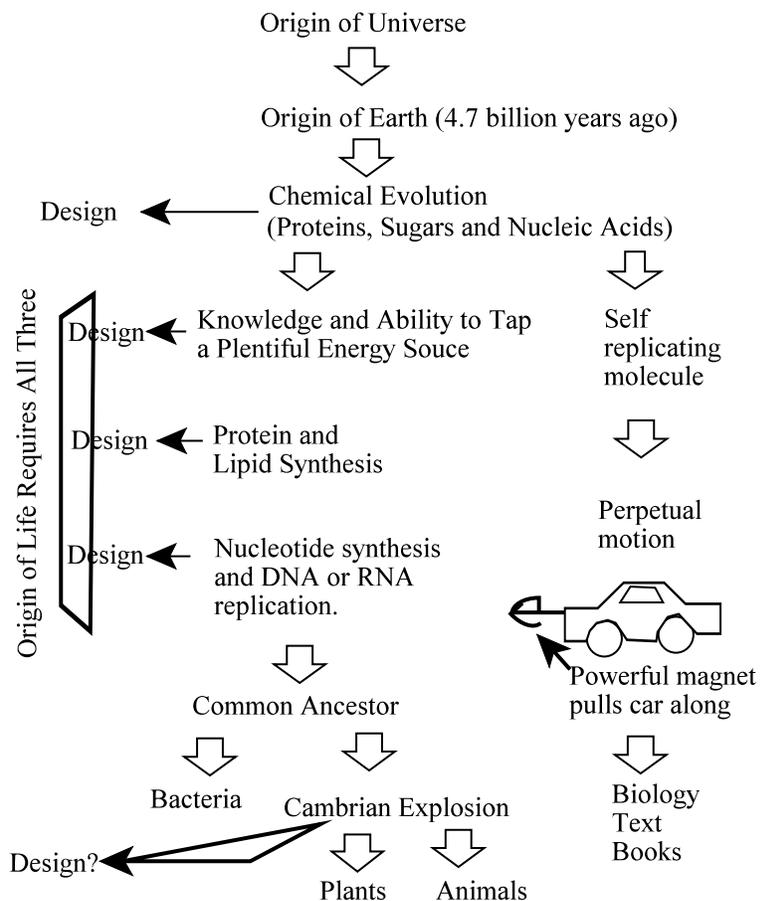
In figure 2, the self replicating molecule leads to a perpetual motion machine. The car in this figure that is pulled along by the powerful magnet is just one example of such a machine. Since perpetual motion machines do not exist, this pathway is not a very promising solution to the mystery of life's origin. It is far more likely that life arose all at once.

The difficulties associated with chemical evolution suggest that the biological precursors necessary for life would have been scarce if they existed at all, and this scarcity suggests that the first living thing was able to synthesize all of the chemicals that it needed for replication and drive this synthesis with a plentiful energy source. Today, life can only tap plentiful energy sources with the help of proteins and lipids, and this suggests that the first living thing was probably also able to synthesize proteins and lipids. Therefore, the first living thing was probably not a simple self replicating chemical, but rather a living cell very similar to life as it exists today.

While the idea that life arose all at once is not a popular one as it is contrary to Darwinian evolution, the evidence suggests that it did.

Notice in figure 2 that the Cambrian explosion may also imply design (Meyer: 2005). The fossil record indicates that almost every major biological classification (phylum) arose in a very short time span about 500 million years ago. The question mark is meant to show that any design inference based on the Cambrian explosion is subjective because no scientist has yet to accurately model the probabilities of such an event.

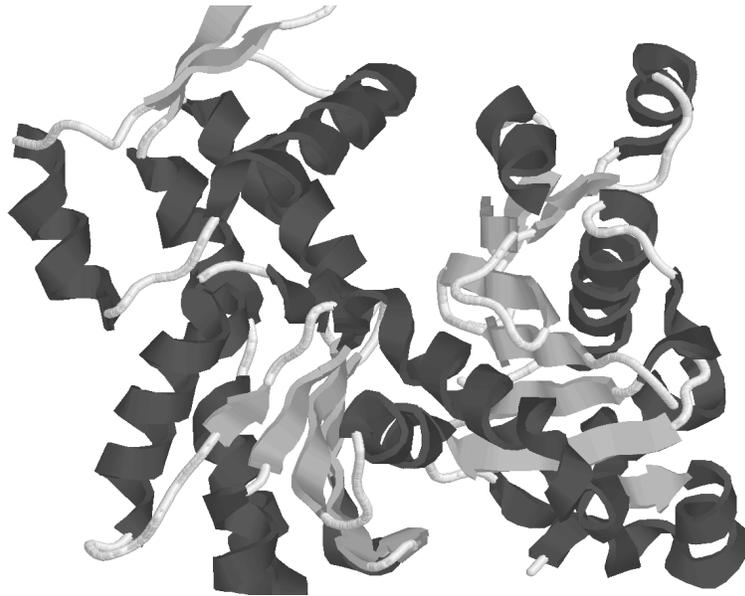
Figure 2: The History of Evolution



## References:

- 1) Richard Dawkins, The Blind Watchmaker, Norton and Company, 1996.
- 2) Fox, Dose, Molecular Evolution and the Origin of Life, Freeman and Company, 1972.
- 3) Shapiro, "The Prebiotic Role of Adenine: A Critical Analysis", Origins of Life and the Evolution of the Biosphere, 25: 83-98, 1995.
- 4) Thaxton, Bradley, Olsen, The Mystery of Life's Origin: Reassessing Current Theories, Philosophical Library, 1984.
- 5) Levy, Miller, "The Stability of the RNA bases: Implications for the Origin of Life," 95: 7933-7938, PNAS, 1998.
- 6) Shapiro, "Prebiotic Cytosine Synthesis: A Critical Analysis and Implications for the Origin of Life," 96: 4396-4401, PNAS, 1999.
- 7) Larralde, Robertson, Miller, "Rates of decomposition of Ribose and other Sugars: Implications for chemical Evolution," 92:8158-8160, PNAS, 1995.
- 8) Joyce, Schwartz, Miller, Orgel, "The Case for an Ancestral Genetic System Involving Simple Analogues of the Nucleotides," PNAS, 84:4398-4402, 1987.
- 9) Joyce and Orgel, The RNA World, Gesteland, Cech, Atkins, Cold Spring Harbor, "Prospects for Understanding the Origins of the RNA World," 1999. 10) Ohno, Evolution by Gene Duplication, Springer Verlag, 1970.
- 11) Fuller, Sanchez, Orgel, "Solid state Synthesis of Purine Nucleotides," Journal of Molecular Evolution, 1975.
- 12) Meyer, "The Origin of Biological Information and the Higher Taxonomic Categories," Proc. of the Biological Society of Washington, 117: 213-239, 2005.
- 13) Miller, "Which Organic Compounds Could Have Occurred on the Prebiotic Earth?", Cold Spring Harbor Symposia on Quantitative Biology, Volume L11, 17-25, 1987.
- 14) Spetner, Not By Chance! Shattering the modern Theory of Evolution, 1997.

## **Part 1: The Evolution of Knowledge and Information**



Cartoon representation of the crystal structure of the protein, actin.  
Actin is one of the proteins required for muscle contraction.

### **Chapter 1: Information vs. Knowledge**

---

One of the goals in this book is to investigate how molecular knowledge evolves in biological systems. To find the molecular knowledge possessed by a chemical like DNA, it is first necessary to quantify the amount of information. This is true because molecular knowledge has been defined using information.

The scientific definition for information is very different from the common one. The everyday definition implies that information should be useful or at least convey some amount of knowledge. The scientific definition does not make any distinction between useful and useless information. For example, consider the following sentences:

The brown dog likes to fetch a tennis ball.

Zxrd zgbzbue awfllt jhzhwzhg zwnzi oppwnni wyxaz.

If information theory is applied to these two sentences, the results will indicate that the second sentence contains much more information than the first. Not only is the second sentence longer, but it contains many letters that are rarely used in English (z, x, and w). The first sentence contains useful information. The second message contains no useful information. Yet information theory asserts that the second contains more information than the first. How can this be?

To understand why, consider why scientists developed information theory. The theory was developed by an engineer, Claude Shannon, who was interesting in transmitting information. The second sentence takes longer to transmit than the first, so it contains more information. This definition is clearly not useful to biologists studying evolution.

Evolution involves the creation of information that provides a selective advantage. That is the organism that possesses the new information has an edge over those that do not. Therefore, the information must be useful. This is why the word knowledge is preferable. Knowledge implies that information is useful.

In communication systems, information does not have to contain knowledge. In general, the same cannot be said for biological systems. Information that does not provide a selective advantage is often lost. Thus, the information found in biological systems usually conveys knowledge, and this knowledge provides a selective advantage. In biological systems, knowledge and information are often related, but they are not necessarily equal. Consider the following two sentences:

I have a dog. His name is Bubba. He is a black lab. He is 13 years old.

My black lab, Bubba, is 13.

Both sentences describe four identical concepts, so the knowledge conveyed by both is identical, but the first sentence contains much more information than the second. Because information has a precise mathematical definition, it can be determined rather easily. Furthermore, it is possible to define molecular knowledge in terms of information. The proposed definition is as follows:

**Molecular Knowledge:** the minimum amount of information necessary to enable a chemical (or group of chemicals) to accomplish some task or to specify some trait. The only stringent requirement is that molecular knowledge must confer a selective advantage so that natural selection can preserve it.

Because molecular knowledge is now defined in terms of information, information theory can be used in conjunction with human insight and experimental data to calculate knowledge. The rest of this chapter will explore information and its properties.

### **The Nature of Information**

Mathematically, information is defined as a reduction in uncertainty. Consider a scientist trapped in a room. He has a coin and a telephone. He is told to flip the coin and then tell his colleagues who are 500 miles away the results using the telephone. He is to repeat this process until he is told otherwise.

Before the scientist flips the coin, he does not know whether it will land heads or tails. There are two possible outcomes, and the scientist does not know which will happen until he observes the results. Suppose that the first toss is heads. As soon as the scientist observes this result, he has information. Two possibilities have been reduced to one. Before observing the coin, the scientist was uncertain of the outcome. After he observes the result, he is certain of the outcome. His colleagues do not have any information until he tells them that the coin landed heads.

A unit of information is called a bit. Whenever two possible outcomes are reduced to one, one bit of information is created. Thus, the scientist acquires one bit of information each time he tosses the coin and observes the result.

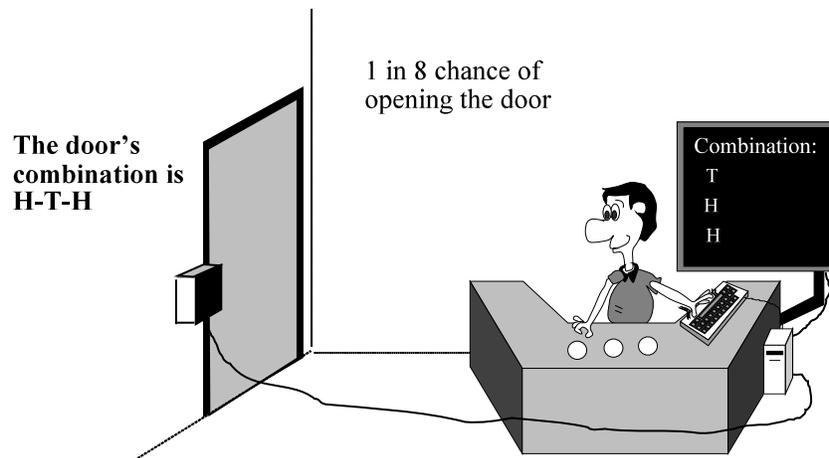
Uncertainty depends on the number of possible outcomes. For example, if the scientist is given a die to roll instead of a coin his uncertainty increases. With the coin, there are only two possible outcomes, and with a die there are six. The reduction in uncertainty for a die (six possible outcomes reduced to one outcome) is greater than it is for the coin (two possible outcomes reduced to one outcome).

Thus, the scientist acquires more information when he observes the results of tossing the die than when he observes the results of tossing the coin.

### Trapped Scientist with Three Coins

Consider a scientist trapped in a room. He is given three coins, instructed to toss the coins one at a time, and enter the results into the computer (figure 1.1). If the coin lands tails, he is instructed to enter the letter *T*, and if it lands heads, he is instructed to enter the letter *H*. The combination for the door is H-T-H. The scientist has a 1 in 8 chance of opening the door.

Figure 1.1: Trapped Scientist with Three Coins



When three coins are tossed, there are eight possible outcomes:

<u>Coin 1</u>	<u>Coin 2</u>	<u>Coin 3</u>	
H	H	H	
H	H	T	
<b>H</b>	<b>T</b>	<b>H</b>	----->opens the door
H	T	T	
T	H	H	
T	H	T	
T	T	H	
T	T	T	

When the first coin is tossed, it has two ways to land, heads or tails. The same rules apply for the second and third. So the total number of possible outcomes is  $2 \times 2 \times 2 = 8$ .

Whenever the scientist tosses a coin and observes the results, he acquires information. When he tosses the first coin and observes the result, he acquires one bit of information. After he observes the result of the second coin, he possesses 2 bits of information, and after the third, he possesses 3 bits. Suppose on his first try to open the door, all three coins land heads. After observing this event, the scientist will possess 3 bits of information. He keeps trying, and after a few more tries, the first coin lands head, the second lands tails and the third lands heads. When he enters this result into the computer, the door opens. The scientist has acquired knowledge. The combination for the door is H-T-H, and he now knows the combination.

Notice that every time the scientist tosses the coin he creates information, but only one specific outcome creates useful information or knowledge.

One bit of information corresponds to each coin. In figure 1.1, all results contain 3 bits of information. One result, H-T-H, contains 3 bits of knowledge.

Suppose that the combination is changed to H-T-H-H-H-H-T-H-H-H-H-H-H-H-H-H-H-H. The scientist is given 20 coins, told to toss all 20, enter the results into the computer and observe the door. How much information is generated every time the scientist tosses 20 coins and observes the result? Answer: 20 bits because there are 20 coins. While 20 bits of information are generated with each attempt to open the door, only one possible outcome will open the door. This is the only outcome that contains both information and knowledge.

With 20 coins, what is the probability that the scientist will find the correct combination on the first try? Answer: multiply 2 by itself 20 times to determine the total number of possible outcomes (1,048,576). Because only one of these outcomes will open the door, the odds are 1 in 1,048,576 or approximately 1 in a million.

Exponents are a useful shorthand for representing a number multiplied by itself many times. The phrase 2 multiplied by itself 20 times can be written as  $2^{20}$ . The number 10 multiplied by itself 86 times can be written as  $10^{86}$ . See appendix three for a review of exponents.

### **Information Is Closely Related to Probability**

Suppose that the scientist has 100 coins, and he is told that he has a 1 in 64 chance of opening the door if he tosses the correct number of coins and enters an *H* or *T* into the computer. He is told not to toss all 100 coins because the combination for the door is not that long. He is also told that the door only has one correct combination.

How does the scientist figure out how many coins to toss? He needs to convert the odds of opening the door into an equivalent number of bits. One way is trial and error. He can compose a table like table 1.1. If the scientist tosses 6 coins, he will have a 1 in 64 chance of opening the door.

Table 1.1- Information Contained in Coins

Number of Coins	information in bits	possible outcomes	Odds
0	0	$2^0=1$	1 in 1
1	1	$2^1=2$	1 in 2
2	2	$2^2=2 \times 2 = 4$	1 in 4
3	3	$2^3=2 \times 2 \times 2 = 8$	1 in 8
4	4	$2^4=2 \times 2 \times 2 \times 2 = 16$	1 in 16
5	5	$2^5=2 \times 2 \times 2 \times 2 \times 2 = 32$	1 in 32
<b>6</b>	<b>6</b>	<b><math>2^6= 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64</math></b>	<b>1 in 64</b>
7	7	$2^7 = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 128$	1 in 128
8	8	$2^8 = 2 \times 2 = 256$	1 in 256
13	13	$2^{13} = 2 \times 2 \times 2 \times 2 \dots \times 2 = 8192$	1 in 8192

Notice that there is a definite relationship between the number of bits in the door's combination and the odds that the scientist will open the door. Probability theory and information theory are closely related.

## Mathematical Definition of Information

The following equation defines information:

Equation 1

$$2^{(\text{information})} = \frac{\text{Total Possible Outcomes}}{\text{Observed Outcome (s)}}$$

Example 1: a scientist is told to toss a coin 3 times and remember the results. How much information does he acquire when he observes the result? Answer: The total number of possible outcomes is  $2 \times 2 \times 2 = 8$ , and only 1 outcome will be observed. So  $2^{(\text{information})} = 8/1$ . Because  $2^3 = 2 \times 2 \times 2 = 8$ , the scientist acquires 3 bits of information.

Example 2: Suppose the result in example 1 is H-H-T, but the scientist is unsure about the outcome because he cannot remember whether the first coin landed head or tails. How much information has he acquired? Answer: there are still 8 possible outcomes. The scientist observed either H-H-T or T-H-T, but he is not sure which. So both must be counted as observed outcomes. So  $2^{(\text{information})} = 8/2 = 4$ . Because  $2^2 = 2 \times 2 = 4$ , the scientist acquires 2 bits of information.

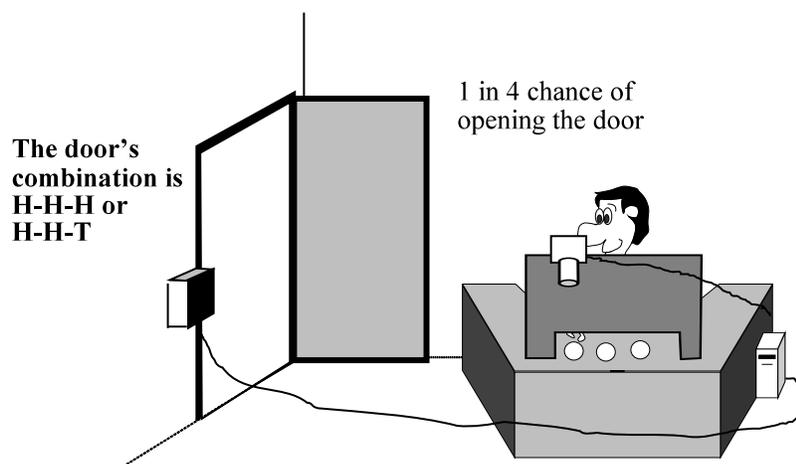
Note that the odds of an event like winning the lottery are often expressed as one in some number, like a million. The one is the observed outcome, and the million represents the total number of possible outcomes. So the information acquired by knowing who wins this lottery is easy to calculate:  $2^{(\text{information})} = (1 \text{ million outcomes} / 1 \text{ outcome})$ . Because  $2^{20} = 1,048,576$ , approximately 20 bits of information are acquired when the outcome of this lottery is observed.

The equation for information can be solved explicitly for information with the use of logarithms. Since many calculators have a log function, the next equation is often easier to use, but less intuitive than the first. See appendix three for a review of logarithms.

### **Trapped Scientist Who Cannot See the Results**

Suppose that the scientist in figure 1.1 cannot see the results of the coin toss because a screen is placed between him and the coins (figure 1.2).

Figure 1.2: Trapped Scientist Who Cannot See the Results



In figure 1.2, the camera monitors the results of the coin toss and sends the results to the computer. Depending on the results of the coin toss, the computer is programmed to do four things: open the door, close the door, beep once, and beep twice.

<u>Coin 1</u>	<u>Coin 2</u>	<u>Coin 3</u>	
H	H	H	----->opens the door
H	H	T	----->opens the door
H	T	H	----->closes the door
H	T	T	----->closes the door
T	H	H	----->computer beeps twice
T	H	T	----->computer beeps twice
T	T	H	----->computer beeps once
T	T	T	----->computer beeps once

The scientist cannot observe the results of the coins. He can only observe what the door and computer do after he tosses all three coins. The door opens for 2 of the 8 possible results. If the door is already open, and the camera observes H-H-H or H-H-T then the door will stay open. Two results will close the door if it is open, and have no effect if the door is already shut. Two results will cause the computer to beep once, and two results will cause the computer to beep twice. Does the scientist still acquire 3 bits of information when he tosses three coins?

**Case 1:** the door opens or stays open.

$$2^{(\text{information})} = (8 \text{ possible outcomes} / 2 \text{ outcomes that cause this result}) = 4.$$

Since  $2^2 = 4$ , 2 bits of information are acquired.

**Case 2:** the door closes or stays closed.

$$2^{(\text{information})} = (8 \text{ possible outcomes} / 2 \text{ outcomes that cause this result}) = 4.$$

So in the case, 2 bits of information are acquired.

**Case 3:** the computer beeps once, 2 bits of information are acquired.

**Case 4:** the computer beeps twice, 2 bits of information are acquired.

The average amount of information acquired each time the scientist tosses all 3 coins is now 2 bits. He is using 3 coins or 3 bits to transmit 2 bits of information. He must do this because the code that translates the result of the coin toss into what the door and computer do is not the optimal code. The optimal code should only require 2 coins to transmit 2 bits. One possible optimal code is as follows:

<u>Coin 1</u>	<u>Coin 2</u>	
H	H	----->opens the door
H	T	----->closes the door
T	H	----->computer beeps once
T	T	----->computer beeps twice

The average uncertainty per symbol (or coin in this example) is called the Shannon entropy. Shannon entropy\* measures on average how much each observed symbol or coin decreases uncertainty. Because information corresponds to a reduction in uncertainty, Shannon entropy is also a measure of information. When 3 coins are used to transmit 2 bits (non-optimal code), the Shannon entropy is 2/3 of a bit per coin. With the optimal code, the Shannon entropy becomes 1 bit per coin. The total information transmitted in both cases is the same because 3 coins x 2/3 bit per coin = 2 coins x 1 bit per coin = 2 bits.

\*Shannon entropy should not be confused with the term entropy as it is used in chemistry and physics. Shannon entropy does not depend on temperature. Therefore, it is not the same as thermodynamic entropy.<sup>1</sup> Shannon entropy is a more general term that can be used to reflect the uncertainty of any system. Thermodynamic entropy is confined to physical systems.

### **References:**

- 1) Brillouin, Science and Information Theory, 1956.
- 2) Reza, An Introduction to Information, 1961.
- 3) Pierce, An Introduction to Information Theory, Symbols, Signals and Noise, 1961.
- 4) Shannon, A Mathematical Theory of Communications, Bell Labs,1948.

## **Chapter 2: The Evolution of Molecular Knowledge**

---

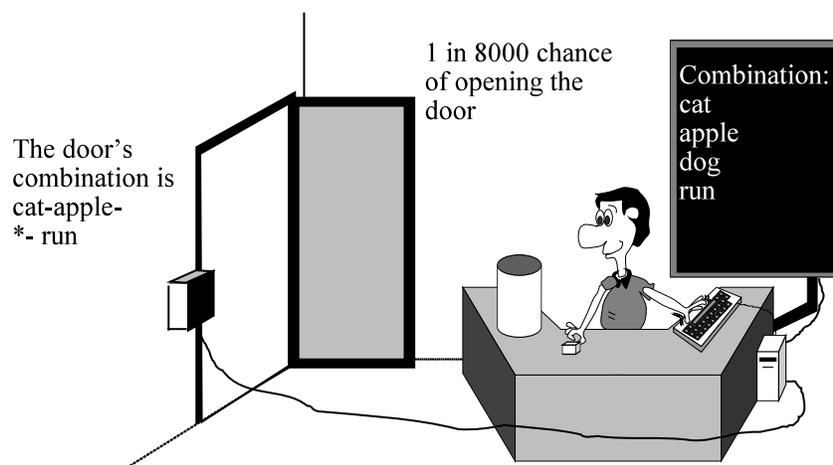
This chapter will introduce several examples to show that knowledge can be created by living organisms. This knowledge is created by chance in steps. The size of the step in knowledge determines whether or not chance will find an appropriate solution (figure 1 on page 6).

### **The Trapped Scientist**

A scientist is placed in a locked room with a computer and a combination lock on the door. He is told that the door will open when he types the correct combination of words into the computer and pushes enter. He is given a basket containing 20 wooden blocks (Figure 2.1). Each block has a single word written on it. The words are as follows: *cat, drink, bike, book, apple, run, man, soon, dog, coconut, zoo, fun, radio, sun, walk, milk, water, pear, plant, computer*. The scientist is instructed to shake the basket and then select a block without looking at it. He is to read the word, enter it in the computer, and place the block back in the basket. He is further instructed to repeat this procedure until he has entered four words into the computer. He is to press enter to see if the door opens.

The combination to the door is cat-apple- \*- run. The asterisk has a special meaning. At this position, any word is acceptable. There are 160,000 possible combinations, and 20 will open the door. It is very unlikely that the scientist will select the correct one on the first try. Each time he enters 4 words on the computer and pushes enter, the door has a 1 in 8000 chance of opening ( 1 in 20 for first word, 1 in 20 for the second, 1 in 1 for the third, and 1 in 20 for the fourth means that the odds are 1 in 20 x 20 x 1 x 20 or 1 in 8,000).

Figure 2.1: Trapped Scientist with 4 Word Combination



After a few thousand unsuccessful tries, the scientist draws cat-apple-dog-run. When he presses enter, the door opens (figure 2.1). This combination of words contains useful information, the knowledge that is needed to open the door. This knowledge was found by chance. Furthermore, this knowledge confers an advantage in that it allows the scientist to leave the room. In all of these examples, a door opening will represent a step in knowledge that confers an advantage. This book will call such a step an infon. An infon can be defined in terms of either knowledge or information.

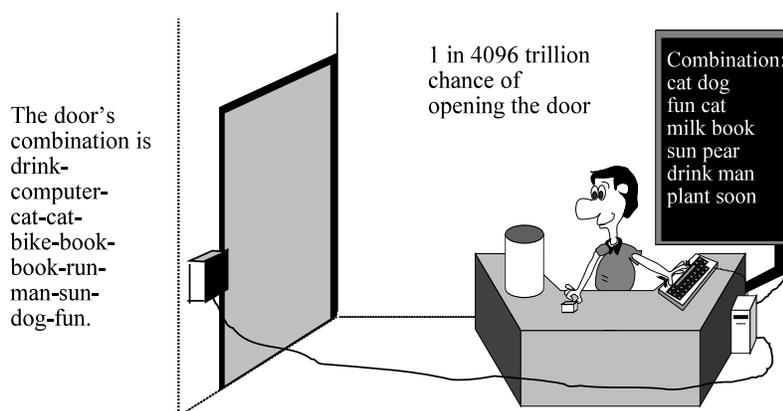
**Definition:** an infon is a step in molecular knowledge.

**Definition:** an infon is the minimum step in information that confers a selective advantage.

Now consider the same example with a much longer combination. The combination is now drink-computer-cat-cat-bike-book-book-run-man-sun-dog-fun. The scientist is instructed to repeat the procedure of drawing words from the basket until he enters 12 words.

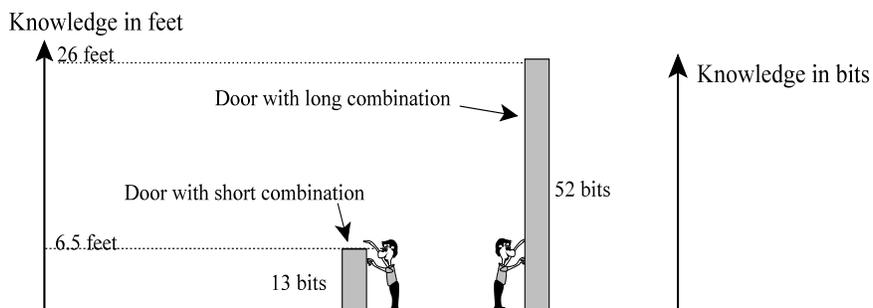
The number of possible combinations is now 4096 trillion. The scientist has a 1 in 4096 trillion chance of opening the door every time he presses enter. If he draws words for the rest of his life (even if he lives for a 100 million years), he will probably never open the door. The odds of him finding the correct combination are now just too remote. The knowledge needed to open the door can no longer be found by chance because the infon contains too much knowledge (figure 2.2).

Figure 2.2: Trapped Scientist with 12 Word Combination



In figure 2.2, the information per word is given by equation 2 on page 25.  $\text{Information} = 3.32 \times \log(20/1) = 4.32$  bits per word. Since fractions are easier to use, 4.32 is approximately  $4\frac{1}{3}$  bits; therefore, the door with a 12 word combination requires 52 bits of information to open ( $12 \text{ words} \times 4\frac{1}{3} \text{ bits per word}$ ). The door with the 4 word combination only needs 3 words to be correct; therefore, its combination contains 13 bits of knowledge ( $3 \times 4\frac{1}{3}$ ). Figure 2.3 shows how the number of bits influences chance. Each combination is represented by a wall, and each bit adds 6 inches to the wall. If chance is represented by the scientist, then he can climb over the small wall (6.5 feet high), but he cannot climb over the 26 foot wall.

Figure 2.3: Knowledge Represented by a Wall to Climb



### Information, Knowledge, Complexity and Order

Complexity and order are often confused as the same thing, but they are antonyms. Likewise, information and complexity are often considered the same, but they are not. Using the trapped scientist, these terms will now be defined.

Order is the easiest to understand. If the scientist draws the word cat 12 times in a row from the basket and enters this into the computer, then the message he enters is ordered. Ordered messages (or combinations in this example) contain patterns that allow them to be simplified. The combination cat-dog-cat-dog-cat-dog is ordered.

Complex messages are messages that are not ordered. The combination cat-drink-bike-book-apple-run-man-soon-dog-coconut-zoo-fun is complex because it is not ordered.

Information is any change in uncertainty. Once the scientist observes the result of any draw, he acquires information. The results do not influence information because each word reduces his uncertainty by the same amount. Only the combination that opens the door conveys knowledge. In figure 2.2, this combination is also complex, but it does not have to be.

The term specified complexity<sup>1</sup> is often used in intelligent design literature. This terminology is misleading because biological messages do not have to be complex. They can also be ordered.



Figure 2.4: Trapped Scientist with 4 Doors and Short Combinations

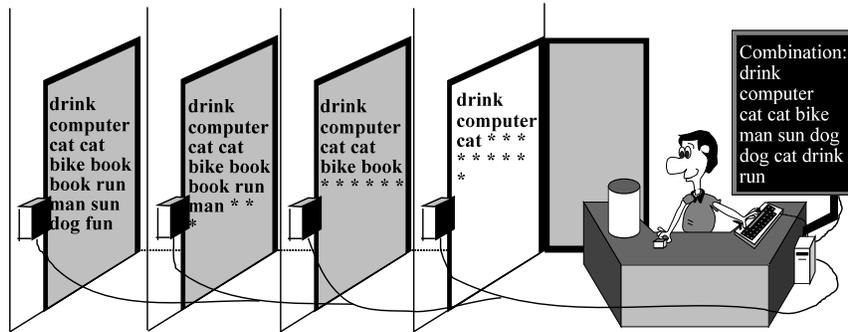
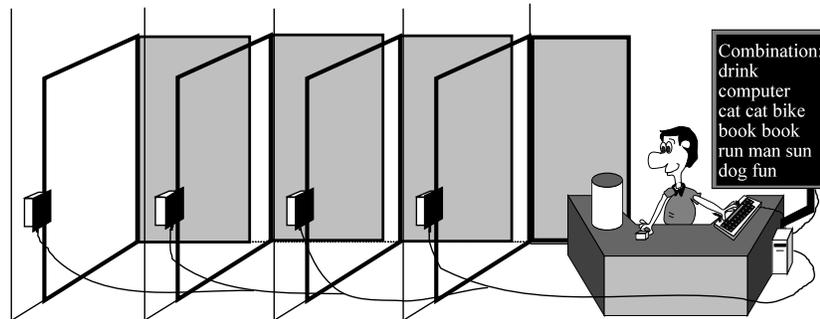


Figure 2.5: Trapped Scientist with 4 Doors Open



Notice that the knowledge required to open the last door in figure 2.5 is identical to that required in figure 2.2. The scientist finds the knowledge required in this example, and he fails in the previous one. Why? Each door represents a step in knowledge. All steps in figure 2.4 and 2.5 are small (3 words). Furthermore, because the computer preserves the combinations that open doors, only combinations that are close to the desired combination are preserved.

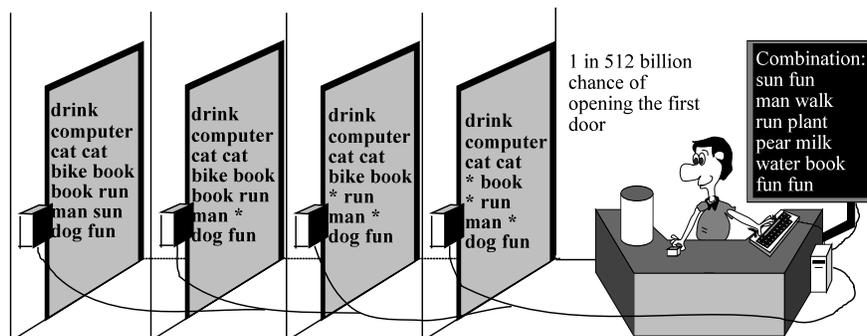
The same amount of knowledge is needed to open the last door in this example, but the correct combination is found because the steps in knowledge needed to find it are kept small by using 4 doors instead of one. Now suppose the combinations are as follows:

Inner door	drink-computer-cat-cat-*-book-*- run- man-*-dog-fun.
Second door	drink- computer- cat- cat- bike- book-*-run-man-*- dog- fun.
Third door	drink-computer-cat-cat-bike-book-book-run-man-*- dog-fun.
Last door	drink-computer-cat-cat-bike-book-book-run-man-sun- dog-fun.

The odds of opening the first door are now 1 in 512 billion. The scientist never opens the first door ( Figure 2.6).

These examples show that the number of new words needed to open a door determines whether or not chance will open the door. Each combination represents a step in knowledge. If the steps are small, (three new words or less), then the combinations are easily found by drawing the words from the basket. If the steps are large, (nine new words or more), then chance can no longer reliably find the combination.

Figure 2.6: Trapped Scientist with 4 Doors but One Large Step



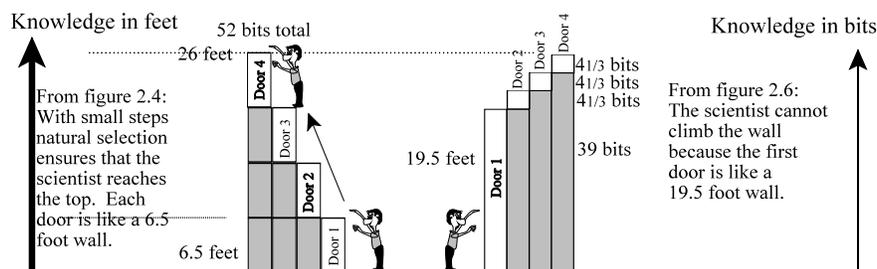
When Darwin introduced the theory of evolution, he did not consider information, but he did mandate that the changes must be slight and continuous. So he did at least understand the nature of the problem. Darwin's premise is simple. Small changes that provide an advantage are preserved by nature. Over millions of years, these changes are cumulative and thus lead to very large changes. His theory works if and only if the steps in knowledge are small.

The scientist in the locked room with 4 doors models natural selection. When the scientist opens a door, the computer preserves the combination that opened the door. That is once a small amount of knowledge is created by chance, the computer preserves it. This preservation is fully analogous to natural selection in biological systems. With the help of natural selection, the scientist can easily find the combination to a hundred doors or more (as long as the combinations are three words or less). But if the first door's combination is large, 9 words or more as in figure 2.6, then chance is no longer effective. All doors remain closed.

Natural selection is not effective when the steps are large because chance never finds the correct combination, and there is no knowledge for natural selection to preserve. So the size of the first step is critical. It completely determines whether or not the scientist can find the correct combination.

Just like before, each door can be represented by a wall (figure 2.7). Since there are 4 doors, there are 4 walls. The walls are now pushed against each other so that they form a series of steps. If all of the steps are small then the scientist can easily climb to the top - even if there are a thousand steps, but just one large step can present a serious problem. The 19.5 foot wall is very difficult for the scientist to climb. For the scientist to get over this wall, he will need some help.

Figure 2.7: Steps in Knowledge: Doors Represented as Walls



### Important Definitions:

*Darwinian evolution* - the steps in knowledge are small so chance creates useful information and this knowledge is preserved by natural selection. Darwinian evolution happens just like Charles Darwin theorized.

*Evolution by design* - the steps in knowledge are very large; as a result, chance never creates any new useful information. Natural selection is irrelevant because no knowledge is created for it to preserve and optimize.

The goal of this book is to analyze the steps in knowledge required for evolution. The results will show that many steps are similar to a locked door with a fifty-word combination. Others are similar to a locked door with a one word combination. The small steps are navigated just like Darwin theorized. Small steps do not present a barrier to chance. On the other hand, the large steps are a significant barrier. How evolution crossed this large barrier is not known.

### **Probability and Information**

The information found in a molecule of DNA or in a protein can only be associated with a probability for evolution when natural selection is excluded. Because of natural selection, it is not possible to relate the information found in the combination of the last door to a probability of evolution. The trapped scientist examples do an excellent job of explaining why this is true (compare figure 2.2 to figure 2.5). The last door in both figures has the same combination, yet one is opened easily and the other remains closed.

A probability can be associated with each individual door opening because by definition natural selection is not active before the molecular knowledge exists. Obviously, the door with the largest step in knowledge will determine whether or not the scientist escapes. This is always the door with the most unknown words (it is seldom the door with the longest combination). In most cases, the odds associated with the evolution of knowledge will depend entirely on a single door, and this is usually the first door because this is the door that creates the initial knowledge (refer to figure 1 on page 6 and figure 2.7). All doors after the initial door only need to optimize the existing knowledge. Thus, these steps are generally much smaller.

## Information in Biological Systems:

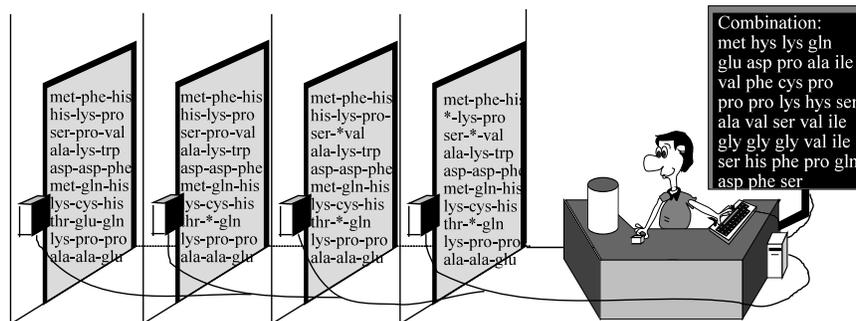
Life does not use words like *cat*, *dog* and *computer*. The words that life uses are chemicals called amino acids. Amino acids are the building blocks of proteins. Proteins implement the molecular knowledge that enables life to live. DNA merely contains the information needed to build proteins. Proteins do all of the work.

The trapped scientist example uses 20 words because there are 20 amino acids used by life to build proteins. Each word corresponds to an amino acid. So if the 20 words in these examples are replaced with the names of the 20 amino acids, then the examples will more accurately model the evolution of a protein. In this example, the combination of the doors is composed of the following words: serine, arginine, proline, leucine, valine, isoleucine, alanine, glycine, cysteine, lysine, tryptophan, tyrosine, methionine, glutamate, aspartate, asparagine, glutamine, histidine, threonine, and phenylalanine. These are the names of the 20 amino acids that are found in proteins. Each amino acid is represented with a three letter abbreviation: ser, arg, pro, leu, val, ile, ala, gly, cys, lys, trp, tyr, met, glu, asp, asn, gln, his, thr, and phe. The blocks in the basket now have these abbreviations painted on them. Nothing else has changed.

Suppose for the protein under consideration to have any function, 27 of the 30 amino acids must be correct. This corresponds to a combination for the first door as follows: met-phe-his-\*-lys-pro-ser-\*-val-ala-lys-trp-asp-asp-phe-met-gln-his-lys-cys-his-thr-\*-gln-lys-pro-pro-ala-ala-gln. Once this protein is created by chance, natural selection takes over, preserving and optimizing the sequence. The asterisks are eventually changed to the correct amino acids as shown on the last door in figure 2.8. The process of asterisk replacement optimizes the protein.

Since the protein exists, the scientist should be able to break the combination. So does he ever find the combination to the first door? Given 50 billion years with 100 million tries a year, the probability of the scientist opening the first door is only 1 in 27,000 trillion. He does not open the door.

Figure 2.8: Using Amino Acids for Words



Reference:

- 1) Dembski, Intelligent Design, InterVarsity Press, 1999.