Some Scribble of Things

Preamble

This article is the result of my quest for some fundamental philosophical questions, inspired by my quest for the future of artificial intelligence. All Intelligences, artificial or not, resemble a lot of similarities. There is a sister article related to this one, called “A few predictions on artificial intelligence”. That article purely focuses on the practical discussions of artificial intelligence, but also reflects some ideas presented in this article. Artificial intelligence is not emphasized in this article. However, some comparisons with them are sparsely discussed in various sections.

This is a lifelong quest. This article only reflects my thoughts at the end of 2017.

Please be forewarned that some of the statements in this article are very provocative. Please read with your own discretion.

The Premise

Early in my childhood, I have believed that the societies evolve from less efficient to more efficient[1]. That said, productivity itself advances from less efficient to more efficient, which drives everything else to create a least friction society, given productivity.

As an extrapolation, any living organism evolves from less efficient to more efficient, given the environment.

Another premise I make is that the past is before the present, and the present is before the future[2], in one episode of the universe[3].

Together, we may conclude that the intelligence evolves from less efficient to more efficient, from the past, to the present, to the future.

All these are driven by simply the survival of the fittest[4], which can be extrapolated to the survival of the most efficient.

The Stages of Evolution

In this section, I will first describe how I divide the evolution in stages without a lot of systematic discussions. The later sections discuss several aspects of the evolution across different stages.

---

[1] This is partially influenced by the “means of production” and the “relations of production”.
[2] I do not suggest absolute time. Time is a very local concept. However, the direction from past to present, to future is the same.
[3] The definition of episode will be apparent by the end of the article.
[4] It is different from Darwin’s interpretation.

In those sections, the stages of the evolution are abbreviated as S<n>, as the subsection title indicates. The abbreviations are referenced frequently in the later sections.

Please note, evolution is a continuous process. Organisms at the boundary of two stages may exhibit the characteristics of both stages.

In my point of view, the metrics to measure the evolution stages of an organism may be condensed to the answers to two questions:

- How the organism behaves in its lifetime (the ability to learn).
- How the behaviors are passed/changed from generation to generation (the ability to evolve). .

And, survival of the fittest is the ultimate driver of evolution.

S0: Behavior indifferent to environment, evolution indifferent to environment

In this stage, evolution is completely by chance, and there is no direction in evolution.

If we look back to history, this is the time span from the beginning of the universe to after the first protein appear on earth, before the first cell division or replication. The process is completely driven by the physical/chemical process rather than the biology process.

If you are familiar with science and biology, you may know how rare it is to create the first protein, especially the creation of counter clockwise dominated proteins. There are gazillions of tries and failures.

This is like shooting in the dark, aimlessly. There is no living organisms, not to say survivors. It takes billions of years to reach the next stage.

S0.5: Behavior reactive to evolution, evolution indifferent to environment

For completeness, I put down this stage. In this stage, some complicated structure may be created by chance, which may react to the environment. However, there is no simpler process to recreate the structure. Once the structure is no longer active, its species is extinct.

Those structures are destined not selected by nature and cannot survive more than one generation.

S1: Behavior indifferent to environment, evolution reactive to environment

Things are different when proteins can replicate itself. This behavior comes from the inner structure of the protein and is not related to the environment. However, when the environment change, some errors in replication create variations and some variations are more adaptable to

the change and survive more. After repeated change of the environment, the organisms may bake the change in their DNA and adapt the change between generations.

This stage spans from single cell organism to some plants, many of which still exist today.

S2: Behavior reactive to environment, evolution reactive to environment

As species evolve, the organisms may adapt to the slowly changing environment, such as four seasons. In this stage, the plants bloom in spring, grow in summer, ripen in autumn, and wither in winter. The change itself is not an indication that they are in S2. Plants in S1 can do the same. However, if the spring comes late a few weeks and the plant blooms late accordly, the plant is in S2.

The rise of the animals is one more step in the evolution. They can mobile themselves in reaction to the changing environment. The migrant bird is one example that they fly thousands of miles to search habitable weather.

One key characteristic for the species in this stage is that their behaviors are always reactions to the changing environment. After many repetitions, they memorize the repeated behaviors, and later on, such repetitions are carved in their DNA so they can repeat the behaviors right after they are born. Thus, even in some cases they appear to prepare for the changing environment, they solely remember what has happened before or in previous generations.

S3: Behavior proactive to environment, evolution reactive to environment

Then comes mammals and humans, who can truly predict what is going to happen and acting accordingly, to different degrees.

Human, as the dominant species in the world, is capable of predicting much further in the future, and can plan complicated strategies to achieve the goal, whatever the goal is. Human is conscious.

In addition, human is no slave of the environment. Instead, they proactively change the environment to make it more convenient (more efficient) for them.

However, even though a human learns a lot in her lifetime, she cannot pass down those valuable skills to the next generation without referring to external assistance. The evolution of human is still a slow and random process. The efficiency gap between the behavior and the evolution is the biggest inefficiency of human beings, in my opinion.

We have seen all the stages above. So they are easier to understand. However, we have not seen any of the stages below. They are merely extrapolations of what we have seen before.

S3.5: Behavior reactive to environment, evolution proactive to environment

We have not seen this stage. But most likely because this stage is unstable and cannot last long. Conceptually, evolution is much harder than changing behavior. Thus, evolution usually drags in terms of efficiency. It is difficult to find a state that is more efficient to evolve than learn.

If some beings evolve to this stage, they are very stupid throughout their lifetimes, but they become noticeable smarter between generations. If the trend continues, soon their behavior cannot be just reactive to the environment because they are smarter. They will quickly reach the next stage.

S4: Behavior proactive to environment, evolution proactive to environment

This is a hypothetical stage. Like humans, those higher beings can proactively change the environment, making the world a better place to live.

Better than humans, those higher beings can proactively change the information encodings they pass to their offsprings, in the hope that their offsprings will be faster, higher, stronger, or better in any aspect than them.

In this stage, the boundary between male and female is blurred, as the evolution is no longer based on permutation, With a gift like this, the diversity of the higher beings first explode and then dwindle dramatically. There will be many species for the higher beings, since the self-directed evolution make several, if not many, reproduction patterns. The success of the high being species really depends on how far they can predict the future, as the consequence of some important evolution decisions will not become obvious until many generations later, at which time, it may be too late to change the course.

Human, as of now, just starts to manipulate the DNA via genetic engineering. There is still a long way to go.

S5: Behavior proactive to environment, evolution is morphed to self-evolution

In this hypothetical stage, the evolution has reached yet another milestone: death due to natural cause is the eliminated. In another word, the higher being can live forever. It doesn't mean the evolution has come to an end. Instead, the evolution is performed at a component level that anything breaks can be repaired or replaced. The thoughts are contiguous even though the physical material embedding the thoughts may wear out and be replaced.

However, accidents do happen. At that time, the higher beings may still die, even though the likelihood of accidents is very low. This is yet another indication that the higher beings fail to

predict the future long enough for foresee the accident. Or the higher being is incapable of adapting abrupt changes of the environment.

S6: Behavior controls to environment, life is eternal

In this hypothetical stage, the higher beings simply do not die. They live forever. They can predict far enough in the future that no accident can happen on them. Even if something unexpected happen to them, they are very quick to respond, and have safeguards to guarantee the continual survival.

One key characteristic of the higher beings in this stage is that they do not have physical presence. In other words, they are ascended.

Why is it so? Let's consider a hypothetical scenario that shows any being with physical appearance may not guarantee to live forever.

As we all know, in this universe, nothing travels faster than light, which is also the speed of causality, except the expansion of the universe itself at some point in time. Hypothetically, during the hyperexpansion phase of the universe, a death light is created instantaneously. Anything the death light reaches is dead, including the higher being. Because the death light is at the speed of light, there is no warnings that the higher being can receive, and there is no time to prepare. The only time the higher being is aware of the death light is the time the death light hit something (after that the higher being can predict the course of the death light and try to avoid it). However, if that something is a higher being, that higher being has no way to protect herself, but be dead.

Put it in another way, everything has a light cone. If the death light is not in the past light cone of the higher being, but intercept the higher being at present, the higher being has no time to react.

This demonstrates that any eternal higher being does not have a physical body.

S7: Behavior is the environment

This is the final hypothetical stage. At this stage, the higher being can predict perfectly to the end of the time. The higher being knows the past, the present, and the future, she is the GOD.

In a narrow sense, it is unlikely that two higher beings can coexist in a universe with overlapping light cones, because one higher being is unlikely to predict the behavior of the other higher being perfectly to the end of the time (otherwise, the other high being is just a puppet). It is however possible that multiple higher beings live in separate light cones that do not intersect (such as one lives in a singularity, e.g. black hole).

Even though the S7 beings can predict perfectly to the end of time, I do not suggest that the world is determined. However, since the S7 beings are all ascended, they can be viewed as

some wave functions. And since there is no isolated observer in quantum mechanics, their wave functions are superposed to the wave functions of the real environment. Thus, in a broader sense, multiple S7 beings can co-exist in a universe (likely highly correlated). The wave functions we see are just the quantum superposition of the wave functions of the real environment plus all S7 beings.

Since the higher being is capable to predict to the end of the time, she no longer needs evolution.

Survival of the Fittest

The phrase “survival of the fittest” is first used by Herbert Spencer and then later adopted by Charles Darwin, as a synonym to “natural selection”.

In my case, however, the phrase “survival of the fittest” is a tautology, merely meaning “survival of those who are better equipped for surviving”.

Intellects live longer

Instead of focusing on competition or cooperation, the phrase implies that any living organism desires to live longer, and the one who succeeds living longer is the one who is most equipped to live longer.

Human is not the longest living species. Some animals like turtle can live much longer than human. Some trees can even live thousand of years. However, human is still the dominant species so far, why?

I think the answer lies in the fact that we cannot simply focus on the longest living individual in a species. There is a long tail of everything. Instead, we should focus on the distribution of the lifespan of all individuals and calculate the mean or average lifespan for the species. If we take that into consideration, a tree can easily reproduce hundreds of millions of seeds in its lifespan, with only a tiny fraction sprout. A fish can easily reproduce millions of eggs. For humans, the number is in the lower hundreds (or single digit if only considering fertilized eggs), but the survival rate is much higher. I haven’t done the calculation, but it seems that human is among the top if not the first in terms of average (median) lifespan[5].

Please note, the survival rate not only depends on what has happened, what is happening, but depends more on what is going to happen. Thus, for intellects, especially the advanced intellects, the survival rate is closely coupled with the ability to predict. If an intellect can predict what’s going to happen, she can eliminate the threat preemptively, and thus survive longer.

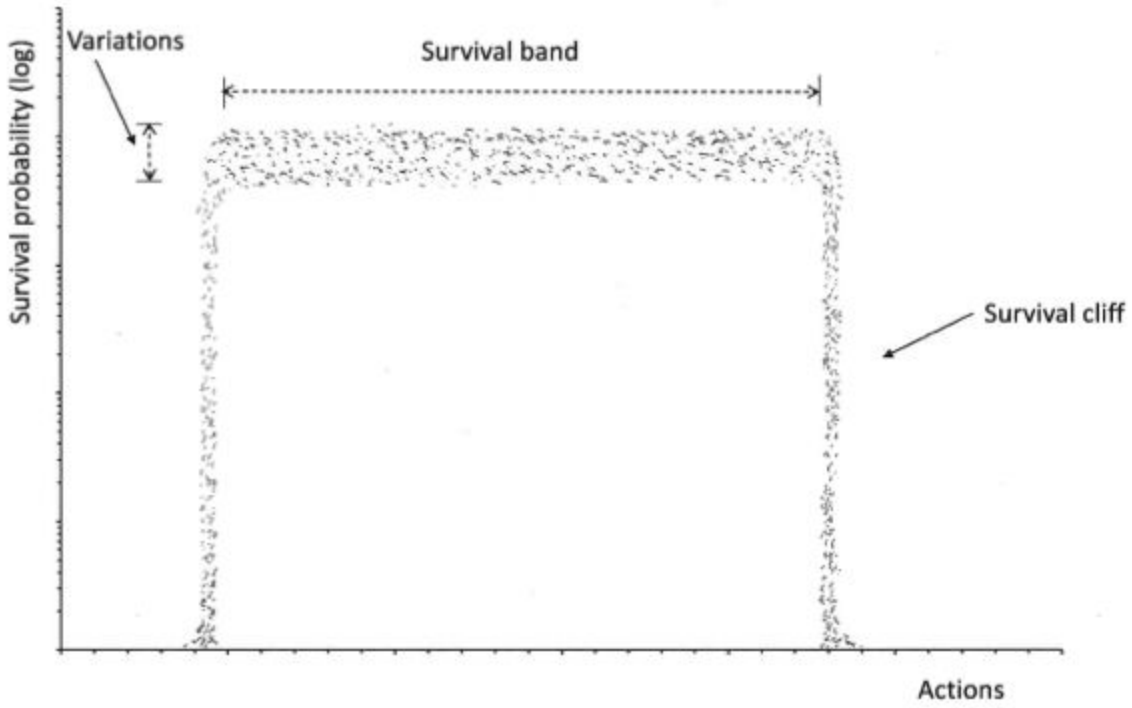
[5] To be more accurate, we should also exclude dormant time, only counting the time the species actively interacting with the environment.

In this sense, the more advanced the intellects are, the better they can predict the future, and the more likely they can survive longer.

The survival band

Depending on the evolution stage, a living organism may make zero to trillions of actions in her lifetime. Each action may impact the probability of survival. If we consider the unlimited actions a living organism can make under unlimited environment scenarios[6] are just dots in an unlimited dimension space, and we find a transformation to map those actions to two dimensions, action and environment, we can plot their impact on the long term survival probability in a 3-D figure.

Here, however, because I'm not experienced in making 3-D figures in my mind 2-D, I make two 2-D figures, one is a snapshot along the action dimension, the other is the average along the environment dimension. It is not as accurate as the 3-D figure. But you can use your imagination to plot the 3-D in your mind. The plots are shown below.



This is the plot under fixed environment. The shape of the plot is a reversed "U" shape. A lot of the actions do not significantly change the long term survival probability[7]. They are just plateaued at some point, which I call it the "survival band". The "survival band" has certain

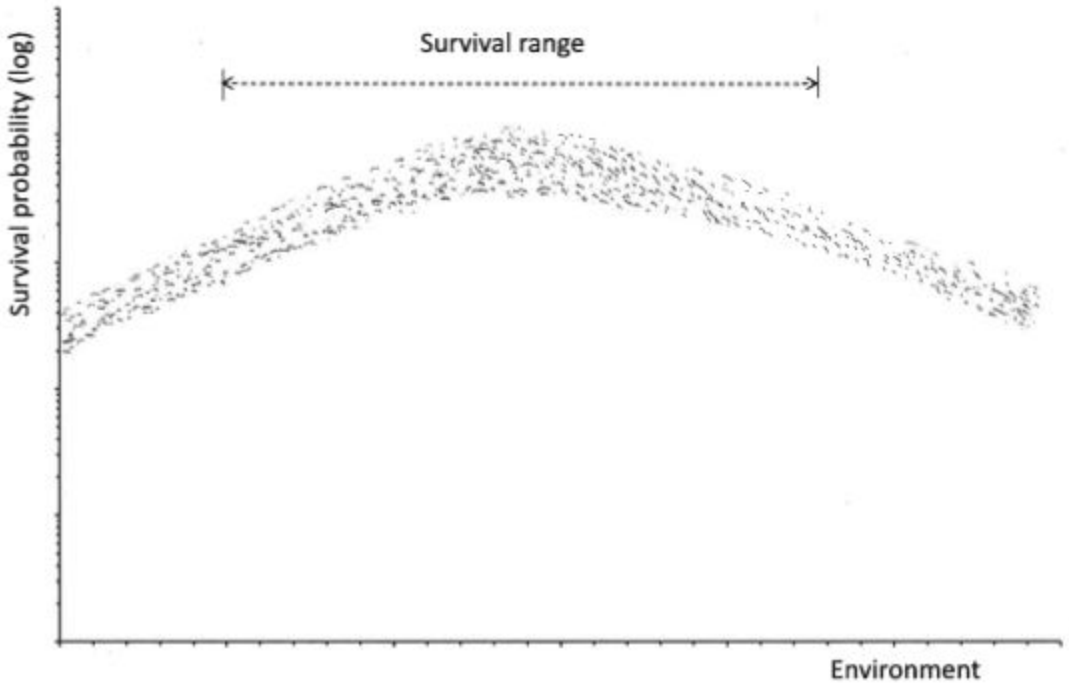
[6] Here the environment means the environment external to the species, which also includes the external environment the species proactively builds. However, it does not include the environment within the species, such as social environment.

[7] The impact of an action to the long term survival probability is the immediate impact of the action plus the discounted impact of all future possible actions. The second term is usually determined by the structure of the living organism.

thickness, which is the variations of the impact on the actions. A living organism can improve the impact of the actions on the survival probability by learning within the variation.

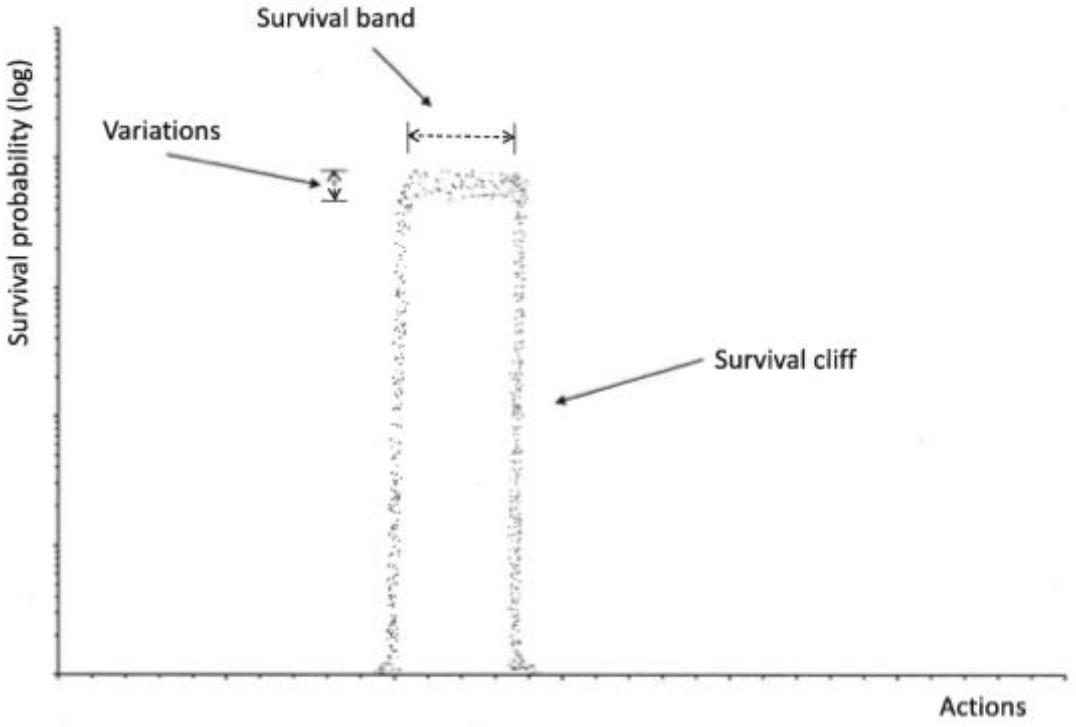
However, actions outside the survival band significantly reduce the long term survival probability, for example, jumping from a 7-story building. They fall into the survival cliff.

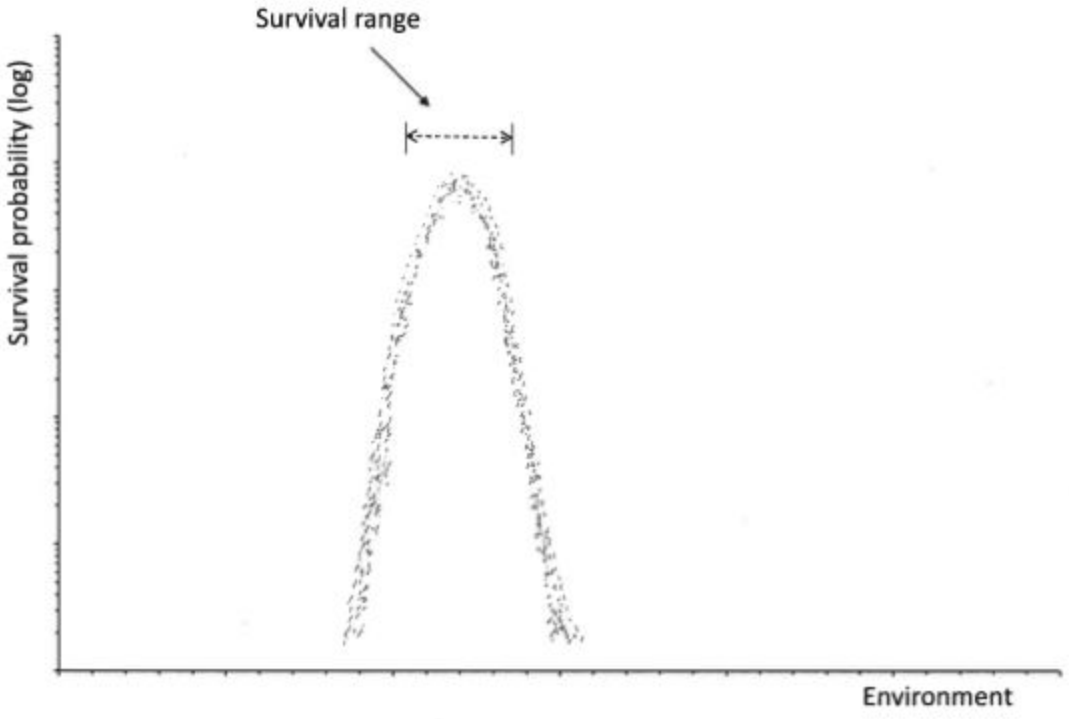
The plot along the environment dimension and average under all possible actions is shown below:



The plot is asymmetric bell shaped, which is quite obvious. More actions are effective under desirable environment, but less actions are effective under hostile environment. Let's define the width of the bell "survival range".

Let's assume the plot above is the plot for humans, which represents an S3 survivability. The plots below are plots for trees, which represents an S1-S2 survivability. What is different?





There are three notable differences:

- The width of the “survival band” is less. This is because a tree can perform far less actions than a human. A lot of the actions are baked in the tree’s gene and the tree cannot choose at all. The width merely indicates the variation of the tree’s genes.

- The depth of the “variations” is also less. This is because the variations of a tree’s action is much less than a human (as a tree’s behavior is indifferent or reactive to the environment). Thus, a tree cannot learn as much as a human.
- The width of the “survival range” is less. It is much easier to knock a tree out of the survival range than a human. For example, in case of a severe drought, a tree’s best self-saving action is to conserve water, when the drought persists, the tree dies. A human, however, can move to a more habitable place, and survive. As you can imagine, the bell curve is flatter when the intellect is more advanced, and when an intellect reaches S7, the curve is essentially flat at infinity (essentially there is no curve).

Please note, it is a mistake to consider the 3-D shape plot to be similar to the smooth surface bell shape. The “survival range” described above is just an approximation by averaging over all possible actions. Some action may work very well for one or a few environment scenarios but work poorly for the rest of the scenarios. The 3-D shape is NOT convex.

The advantage of the intellects in S3 and above is that they can quickly navigate through the non-convex shapes and find the best actions to take for the environment, which makes them survive longer.

Common ancestors

Some not so recent study have found that all men living today have a common ancestor, and all women living today also have a common ancestor. It doesn’t mean they are Adam and Eva. They are separated by tens of thousands of years. Human has 23 pairs of chromosomes, we may find 47 common ancestors, one for each chromosome (46 X + 1 y), each living in a separate distant past[8].

This means, all the people living at the same time as those 47 persons do not survive one of their chromosomes. If we generalize this, we may get the conclusion that:

> The majority of the people’s gene do not survive over time.

This seems to be in contrary to the Darwin’s natural selection theory, in which evolution happens at tiny steps, each step is selected by the environment.

I personally believe natural selection alone is more suitable to explain the variations of geologically separated species, but not to explain evolution across stages (e.g. S1->S2, S2->S3)

[8] I’m not an expert on this and I have not done any research on this. It is a mere extrapolation of the findings. It may not be correct, but it doesn’t fundamentally change the argument.

In AI, such phenomenon is called diminishing gradient. The gradient change cannot survive a few layers. However, in AI, we can use various methods to retain the gradient distribution across all neurons in the same layer (or even in different layers). However we cannot do that on human.

The role of mass extinction

In a stable, slow changing environment, the living organisms explore every niche of the environment and take every advantage to their benefits. If there is a small hole in the environment, it is quickly filled by the living organisms. In such an environment, all species live in harmony. The food chain is perfectly balanced, and no real cross stage evolution can ever happen.

This precisely indicates that the living organisms overfit the environment. That is, some features are developed precisely for the environment and cannot generalize well. Those features become useless when the environment changes.

The inter-dependency among the species are hard to break without external force. Any external force makes a hole in the ecosystem and ripples through the entire system. Each of such event comes with the extinction of some species, some adjustments to the rest of the species and more importantly, the emergence of some new species.

Thus, I believe, the continuous mutation without the nature's premature selection in the early periods after the creation of the hole in the ecosystem is the key to cross stage evolution. Let the mutation fly!

The natural selection, however, is only applied after the hole is roughly filled, resources become limited, and the competition among the species become obvious. It is also the time that the exploration of the species has ended and the new surviving species are squeezed by the natural selection.

The size of the hole in the ecosystem correlates very well to the steps of evolution, The maga holes that eliminate the majority of the species (especially the dominant species) represent the greatest opportunity in the leap of evolution.

The earth has gone through five mega extinction events. Each annihilates 70%+ of all species, including the dominant species. Each gives birth to the next dominant species.

And we are undergoing the sixth maga extinction event. Are we, the humans, the cause? Will we, the humans, be the victim?

I believe the answer to both questions are probably yes.

And I believe this will be the last extinction event that affects the dominant species.

Put aside the possibility of an AI dominance, the species in S4/S5 levels live in a self-created environment, and the species in S6/S7 level are ascended. Thus, human, or human equivalent S3 level species is the highest level dominant species seen on earth.

> *As an analogy, in AI, the extinction-recover process is similar to the squeeze-expand process in the SqueezeNet. The squeeze step reduces the capacity of the layer and squeezes out the overfitting, the expand step increases the capacity of the layer and fill in the blanks introduced by the squeezed net (which may introduce some features not directly inherited from the source). It enriches the subtle features being squeezed out by the previous layer.*

A Universal Model of Evolution

Below I explain how I view evolution in detail

An artificial intelligence view of evolution

The process of evolution is very complicated in real life. It is very easy to get distracted by some local variations. I believe the artificial intelligence resembles any kind of intelligence, so we can find out the essence of AI and generalize from that.

If we examine the recent hugely successful deep learning, we may find that the mathematics backing deep learning is just the very simplistic back propagation based on the stochastic gradient descent (SGD) and the chain rule. Every time the weights and biases are updated by the learning rate times the delta of the gradient. It is simply a greedy approach, possibly with some overshoot[9].

Given an architecture, starting from a purely random initialization, using a simple greedy approach, which just tries to optimize based on the neighboring descending slope, the algorithm can achieve near optimum result. This process is the learning process, which seems very straightforward.

The current research mostly focuses on finding better architectures that can achieve better result more efficiently. Currently there is no theory guiding the architecture search. The search is mostly done by intuition and rule of thumbs. This process is the evolution process. It is orders of magnitude more difficult.

In summary, the evolution model consists of the following two steps, which matches nicely to the metrics described in the beginning of the stages of the evolution.

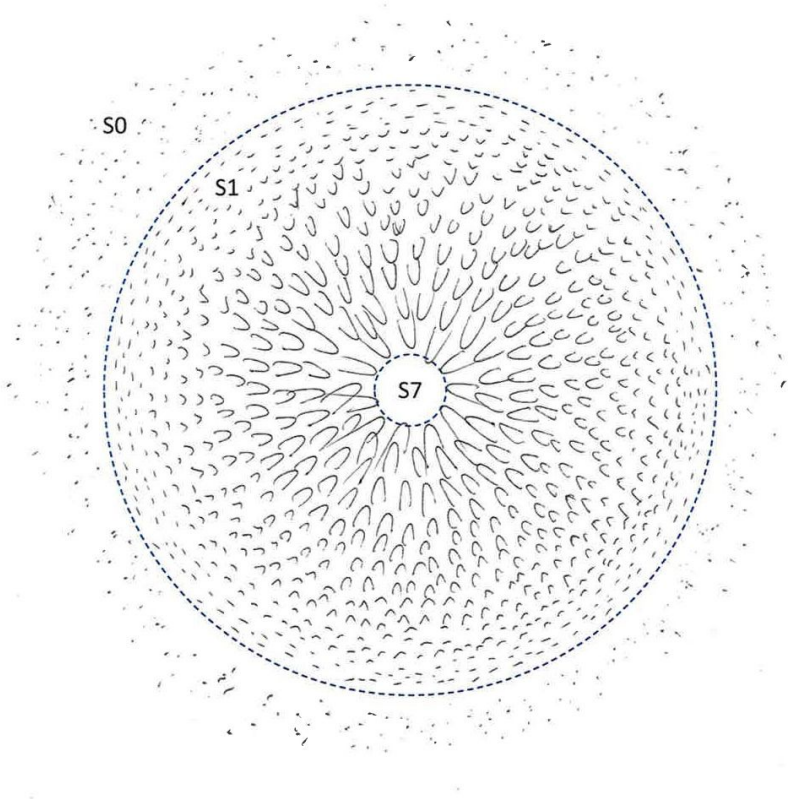
[9] The overshoot makes the learning zig-zag. I'm not sure whether the overshoot is essential in learning. Mostly like it is an inefficiency factor instead.

- Obtain the most efficient implementation for an architecture (the learning process).
- Search better architectures (the evolution process).

Correlate this back to the stages of the evolution, we can find that the 7 stages can be divided to three groups based on the methodology of searching better architectures: a purely random process (S0), permutation and random mutation induced process (S1-S3), and learning directed process (S4-S7). Each group may be further divided based on the learning capability.

The evolution process

The evolution process can be illustrated in the following figure:



Initially the process is purely random, as shown as dots outside the outer circle, which is most in number. Gradually, the process is more directional, as shown by the “U” shapes between the two circles. The “U” shapes indicate the learning capability of the architectures. The greedy algorithm pulls the architecture to the bottom of the “U” shape, but it cannot pull any further because of the restrictions of the architecture. Close to the outer circle, the “U” shape is really shallow, indicating that the architecture limits the ability of learning. Close to the inner circle, the “U” shape is really deep, indicating that learning can significantly improve the efficiency/quality of the architecture. In some cases learning changes the architecture, which essentially merges multiple architectures to one. The number of “U” shape architectures becomes less towards the center, indicating the consolidation of the architectures and fewer variations.

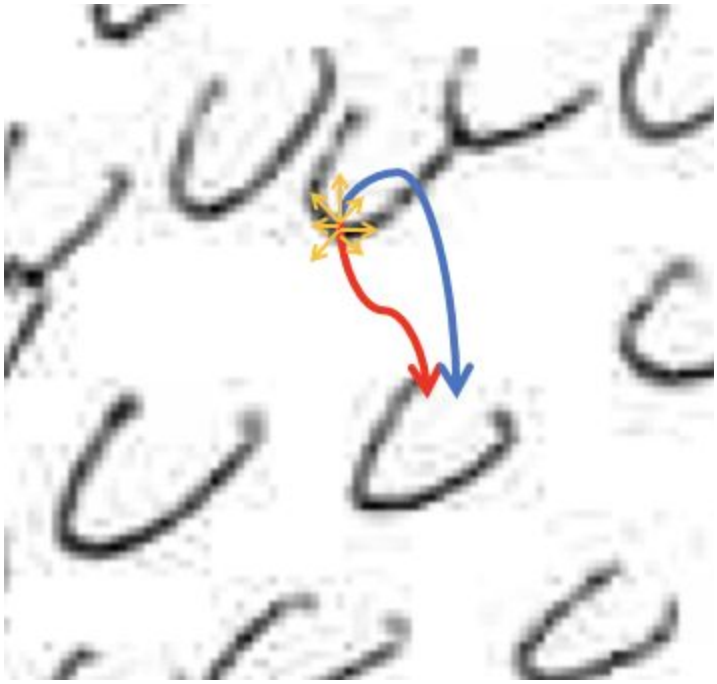
If we mark every dot in the figure with a number indicating the efficiency, the number in S0 is the lowest and the number in S7 is the highest at infinity. The number in the border of the “U” shape is higher than the number within the “U” shape. The number is reverse proportional to entropy.

I believe the direction of evolution is purely to improve efficiency, and the driver is the survival of the fittest.

The survival of the fittest, revisited

In the previous section, we have discussed the thickness of the survival band, indicating that the thickness is the result of the variations of the actions. Now we can say that the thickness is the same as the height of the “U” shape, representing the learning capability of the intellect.

Sine learning is easy, every species is in the bottom of the “U” shape. The species randomly explore from there in all directions, as the orange arrows in the following figure indicates.



The mutations striking through the bottom of the “U” shape is strictly better, and thus most desirable, as the red arrow in the figure shows. However, the fact that the species settles down to the existing “U” shape instead of the mutated “U” shape indicates that such mutation itself is very difficult. There is a thick barrier to strike through.

A more likely path is the blue line, where the initial direction of mutation actually degrades efficiency, only later the mutations find a least resistance path to leap to a more efficient architecture. This path, however, is unlikely in a harmonic environment.

In a well balanced, harmonic environment, the constant force of the natural selection would eliminate any slightly inferior mutation, and thus prevent the first portion of the blue arrow from happening.

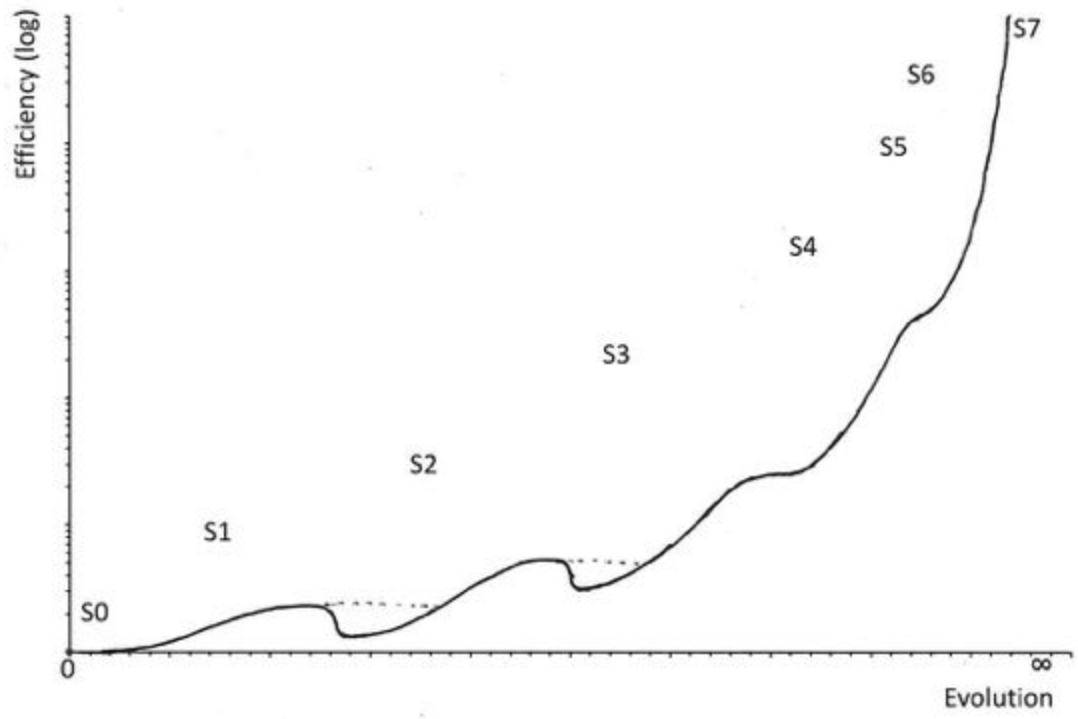
In the event of mega extinction, however, the force of the natural selection is very weak because of the huge hole in the ecosystem. Anything filling it up is good. Only in the second portion of the blue arrow, the force of the natural selection becomes visible, which pushes the surviving species to some more stable state, which sometimes settles to something more efficient.

Thus both the lack of natural selection in the beginning and the application of the natural selection in the end are essential to cross stage evolutions[10].

The evolution of the dominant species

The dominant species is special because it is the only species that can achieve the maximum efficiency. All other species have to find a niche under the shadow of the dominant species.

In the following figure, I hypothesize the evolution of the dominant species in terms of efficiency.

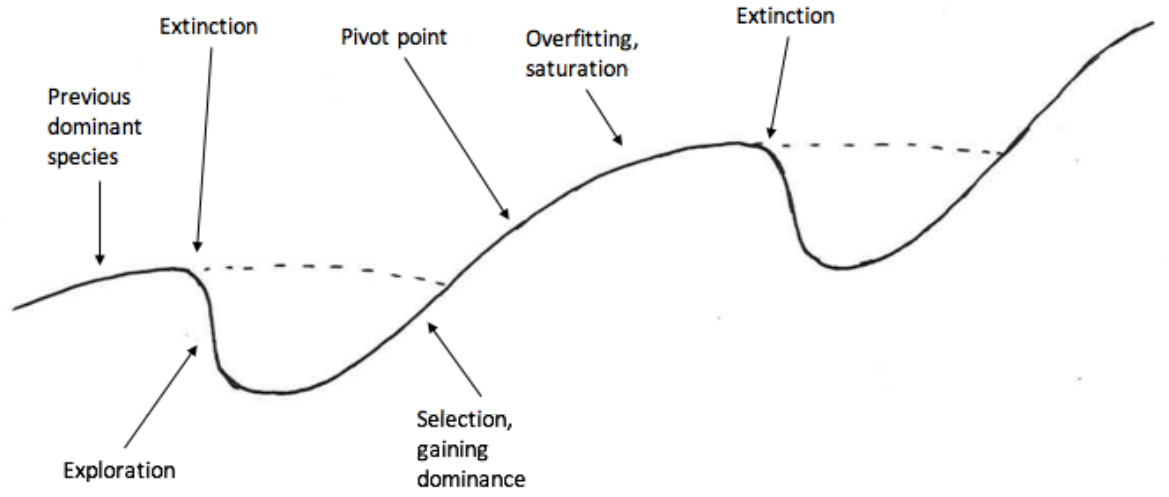


[10] In this sense, we should apply the same methodology to guide the development of the artificial intelligence.

S0-S3

We have experienced S0-S3. In S2 and S3, the efficiency first drops and then recovers. This is the mutation of the species after the global extinction events exploring more suitable architecture. The evolution is a number's game.

The following is a zoom in of the efficiency vs evolution.



When the previous dominant species goes extinct, the ecosystem is left with a huge hole. Other species, being less efficient, try to fill up the hole. This is the exploration stage. Many species try to find the seed of a more efficient architecture. After some time, when the hole is roughly filled, the competition among the species heats up, competing to be the next dominant species. The efficiency is consolidated at a new (higher) ground. One species eventually wins over and gains dominance. Its efficiency improves at an accelerated pace. Later on, however, after the species passes over the pivot point the pace of improving efficiency slows down, and eventually the species is limited by its architecture and efficiency is no longer improved. In the meantime, the species overfits the environment, taking advantage of every niche to its benefit. The species's dominance highly relies on the environment and cannot generalize well. When the environment changes, the species is very likely to become extinct and the cycle repeats.

S4

We may not see a drop in efficiency in the transition from S3 to S4. We may still see a drop. It depends on which route we eventually take. Put aside the possibility that the AI takes over the world, I feel this transition is nevertheless the bloodiest transition ever.

In the previous transitions, the global annihilation events kill most of the species in one or several generations. The pain is quick and equal. Most importantly, they are not as conscious and do not possess a will. After the extinction of most species, the surviving lucky few enjoy the privilege of filling the hole in the ecosystem.

The transition from S3 to S4, however, may last a really long time. It is the first transition that is directed by the species itself. The intellects, human or not, make changes to the chromosome, and set the course for the offsprings. However, when they set a course for many future generations, they do NOT have the capability to foresee the impact of the course many generations in the future. Thus, even though they can direct the evolution, the direction they set are random or pseudorandom.

When those beings direct their offsprings to different directions, their similarities diminish from generation to generation very quickly. We will see an explosion of highly intellectual species and then, depending on the resources, war is very likely triggered among them. It is possible that the evolution ends here and no more intellects would exist on earth.

If they are luckily survive, they will see competitions among the species for many generations. What is grindingly painful for some highly intellectual beings is that they cannot keep up with the level of efficiency of other species, because of the course set up by their ancestors many generations before. At that time their ancestors had no idea, but now they cannot change the course because that would make them in worse position sooner. They are in a hopeless situation. How painful is that! The competition among the species serves as the nature selection.

The number of species quickly reduces. the species form societies[11]. Their competition may last throughout S4.

This is what I envision will happen in the next big leap of evolution.

S5 and above

In S5, when they successfully combat natural death, intellects would get away from each other and find their peace alone, meditating. In S6, they ascend, find their own private yard, and never come back.

Evolution Recap

Up to now, we've only talked about evolution. I have spent great effort talking about evolution. What have we discussed about?

- The criteria of evaluating the evolution and thus, the seven stages of evolution.
- Survival of the fittest is the driver of evolution.
- Intellects live longer by predicting the future.
- Intellects evolve to be more efficient in the past, present and the future.

What is the one take away from evolution? Regardless of the variations and purposes, the end result of the evolution of the intellects is to:

[11] Details see "Society and civilization" section.

Be more efficiently predicting and approximating the environment.

Now let's examine some fundamental questions from this angle.

Causality and Science

When we see a phenomenon, we ask why and want to find out the cause. There is cause, there is effect. This is how the world works.

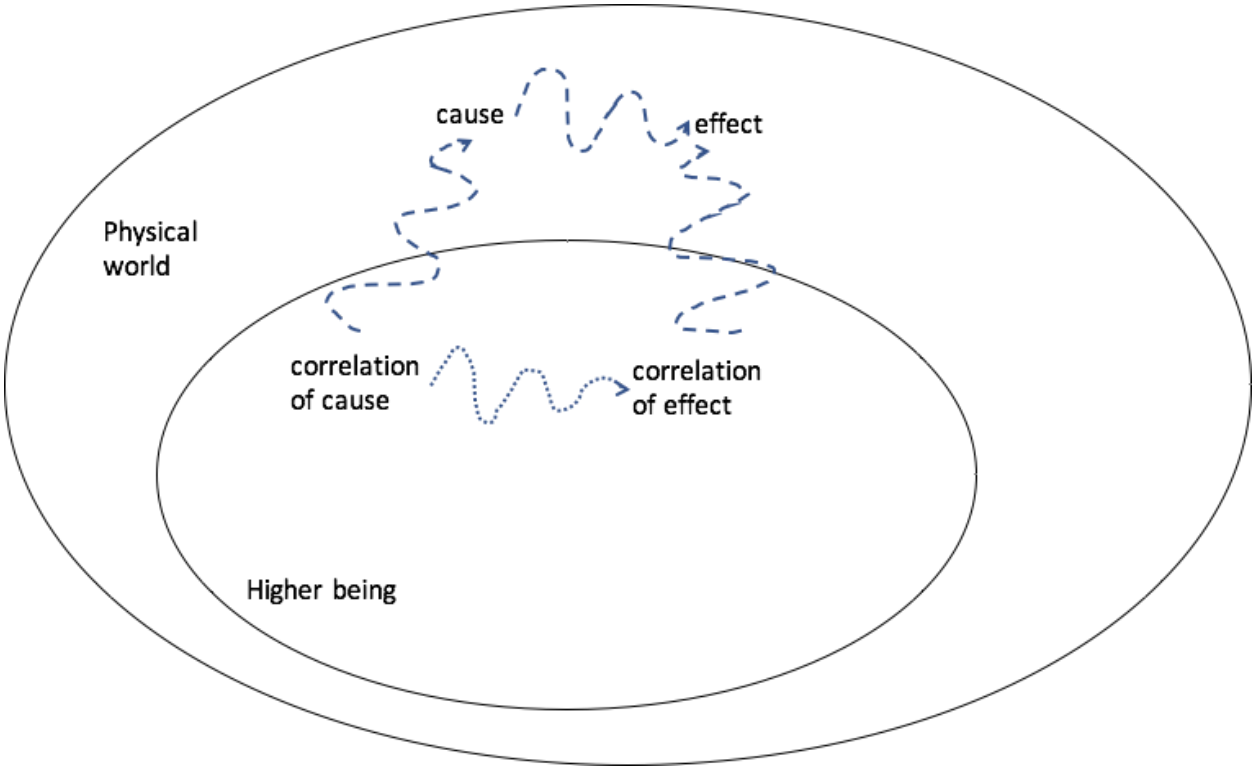
Is it so?

Let's examine the artificial intelligence. It is just the result of the back propagation of some cost function. There is no cause, there is no effect, there is only correlation. Surprisingly, it explains some aspects of the world quite well.

Intellects try to approximate the environment. Do they try to approximate by correlation, or causality? Put it in some other words, is causality just an efficient representation of correlation?

Intellects try to approximate the environment. Do we know that the world we are familiar with is the actual physical world? Or it is just an approximation from some higher beings? If it is the latter, can we know the real cause and effect? Or, we can only know the correlation to the real cause and effect?

The following figure illustrates the problem.



A higher being, living in the physical world, sees a cause, and sees an effect. Being an intellect, she tries to approximate the physical world, so she forms a correlation to the cause and a correlation to the effect in her mind. We, living in the mind of the higher being, see the correlation, can we find out the real causality? Or we will mistake the correlation to be the real one?

Science is the art dealing with causality. But it is both relative (general relativity), and uncertain (quantum mechanics). What if the causality we see do not exist? Science can still find some efficient relations, but it will never find the real cause and effect. This echoes what I have described ten years ago: science cannot reveal all the secrets in the world.

> *If we compare science and ML, we may find a lot of similarities. ML tries to approximate the environment; science tries to explain the environment. The objective of ML is to increase accuracy and reduce complexity. the objective of science is also to increase accuracy (ideally 100%) and reduce complexity. ML has overfitting problem, the number of parameters is reduced to reduce overfitting; science has Ockham's razor, the number of assumptions is reduced to form better theories.*
>
> *Currently ML is still based on induction. But science is mostly based on deduction, starting from the first principles. When the number of parameters is reduced to minimum, induction becomes deduction. This is an indication that ML has a lot of potential to improve. When will ML grasp deduction?*

Society and Civilization

Intellects approximate the environment. One key characteristic of such approximation is that the intellects are inside the environment. Thus, the approximation is:

> A lower capacity subject tries to approximate a higher capacity subject.

To approximate more efficiently, the intellects need to abstract more efficiently than the environment. In addition, the intellects may increase their capacities.

A straightforward way of increasing their capacities is to use the power of the many. A group of intellects collaboratively approximate the environment, each approximating a subset of the environment, and collectively they can approximate the environment more accurately together.

And this forms the societies and civilizations.

S2

At S2, the intellects reacts to the environment. The environment is in the dominant position and the intellects try to survive the harsh environment. Each individual may be very weak, but a herd of them together is formidable. There are countless examples in the nature, such as bees, ants, zebras, and wolves.

Some animals however, act alone, and rely less on their peers to approximate the world together. Instead, they have the notion of territory and try to limit the environment they need to approximate in order to be successful. They attack the approximation problem from a different angle.

S3

The fundamental inefficiency of intellects in S3 is the gap between the knowledge learned in one's lifetime and the knowledge passed between generations. Thus, external means are relied on to pass such knowledge from generation to generation, first by word of mouth, then by common written language. Societies and civilizations are formed as a result.

If we view the societies and civilizations super intellects that have longer lifespans, we may find that they are more efficient than their individual composing intellect. We are just tiny components in a big society and approximate a tiny fraction of the environment.

A society is most stable when its internal bindings among its composing intellects is much stronger than its external bindings to the environment as well as to other societies. The S3 intellects can proactively change the environment. Thus, their relations to the external

environment is not the most important. They may build an artificial environment that the intellects mostly deal with the artificial environment and thus further strengthens the society.

Of course, societies have internal frictions. Not all intellects put effort in the same direction. That is the inefficiency of a society and has to be dealt with by training, evolution or even revolution. We have seen countless examples in the past. When a society's internal structure cannot keep up with the advancement of the productivity, the society is due for a change.

The S3 Society, though much more efficient than individual S3 intellects in approximating the environment, is just a bandage to the fundamental problem for S3 intellects. A final solution has to come from within rather than from the outside.

S4

The S3 intellects are generalists. There is only one species (human) and everyone is more or less the same. This is because their evolution is still reactive to the environment and sufficient number is needed to ensure a sustainable evolution. However, generalists are not efficient enough.

Not so much for S4 intellects. S4 intellects are specialists, specializing one thing and one thing only. Since they can generically change their offsprings, it is easy for them to specialize one thing for many generations. This change has profound impact on the S4 societies.

The S4 societies are made of several to many complementing species coexisting together, specializing different aspects. As an analogy, the society structure is very like a hive structure, composing queens, workers, drones etc. Just each individual is a million times smarter. Because they can so easily change the external environment, they may live exclusively in their artificial society. It's time for many of them to leave mother earth and explore the universe.

S5 and above

When an S4 species has learned the secrets of overcoming death, there is less of a need for it to stay in an S4 society. I see two quite different outcomes from here on.

They may leave the S4 society and explore to approximate the environment alone, as they don't die and have unlimited time to learn to approximate more efficiently. As a result, the rest of the S4 society fade away since they coexist and losing one species is devastating to the entire society.

A more likely outcome is that the S5 intellects teach the rest of the S4 intellects in the society to conquer death. Thus, the entire society is evolved to an S5 society. It is more efficient for them to approximate the environment. However, since they do not die and they coexist, there is no need to call them separate species. They are all components of the same species. Later, when

the S5 society approaching S6, there is no difference between a society and an intellect. All previously S5 intellects can be viewed as components of a larger intellect.

Yes, an S5 society will be evolved to an S5 intellect.

And after that, there will be no more society.

Society, Nation, and Religion

Society, nation, and religion all try to align people's thoughts. Society does so by culture, virtue, and ethics. Nation does so by law enforcement. Religion[12] does so by commandments and religious disciplines. In any case, they persuade people not to explore some action space.

Some actions are essential to the survival of any person, such as breath, eat, and sleep. Those actions are baked in our DNA and we are born with that. However, not so much for many other actions. For example, jumping from a 7-story building is obviously bad. But it doesn't happen frequently enough to embed it inside our DNA. Thus, society, nation, and religion all serve the purpose of minimizing those actions that threaten the survival of the species or a group of people.

Humans, as S3 intellects, are generalists in nature. Society, nation and religion cannot change that physically. However, they can change what people think, and specialize people's mind. In this sense, their existence is an indication that we are evolving from S3 to S4. Even though we are not ready to do so physically, we are already on the journey mentally.

When evolving from S3 to S4, one primary direction is functionality. The high beings are specialized to perform one function more efficiently. Another primary direction is ideology, resulting from the residues of the society, nation, and religion, which may or may not have practical use cases.

I envision that such specialization is unique for S3 intellects. S4 intellects bake such specialization in their DNA, thus rely less on external forces. In this sense, the specialization purpose of the society, nation, and religion is most useful in S3, and less useful in S4 and other stages.

[12] Here I only discuss one practical implication of religion. I do not discuss whether the religion is real or not. Some religion may be really started from messages of S7 intellects. Some religions may be founded from shaky grounds. But their implications are the same.

Consciousness

Conscious is a vague term that is very difficult to define. In wikipedia, conscious is defined as the state or quality of awareness, or, of being are of an external object or something within oneself.

More recently, scholars link conscious with the prediction of future events:

> In the keynote speech of the Hotchips conference, 2017, Phillip Alvelda described what he believed was the root of consciousness[13]: the future prediction of your conscious state.
>
> In the note “The consciousness prior”, Yosha Bengio described the assumption that a conscious thought can encapsulate a statement about the future.

I agree with them, and consider consciousness to be an efficient mechanism to predict the future. The survival of the fittest pushes to the emergence of conscious beings, as I have explained before. So, consciousness is one inevitable step in evolution. From my explanation of the steps of evolution from S0 to S7, the higher the level, the further in the future can be accurately predicted, and thus, the more conscious the beings are.

Prediction is the objective, and conscious is an efficient way of achieving the objective. However, it doesn’t explain how to be conscious.

We are aware of the environment simply because we can compose/decompose our sense to the environment, which significantly reduce the amount of information requiring processing. Unconsciousness simply means the intellect is unable to compose/decompose the environment. With that amount of information, an unconscious being is unlikely to grasp the important, ignore the unimportant, and express the thoughts to the outside.

In my opinion, the ability to compose and abstract is the prerequisite to be conscious, simply because it is also an efficient mechanism to predict. Thus, if we want to make a conscious machine, rather than focusing on the vague term of conscious, we should focus on building a machine that can compose and abstract by itself. It will become conscious all by itself at some point.

> *Suppose an ML algorithm derives the correlations between two temporal and spatial separated events, at first as a black box with billions of weights, bias and MACs. This represents an unconscious level of understanding. We can decompose the relations into simpler ones with millions of weights, bias and MACs, treat them as first principle, and*

[13] https://youtu.be/PVuSHjeh1Os, 57:30 onwards.

then deduce the correlation from there. This is what consciousness is composed of. When observing the decomposed first event, the second event can be predicted based on the deduction alone with reasonable accuracy. This is what consciousness exhibits. This process can be repeated to reduce the amount of computation required by the first principle, and thus reach a more conscious state.

Free Will

According to wikipedia, free will is the ability to choose between different possible courses of action unimpeded. In fact, any will is free. There is no such thing as non-free will. But below I will use the term free will since most people like it.

Whether free will exist, the source of free will, and the free will's relations to the environment are all heavily debated topics in philosophy. Here I just present what I think free will is.

What is free will?

Let's consider the following hypothetical scenario:

> Three persons each is presented a two-to-one choice question. She must make a choice between the two options. If she does not make a choice, she will be executed.
>
> The options presented to person one are: jumping from a seven story building, or being locked in a house on fire.
>
> The options presented to person two are: jumping from a seven story building, or having a seafood dinner.
>
> The options presented to person three are: having a seafood dinner or having a steak dinner.

For person one, she must have projected that neither option is desirable, and she must be calculating which option is more likely to survive.

For person two, she must have projected that option two is much better than option one (except if she is strongly allergic to seafood). Almost certainly that she will choose option two. In her projection, the two choices are not comparable.

For person three, she must also have projected that both options are desirable. She then makes the choice based on whether she likes seafood or steak better. Maybe she will have follow up questions like what kind of seafood and what kind of steak.

This hypothetical scenario tells us that free will is not totally free. It is limited by the physical constraints. Person two is free to choose jumping from a seven story building. But if she does so, she is unlikely to survive and is unlikely to make any further choices, if any.

Thus, the desire to survive strongly influence the decision of a person. If a person chooses not to survive, she is less likely to have offsprings and pass down the choices to later generations.

This is a very extreme example. In less extreme cases, the choices are less obvious. But still, the rule of the survival of the fittest still applies to free will.

In addition, let's revisit what the consciousness is: consciousness is the ability to predict the future state of the agent. In this example, the three persons, when presented the options, all project what the outcome of each choice would be. Their choices are based on their projections of the outcome, instead of the actual outcome.

However, we have to admit that no human being can predict the future perfectly. Thus, anyone's projection is limited. In one's foreseeable future, if the outcomes of the choices are obviously different, she will make an obvious choice. If the outcomes of the choices are about the same or hard to tell, she is difficult to make a choice. If the outcomes of the choices do not turn out to be the same as she projects, she is likely to make a wrong choice. In any case, she makes a choice using her free will based on her projection.

Let's combine the survival of the fittest and the consciousness. This is what I consider free will is:

> Free will is the perception of choosing from multiple choices of about equal consequences, resulting from the agent's inability to predict the future accurately. It is a means of performing random exploration with limited playouts. Survival of the fittest favors the free wills that make more efficient choices.

Suppose all intellects have a common goal, but the goal is so far fetched that no clear path exists (as far as the intelligence's capacity concern). With the cloud of possibilities, the intellects need to pick a path to the goal. But due to the capacity restriction, the path is unclear and can only be optimized by local constraints and the consequence is unclear and may appear to be random. Free will is an efficient way to explore the path. It is difficult to ask the ultimate goal because it is far fetched. When asking why a particular action is taken, it can be explained from the local optimization objectives, mostly by composition from the first principle. The scenario that multiple paths seem to have comparable consequences gives the perception of a free will.

I believe the common goal is to live longer. In other words, all intellects want to be S7 intellects. Nobody knows how to reach there, and free will is an efficient way of exploring the search space to reach the goal. At last, survival of the fittest drives the direction of the exploration by favoring more efficient paths.

Irrational decisions

Irrational decisions are those made without reasoning, seemly obvious inferior to the decisions based on reasoning. Most of such decisions are made with subconscious. It is a representation of free will.

People make irrational decisions all the time. But such decisions are all within a bound. For example, jumping from a seven-story building is irrational, and is obviously out of bound. People doing that will be obviously penalized. However, when the irrational decisions are within the bound, many of such decisions are not penalized (at all). I believe that is mainly due to the following several factors:

- Humans cannot see far enough in the future, which result to many aspects being inefficient. In this case, knowingly inefficient in one area doesn't matter too much since many other aspects are unknowingly inefficient. Thus, in many scenarios people may recover from this knowing inefficiency sooner or later. This is however not true when the entire society and environment become more efficient.
- Irrational decisions are not made from projecting the future prosperity from the decision. Since a human only has one life to experiment on, when she makes an irrational decision, she cannot know how the life would be if she had chosen otherwise. To make a rational decision, a human needs to observe the lives of many other people in many different scenarios, examine the effect of the decisions on their future lives, draw a relations between them, and extrapolate to the decisions for her. This is a very difficult task for many people, and it is another indication of humans' inefficiency.
- The hardware structure of humans hasn't changed much in the past few thousand years. This is because our evolution is still reactive to the environment. First, a few thousand years is a very short period of time in evolution, and humans' inefficiency hasn't changed much. Second, with the common ancestor hypothesis, the evolution we experience is virtually zero. Plus, we haven't experienced any mass extinction. All indicate that the inefficiency is inherent in humans' structure and is unlikely to change.

Irrational decisions is an indication of imperfect evolution. I project, more advanced intellects make less irrational decisions.

Will Jane sit on the sofa?

In this section, I present a hypothetical scenario for an intellect called Jane. From this example we can see how Jane makes a decision from her free will.

Jane stands in front of a sofa, she can make a decision whether to sit down on the sofa or just stand there. The desire comes from Jane's inner free will. However, if she knows that sitting down the sofa would give her 0.0001% more chance of surviving an earthquake, everything else being equal, she would decide to sit down on the sofa. Because of the survival of the fittest, given enough time (many billion years?), the intellects deciding to sit down would survive more

than the intellects deciding to stand, thus giving an advantage to the intellects deciding to sit down. Thus, the appearance of Jane's free will is because she doesn't know the consequence of the action of standing besides a sofa.

However, Jane will have trouble deciding whether to sit down or not if sitting down on the sofa would give her 0.0001% more chance of surviving an earthquake, but standing would give her 0.0001% more chance of surviving a flood, assuming the likelihood of having an earthquake and flood are the same. However, knowing the consequence, Jane would not allow her to be in such a dilemma, she would not stand besides a sofa in the first place. The fact that Jane's standing besides a sofa is an indication that she does not know the consequence of the prior action of standing besides a sofa.

Thus, as Jane better understands the action's consequence to the future, the rule of the survival of the fittest governs that her decision choices become more limited. In appearance, her will becomes less free.

This is a tug-of-war between two forces. On the one hand, if Jane can correctly predict further in the future, she may encounter more uncertainty and she's unsure of the consequence of her current action due to the more uncertainty. As a result, she will have more options of about the same consequences to choose from. Thus, her will appears to be more free. On the other hand, if she can correctly predict further in the future to the extent that the future uncertainty is reduced, she better understands the consequence of her current action, and she will make informed decisions to maximize her survival. As a result, she will have less options of about the same consequences to choose from. Thus, her will appears to be less free.

Jane's decision is not only dependent on prior decisions, but is also dependent on her projection of the consequence of the possible future decisions.

It doesn't account the possibility that Jane could change the possibility of the earthquake or the flood prior her decision to sit or not. But it doesn't fundamentally change the equation. If, with her prior effort, there is no difference between sit down or stand in any respect (note, it is no difference, not no difference that Jane could project), it makes no difference whether Jane sits or stands. She could just flip a coin and decide, and gives the perception of free will. It is unclear, however, whether this scenario can really exist, or it is just an indication that Jane cannot see through the future far enough. Even if that is possible, do we still call it "free will", since nothing is changed over the long run? Suppose that is possible, and Jane is a GOD that can see perfectly to the end of time, the decision she makes may just be some residue factors formed before Jane becomes the GOD. This is because the evolution trajectory is already set before Jane becomes the GOD, and being efficient, Jane has no reason to change the trajectory after she becomes the GOD.

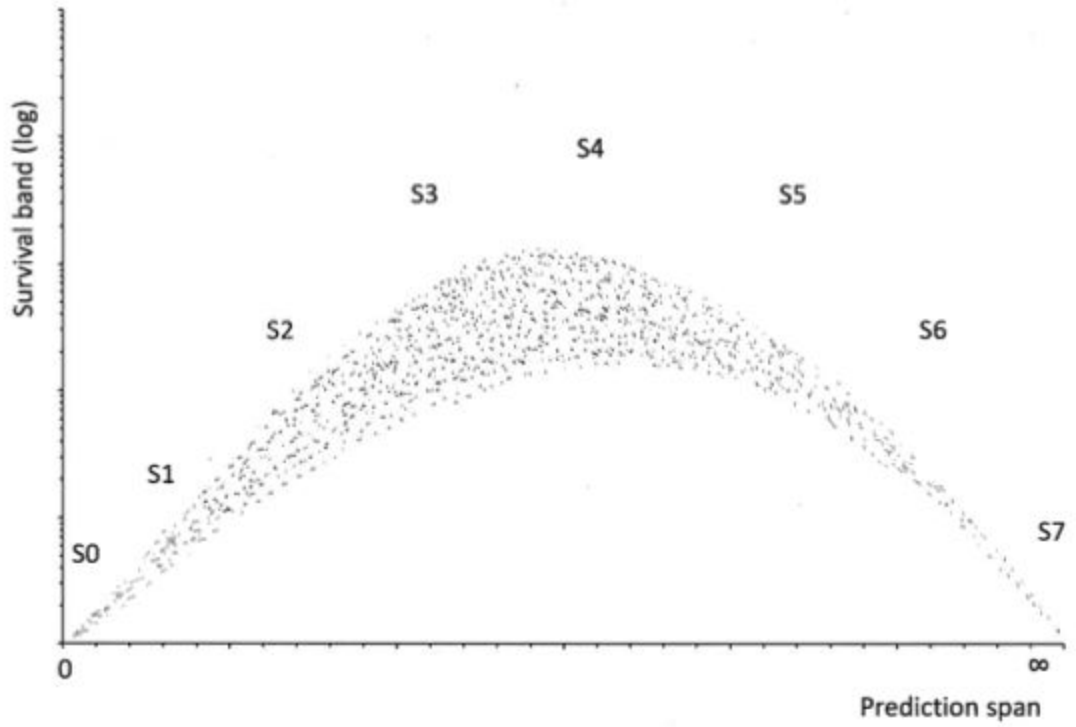
The evolution of free will

Let's conclude what is described in the Jane's example. When an intellect can predict further in the future, she is exposed to greater uncertainty, and she can pick from more choices with equal or about equal consequences in her foreseeable future. But at some point, seeing further in the future becomes a constraint that reduces the number of choices the intellects wants to make, because of the survival of the fittest, and the intellect wants to be the fittest. The will is then less "free". As an extrapolation, when the subject can foresee infinity in the future, there is only one way to achieve that, and the will is no longer free.

So, when is the turning point that the will becomes less free? I believe this depends on when the intellects cannot see the difference among most options. And that is somewhere between S3 and S4.

It is obvious that intellects from S0 to S2 do not have a lot of options to choose from, because of their limited prediction capability. Will is an indication that the action is proactive to the environment. In S3, we the humans are presented a great deal of possibilities, since we are actively changing the environment. We are also generalists, which means the possibilities can come from anywhere, to any person. In S4 however, intellects are more specialized in a narrow domain. Even though they can see much further in the future and encounter greater uncertainty. The greater uncertainty attributes to many S4 species in a society. Each species only gets a small slice of the uncertainty in their speciality domain. Their number of options may not be more than the number of options for S3. When they are evolved to S5, however, they may see an explosion of uncertainty, simply because a society is attributed to one species. I consider it just an accounting issue. At S5, when they see further in the future, the total uncertainty gets reduced. They may still not see more options of about equal consequences.

This is best illustrated in the figure below:



At last, let's make another analogy. In the go game, some positions are strictly superior to some other positions. In this case the player does not have free will, she will sure place the stone to the superior position. Unless she doesn't want to win the game, in which case she will be deprecated by the survival of the fittest. In some cases however, the player cannot read the board clearly and may feel that the moves on several positions are about the same and the player is free to choose which position to place the stone. Her analysis and free will decides which position to place the stone. This is often the result that the player cannot see the consequence of placing stone to the positions far enough to the future of the game. To a much stronger player, such as AlphaGo Zero, some of the positions may be considered strictly inferior than other positions and the stronger player would not place the stone in those positions. Thus, to a stronger player, her free will is less because she can see further in the future and some choices considered equal by some weaker players are no longer equal.

Like go, like life.

Where Are We?

The higher the level of the intellects, the better they can project the future. It means those intellects have a small model of the environment. The higher the level of the intellects are, the more accurate the models are. If they use MCTS (of course they are much smarter than that), they may playout their lives many times before they make a decision for their actual move.

Then here comes a very practical question:

> Are we in a real physical world? In a projection (simulation) of a God or higher being? Or both?

Similar questions have been asked again and again in the past. Some of the famous SiFi movies are based on such quest, such as the Matrix trilogy and the thirteenth floor.

This question may deserve its own article. But here are some hints from my perspective: start thinking from the assumptions of the fundamental physics (empirical observations), ponder why they are in place, evaluate various levels of abstractions from there, and compare them against the fundamental restrictions of projection (simulation).

Nothing is certain. But at least you can be more confident on your preference.

Or, does it matter?

Belief is everything.

V0.1 - Fei Sun, Dec. 2017 @ San Jose