Information Value and Darwin's Theory of Evolution

Abstract

Many scientists regard Darwin's "evolution through the natural selection of heritable differences arising at random in each generation" as a plausible theory for the evolution of a complex genus from more primitive predecessors. This paper looks at human evolution using an information-theory approach.

Verifying Darwin's theory using an information theory approach illuminates major open issues, for which no easy answers appear to exist. This does not invalidate Darwin's theory, but establishes the need for minor or major extensions/improvements. It also establishes the need for alternatives. Additional research is required to settle the issues.

The Design File

When an engineer develops a new complex structure he does this by creating a design file. If he develops a new computer program, the design file consists of high level source code in a language like C/C++, Cobol, Visual Basic or Java, that a compiler converts into a machine program (binary), which can be executed on a computer.

Similarly, when he wants to create a new microprocessor, he creates a description in a language like VHDL or Verilog. Through steps like synthesis, wiring and placement a mask set is generated, which a silicon foundry uses to manufacture the new microprocessor.

The design file of a living being is called a genome. The genome is a complete set of genetic information, stored as DNA sequences within a number of chromosome pairs of the cell nucleus and in a small DNA molecule within the mitochondrion. Deoxyribonucleic acid (DNA) is a molecule that encodes the genetic instructions used in the development of all known living organisms

DNA molecules are double-stranded helices, consisting of

- sequences of "nucleobase" (guanine, adenine, thymine, and cytosine), recorded using the letters G, A, T, and C, as well as
- a backbone to which the nucleobases (G, A, T, C) attach at regular intervals.

It is the sequence of these four nucleobases along the backbone that encodes genetic information. This information is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.

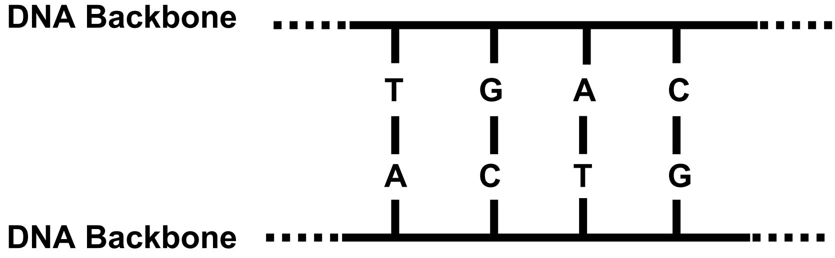


Fig. 1 : DNA

The nucleobases (informally, bases) form pairs between the two backbones: A with T and C with G. Since there are 4 bases, each base pair encodes 2 bits of information.

The human genome contains over 3 billion base pairs organized into 22 paired chromosomes, the X chromosome (one in males, two in females) and, in males only, one Y chromosome, contained within the cell nucleus. It also includes the mitochondria DNA. Together this represents over 6×10^9 bits of information.

It is generally assumed that the homo sapiens genome and the chimpanzee genome differ by only a few percent or between 10^7-10^8 bits. The bulk of this probably goes into the part of the design file which describes the human brain. Thus, 10^7 bits is a reasonably lower number.

10^7-10^8 bits are not a particularly large amount of information, and represent some rather tight coding. For example, a minimum Linux Operating system requires some 7×10^8 Bytes or 5×10^9 bits. A normal Linux distribution would be larger by something like a factor 5. A design file in VHDL or Verilog used to implement a new microprocessor would be somewhat of the same size.

The human brain is assumed to contain about 10^{11} neurons (nerve cells). The complexity of the brain is due to the massive number of highly interconnected units working in parallel. The number of interconnections between these 10^{11} neurons is not known very well, but is estimated to be in the 10^{14} and 10^{15} range. The representation of each connection requires several bits of information. Thus, when a new baby is born, the 10^7-10^8 bits in the design file need to be expanded into the 10^{14} and 10^{15} bits range.

Human evolution

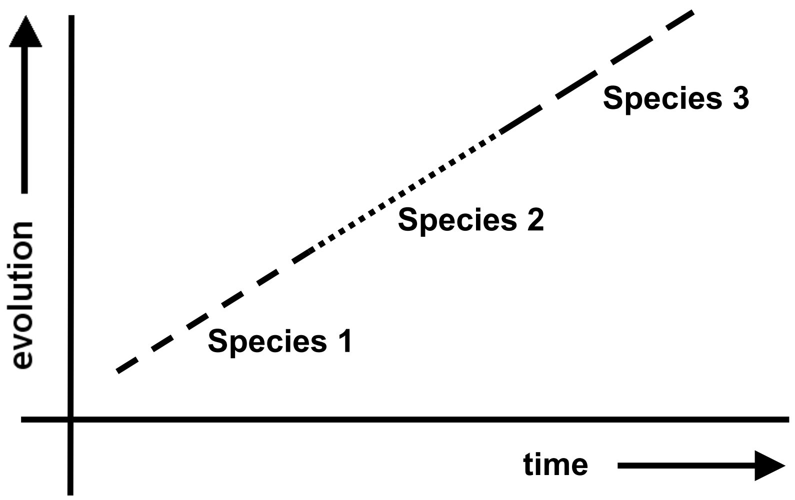


Fig. 2 : Gradual evolution

Until the 1970's, it was generally thought that slight genetic changes in a population from generation to generation had the effect that past species evolved gradually into other species over millions of years.

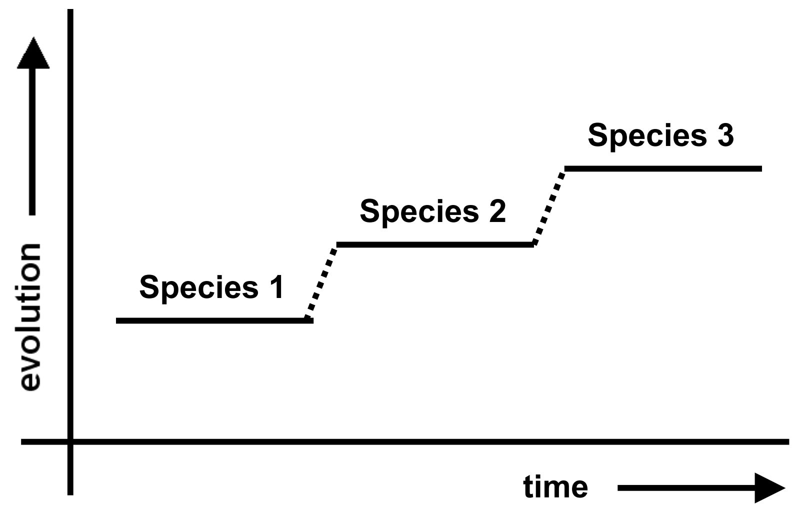


Fig. 3 : Staircase model

The evidence in the fossil record indicates that most species remained essentially the same for millions of years and then underwent short periods of very rapid, major change. The fossil record does not convincingly support transitions from one species to another.

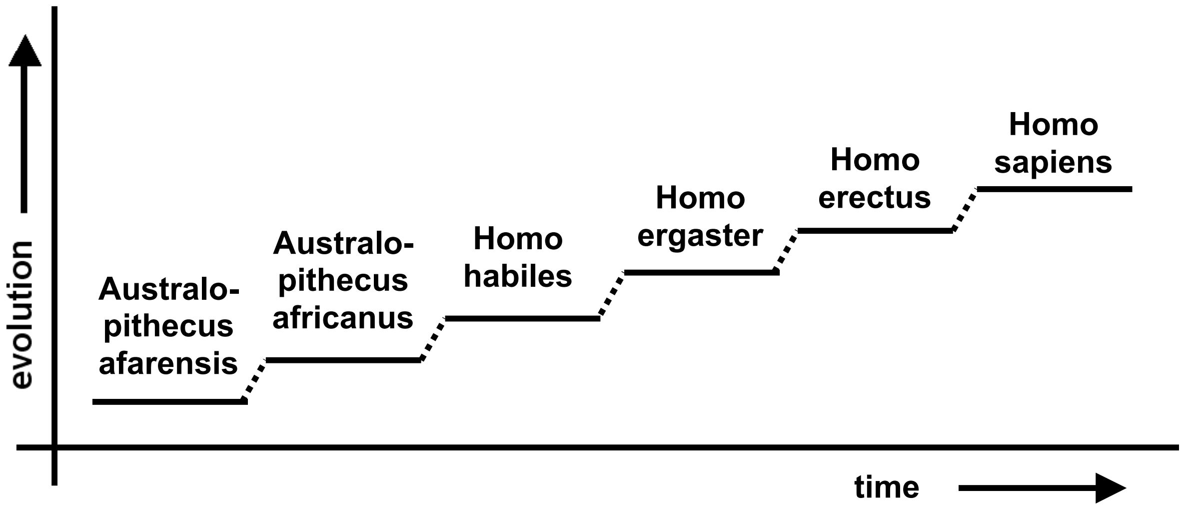


Fig. 4. : Simplified model of human evolution

The human evolution shown in fig. 4 matches the staircase model shown in fig. 3. Fossil discoveries of many individuals in many sites support fig. 4. Additional sites have yielded remains of single or a few individuals. There is disagreement which ones of these are ancestors of homo sapiens.

Australopithecus afarensis lived approx. 4 x 10^6 years ago. If we assume 10 years as the length of an average generation, then the evolution from australopithecus to homo sapiens took a total of 4 x 10^5 generations.

We have to guess at the total human population during this timeframe. Most estimates assume a total population of less than 100 000 individuals, with less than 10 000 at some times. For the purpose of the following discussion let us assume a constant average population of 10^5 humans during the last 4 million years, until the expansion of homo sapiens.

4 x 10^5 generations, and 10^5 individuals pro generation means a total of 4 x 10^{10} individuals since the appearance of australopithecus afarensis about 4 Million years ago.

The individuals of each new generation may suffer a mutation of their genetic inheritance. Most mutations will be either harmful or irrelevant. Some mutations may be useful. We call these mutations “useful mutations”.

Thus, a total of 4 x 10^{10} humans have lived since the appearance of australopithecus afarensis. If we assume that 10 % of those obtained a useful mutation, the total number of useful mutations has been 4 x 10^9.

A "useful mutation" may be either successful or unsuccessful. The vast majority of useful mutations will be unsuccessful. It is successful, it succeeds to create a new human species that is sufficiently superior to the preceding one that it succeeds in replacing it. An "evolution step" is the result of a successful useful mutation to the human population. However, only a very small percentage of all useful mutations will be successful.

The number 4x10^9 is not particularly large. A 4 GHz x86 computer is able to execute about 10^9 machine instructions within each second.

Information Value

Information Theory – the entropy of Information - was developed by Claude E. Shannon in 1948 [SHA48]. Information Theory has proven useful for the encoding and transmission of information over a noisy communication channel. There are also parallels to the entropy concept in thermodynamics. To the disappointment of many scientists, information theory has yielded few additional useful application.

An exception is a paper by Kelly, published in 1956 [KEL56] . Kelly's paper deals with the issue of “value of information”. It considers the case of a gambler, playing a game of chance. The gambler receives à priori information from an entity called the "advisor", who, through some faculty, has previous knowledge as to the outcome of the next game of chance prior to placing a bet. The paper then calculates the value of the information obtained.

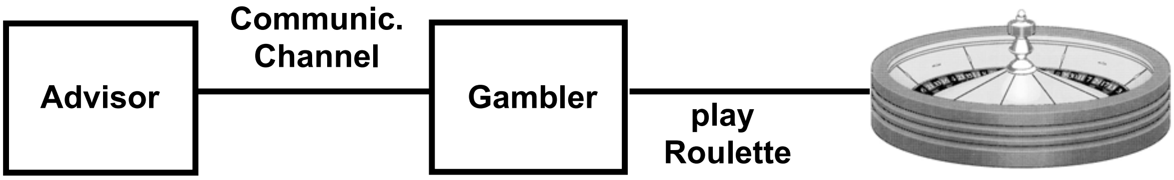


Fig. 5. : Original gambling set up

Assume the gambler plays even/odd chances (even money bets, e.g. betting red or black). Receiving a message from the advisor containing 1 bit of information guarantees a win with certainty. After n bets he would have 2^n times his original bankroll. This permits an exact calculation of the “value of information” obtained via the communication channel.

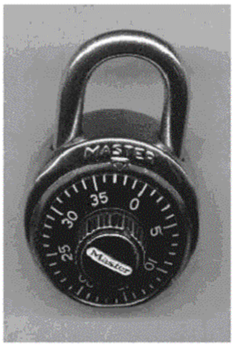
There are several important prerequisites for this to work. First, the gambler has to know how to decode a possibly lengthy message to extract the one important bit. Then, the information value depends on the gamblers starting capital. The higher his starting capital, the higher his gain.

Finally, the gambler has to know how to play roulette, and not to use the information in a room next door at a game of black jack. This is important: the information value depends on what the receiver can do with it.

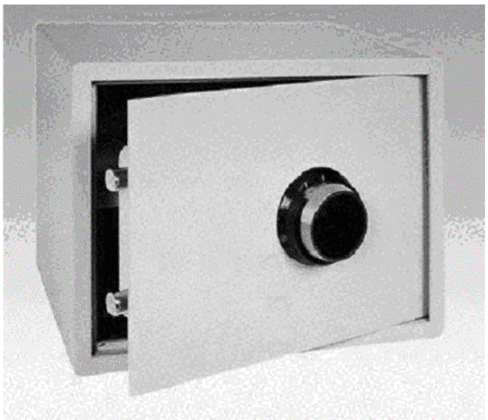
Now assume the roulette wheel has 64 slots instead of the usual 36, and no 0 or 00 slots ("even odds"), and the gambler places his bet on a particular number. Without receiving a message from the advisor, the probability to correctly guess the outcome of the next roulette wheel spin is P=1/64. 64 is a binary number ($2^6 = 64$), that can be expresses with 6 bits of information. If the gambler receives the 6 bit number via the communication channel from the advisor, this again guarantees a win with certainty.

Kelly's paper extends this analysis to a more general case where the information is transmitted over a noisy communication channel. He demonstrates that the information value is in accordance with Shannon´s information rate over a noisy channel.

Now let us modify the model shown in fig. 5 . First we replace the roulette wheel with a numeric combination lock, as shown in fig. 6 . Assume the combination lock has 64 positions. The game consists of guessing/playing the correct combination lock number (the key).



Numeric Combination Padlock



Safe with Combination Lock

Fig. 6. : Numeric combination locks

Next we assume the role of the gambler is being performed by a computer. The computer has a 6 bit buffer. The 6 bit buffer stores the number, the computer/gambler will use for his next bet. Also assume the combination lock has 64 positions. This configuration is shown in fig. 7.

We consider 2 alternatives. In alternative 1, the advisor knows the combination lock number, and transmits a message to the gambling computer. The gambling computer knows how to extract the number from the message and stores it in his buffer. Using this number, the gambling computer is able to open the combination lock safe at the next trial.

In alternative 2, the advisor does not know the combination lock number. It uses a random number generator to generate a sequence of 6 bit numbers and transmits them to the gambling computer. Let us assume it does not remember, which numbers it has previously transmitted, and thus may transmit the same number twice. The gambling computer stores the number in its buffer and tries again, betting a small fraction of his starting capital. On the average it will be successful after 64 trials, recovering the losses of his previous bets.

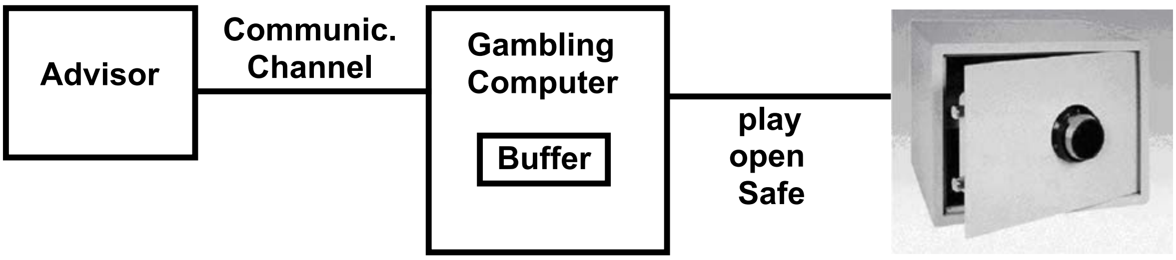


Fig. 7. : Combination Lock Safe Model

In both cases the gambling computer will remember the correct 6 bit number. In the future it will be able to successfully open the combination lock safe on the first try.

Now assume, the combination lock safe is sitting in a plain, populated by a horde of homo erectus. It is refilled at regular intervals with some food, after which the door of the safe is closed, if it was open before. Each individual plays the role of the gambling computer. Clearly, the individual, who has stored the correct combination lock key in his buffer has an evolutionary advantage.

The information value is determined by the number of useful mutations (successful or not). A useful mutation corresponds to a trial opening the safe (probably unsuccessful). A successful useful mutation corresponds to opening the safe. In real life, the successful useful mutation will provide some other usefulness not available before.

Darwin's "evolution through the natural selection of heritable differences arising at random in each generation" adds usefulness to the genome via a successful useful mutation. This is the same as adding information value to the genome. It does not matter if the usefulness is obtained by opening a safe or by anything else.

You may want to consider more complex configurations than shown in fig. 7, Please remember, Kelly's paper considers a game of even odds. He defines as "Information Value" the number of bits, which the advisor has to transmit to change a game of chance into a win with 100 % certainty. You may want to consider more complex configurations than shown in fig. 7, but you have to maintain the same definition of information value. Please remember that generations of roulette players have unsuccessfully tried to invent "systems", that "beat the house". It is not at all obvious, why Kelly's concept of information value should not apply to any kind of human evolution.

This is the message: You can receive an information value of n bits by obtaining it from somewhere else. Or you can develop an information value of n bits by 2^n random experiments.

Information Value of a Mutation

We have assumed:

- **The total number of useful mutations during the last 4 Million years has been 4 x 10^9.**
- **There have been 6 different species (6 evolution steps) between australopithecus afarensis and homo sapiens. Thus there have been 4 / 5 x 10^9 useful mutations to change (single evolution step) from one species to the next.**
- **They represent 4 / 5 x 10^9 trials for each evolution step.**
- **4 / 5 x 10^9 trials equals about 2^{29} trials. This represents about 29 bits of information value for each evolution step, or 5 x 29 = 145 bits of information value for the total evolution from australopithecus afarensis to homo sapiens.**

As a lower limit we also assume 10^7-10^8 bits for the part of the design file responsible for the development of the human brain. Thus, when a new baby is born, the 10^7-10^8 bits in the design file need to be expanded into wiring information in the 10^{14} to 10^{15} bits range. This in itself is an achievement where we do not know how it is done. However, we cannot assume a value less than 10^7-10^8 bits.

Darwin's theory generates 29 bits of information value for each evaluation step. This is a very small amount of information. In other words, adding or changing 29 bits of information within 10^7-10^8 bits design file is supposed to account for the growth in intellectual capability between homo erectus and homo sapiens. Based on the existing body of scientific knowledge, this is inconceivable.

Also, a total of 145 bits of information value is supposed to account for the brain evolution from australopithecus afarensis to homo sapiens. Again, based on the existing body of scientific knowledge, this is inconceivable.

You probably do not like my assumptions or my numbers. I invite you to substitute your own values. You will be surprised, the result will not be that much different.

The Architecture of Complexity

A paper by Nobel Laureate Herbert A. Simon offers perhaps a solution [SIM62].

In his paper Simon discusses two watch makers, named Hora and Tempus, who manufacture very fine watches, which are in high demand. They spend their time assembling watches from exactly 1 000 parts each, procured elsewhere. Tempus's watch design has the characteristic, that a partially finished assembly falls apart, and Tempus has to restart from scratch, when he is interrupted in his work by a customer phone call, ordering another watch.

Hora has designed his watch so that he could put together subsubassemlies of 10 parts each. Ten of these subsubassemblies could be put together into a larger subassembly, and a system of ten larger subassemblies constitute the whole watch. Hence, when Hora has to put down a partly assembled watch in order to answer the phone, he loses only a small part of his work, and on the average he assembles his watches in only a fraction of the time it took Tempus.

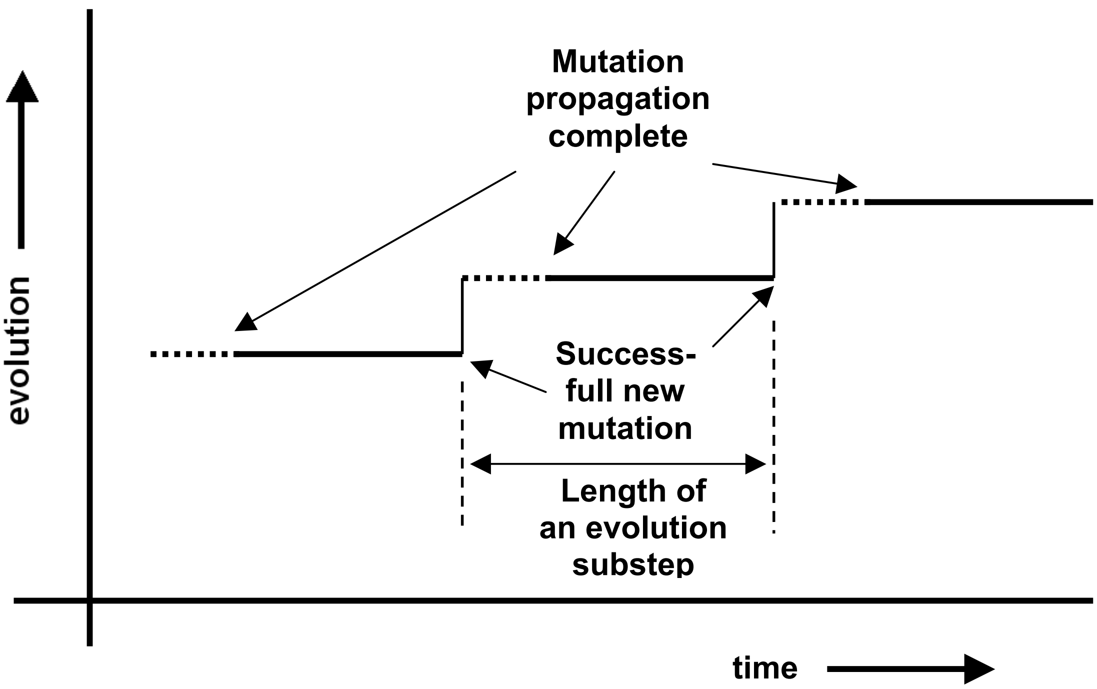


Fig. 8 : Substep model

We can apply this concept to the evolution Model shown in fig 8. The model is limited to a small part pf the population of for example 10 000 individuals, living within an isolated limited-size geographical area.

Evolution occurs as a sequence of substeps. A successful new mutation starts a substep. It provides superior survival characteristics to all individuals carrying the successful new mutation in their genes. An evolution step consists of a sequence of substeps. While an evolution step involves the creation of a new species, a substep does not.

A substep consists of 2 phases. During the first phase, the successful new mutation is inherited by a growing number of individuals in the population. In the second phase, all individuals have inherited the successful new mutation, and the next successful new mutation can occur.

If there is just a single substep within an evolution step, 4 / 5 x 10^9 or ≈ 2^{29} useful mutations occur during its lifetime, generating an information value of 29 bits. This is the same information value for both the substep and the evlution step.

If there are 10 substeps within an evolution step, 4 / 5 x 10^8 or ≈ 2^{26} useful mutations occur during each substep, generating an information value of 26 bits, and an information value of 260 bits for the evolution step.

If there are 100 substeps within an evolution step, 4 / 5 x 10^7 or ≈ 2^{23} useful mutations occur during each substep, generating an information value of 23 bits, and an information value of 2300 bits for the evolution step.

And so on.

For different substep numbers the results are summarized in table 1

1 Years pro substep	2 Number of substeps	3 Number of useful mutations decimal	4 Number of useful mutations binary	5 Substep information value	6 Evolution step information value
800 000	1	4/5 x 10^9	2^{29}	29 bits	29 bits
80 000	10	4/5 x 10^8	2^{26}	26 bits	260 bits
8 000	100	4/5 x 10^7	2^{23}	23 bits	2300 bits
800	1 000	4/5 x 10^6	2^{20}	20 bits	20 000 bits
80	10 000	4/5 x 10^5	2^{17}	17 bits	170 000 bits

Table 1 : Substep model

Thus there is progress. With 10 000 substeps the information value for the evolution step is 170 000 bits. This approaches the gradual evolution shown in fig. 2 . It also starts to approach the postulated minimum for the design file of 10^7-10^8 bits.

The low value of substep information value (column 5 in Table 1), lower than in fig. 7, makes the substep model even less credible. There are several additional problems with the substep model shown in fig. 8.

- **The fossil record does not support the substep model.**
- **It took the definitely superior homo sapiens several 10 000 years to replace homo erectus in Asia and homo neanderthalis in Europe. How can a replacement or addition of just a few ten bits account for the superior characteristics required for a particular substep population to replace its predecessor ?**
- **In particular the lowest entry in table 1 (80 years between substeps or 8 generations) is difficult to believe, since it will take at least all that time to propagate the successful mutation. Since most of the time only a small fraction of the entire population will carry the successful mutation, there is no time left to collect another successful mutation.**
- **Even 170 000 bits fall far short of the required 10^7-10^8 bits in the design file.**

Thus, based on the existing body of scientific knowledge, the substep model is even less credible than the Combination Lock Safe Model.

Again, you probably do not like my assumptions or my numbers. I invite you to substitute your own values. Just as before, the result will not be that much different.

The missing information

Does all this eliminate Darwin's theory of "evolution through the natural selection of heritable differences arising at random in each generation" ?

We know to little about the world in which we are living. Thus the gaps in Darwin's theory may very well be filled by future discoveries. For example, some computer scientists believe in self-organization of highly parallel structures. In other words, there exists supposedly a law of nature, which says that a sufficiently large parallel structure will exhibit intelligent characteristics by its own. This may very well be true. On the other hand, maybe not. It certainly has never been observed.

There may be other, yet undiscovered factors, which could provide a solution.

Today Darwin's theory is just that – a theory confirmed by the observation of small genetic changes within a species. It is based on speculation, not supported by scientific evidence, for the generation of a new species.

The most frequently discussed alternative assumes the existence of some superior intelligence, which supplies the missing information.

This entity is the Yahweh or Adonai of the Jews, the God of the Christians, the Allah of the Muslims, or the more diffuse entity worshipped by Hindus, Buddhists or Taoists. Or it is an alien civilization from outer space. Many people believe to see in evolution the evidence of a superior intelligence at work.

However, all this again is based on speculation not supported by scientific evidence.

We could try to use Occam's razor to decide between the two competing theories. Occam's razor states that among competing hypotheses, the hypothesis with the fewest assumptions should be selected. In other words, the simplest explanation is usually the correct one. Unfortunately, we know so little about the world in which we live, that this approach fails as well. Also, there may very well a third or a fourth theory, which has not been formulated so far.

So, where does this leave us ?

The approach for the future.

What should we do ? We should continue our research in artificial intelligence, trying to fill the gaps in Darwin's theory. We should be very open minded about new discoveries.

In addition, a superior intelligence has at some point to impact the human genome. We could try to observe this phenomenon.

Unfortunately, this will be very difficult. We have assumed 10^7-10^8 bits for the part of the design file responsible for the development of the human brain. This information has been added over a period of 4 Million years. If we divide 4 million years by the 10^8 bits in the design file, this results in a data rate of approx. 10^{-6} bits per second. It is way beyond the existing technological capabilities to detect or measure such a small data rate within a probably very noisy environment. This does not mean, we should not try. In cosmology, dark energy is an accepted theory, although we presently have no means to verify it.

Finally, we should be very open minded to discuss additional evolution theories. Our scientific horizon is expected to expand each year and will open up new vistas.

Good hunting !

Literature

[KEL56] J. L. Kelly, Jr, A New Interpretation of Information Rate.
Bell System Technical Journal, 35, (1956), 917–926.
www.grx.de/Mirror/Kelly05.pdf.

[SHA48] Claude E. Shannon, A Mathematical Theory of Communication.
The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
www.grx.de/Mirror/shannon1948.pdf.

[SIM62] Herbert A. Simon, Jr, The Architecture of Complexity.
Proceedings of the American Philosophical Society, Vol. 106, No.6, December 1962, p.467.
www.grx.de/Mirror/Simon01.pdf.