

# A Theory Explains Deep Learning

Kenneth Kijun Lee Chase Kihwan Lee

Deduction Theory, LLC.

tf.kenneth@gmail kifhan@gmail.com

## Abstract

This is our journal for developing Deduction Theory and studying Deep Learning and Artificial intelligence. Deduction Theory is a Theory of Deducing World's Relativity by Information Coupling and Asymmetry. We focus on information processing, see intelligence as an information structure that relatively close object-oriented, probability-oriented, unsupervised learning, relativity information processing and massive automated information processing. We see deep learning and machine learning as an attempt to make all types of information processing relatively close to probability information processing. We will discuss about how to understand Deep Learning and Artificial intelligence and why Deep Learning is shown better performance than the other methods by metaphysical logic.

## Study Deep Learning from scratch

We are thinking about how to do programming better as this way.

1. Read and analyze well-written code.
  - a. Analyze it from my standpoint and annotate it.
2. Try porting the analyzed code to another language.
  - a. In this process, I verify that my analysis is accurate that what I understand the process of information processing in the program.
3. Based on what I have learned so far, I plan the goals of the software I want, and implement the information processing process in code that can achieve that goal.

We call this "The royal road of programming". We have used this method to steadily improve our programming skills. We are studying deep learning study in the same way.

Starting in early November 2016, we began analyzing and annotating single-perceptron and multi-perceptron implementation code.

Perceptron is the basic unit of artificial neural network. The source we analyzed was written in C++.

We spent two weeks analyzing and annotating, and then started porting. The language we chose to port to was "[Julia](#)". We took the time to discuss what to choose between [Golang](#) and Julia, and finally we decided to try Julia. The reason for this was that Julia was better at doing mathematical expressions.

### **The reason for we chose Julia as the Porting Language**

1. It is a relatively newer language.
2. Performance is fast.
3. It is convenient for mathematical calculations.

So we started doing that and in the second week of December 2016 we finished porting single perceptron and multi-perceptron code to Julia. We will share the code. There is also the original C++ source link in the github link.

[perceptron-jl - github](#)

## **The reason why we port it rather than learning by the reproduct with the framework**

We have been experimenting with the framework tools and sources for few weeks, starting in October 2016, with TensorFlow, Caffe, Theano, and Keras. I also reproduced the deepmind DQN source. Even so, I have not been able to understand the basic principles of deep learning. So, I chose the precision method to analyze it from the basic unit.

The reason we share our process is that we want to help people who want to explore the basics of deep learning in a similar way. We hope that you are interested and want to give us feedback. Please tell us any mistakes we made, or something to improve.

In the future, we will port and implement CNN, RNN, NLP, DQN, and other complex learning packages and organize them into documents. I hope to have an exchange with people who want to share the same interest.

Please check below to see our research journal we made.

## **Research journal**

As we work on this, we have learned to work on the mathematics that were unfamiliar, such as vectors, matrices, calculus, and topological mathematics.

We were studying logic and mathematics before studying deep learning and artificial intelligence. We were studying the subject under the name "Deduction Theory". If the word "Deduction Theory" comes out in the content to be followed, it would be easy to understand if you think, "Oh, this is the name they gave to the subject they were studying."

The word Deduction Theory could be not familiar who have a first time reading.

Deduction Theory is differ to deductive reasoning in conventional logic. I will explain how it different.

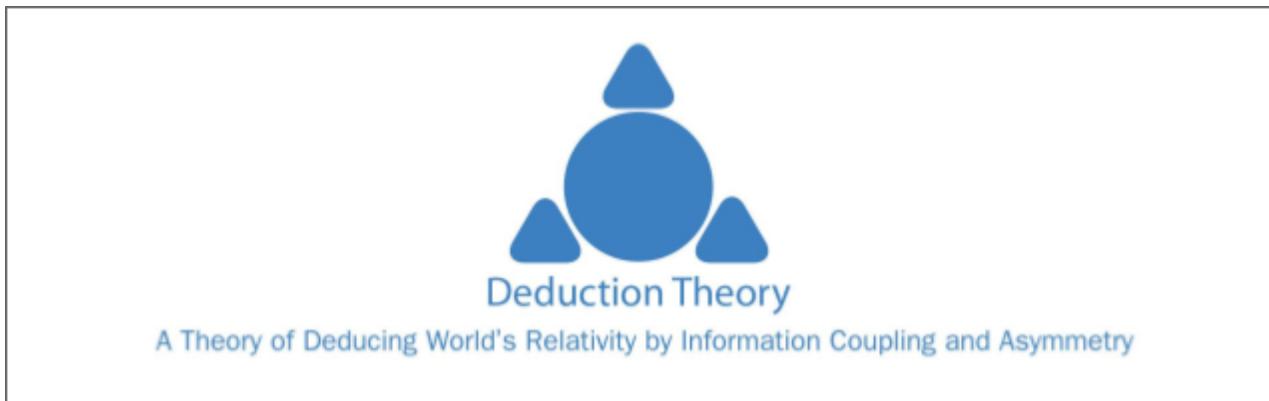
#### [Deductive reasoning \(Logic\)](#) - Wikipedia

Deductive reasoning links premises with conclusions. If all premises are true, the terms are clear, and the rules of deductive logic are followed, then the conclusion reached is necessarily true.

The deductive reasoning is based on a method of verifying the absoluteness of the sentence. It is a way of thinking that assumes that the relationship between truth and false is absolute and leads to premise and conclusion.

In our Deduction Theory, we do not assumes absoluteness. We use a word 'deduce' as a pure exploration of information.

Deduction Theory is a Theory of Deducing World's Relativity by Information Coupling and Asymmetry. This is our slogan.



Deduction Theory is an intellectual property created by Kenneth Lee. Deduction Thoery is a registered trademark of Deduction Theory LLC, in U.S.

We will explain this theory deeper in following journal.

The statements we make in our journals are not absolute facts, but are hypotheses we have come up with in conceiving a worldview that helps us study deep learning. Please give us any feedback if you understand so and have a reasonable opinion.

## What is Deduction Theory?

Deduction Theory is a way of thinking.

How does Deduction Theory come about? What is the process of thinking as Deduction Theory? In conventional logic there is an absolute perspective and a relative perspective about the process of information processing.

The perspective, the worldview, and the way of thinking are the same meaning in processing information.

Let's look at the absolute perspective.

An absolute perspective is that there is an absolute element, an absolute reason and an absolute basis for the process of the information processing. Let's take an example of this. In logic and mathematics, it is assumed that there is "axiom".

An axiom or postulate is a statement that is taken to be true, to serve as a premise or starting point for further reasoning and arguments. - [from Wikipedia](#)

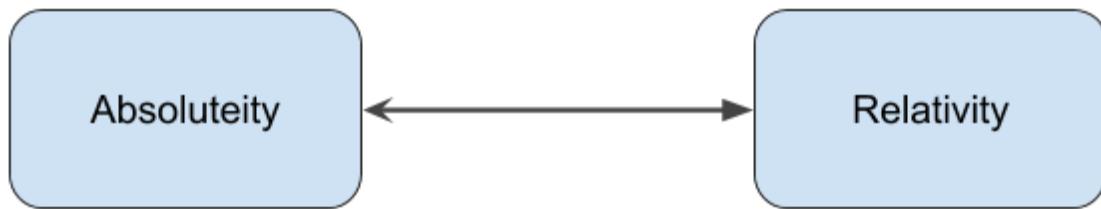
But it is a contradiction to say that there is an axiom, that axiom is an absolute logic that is true even if it is not grounded. Assuming that there is absolute logic, it means it's possible that there is no absolute logic.

We do not have to use this inconsistent statement and just say, "I'll show you one of the top rules to use for future processing of information. If the rule needed to change, we can make a new game. But in the game we process, the top rules remain their role."

They claim to make absolute axiom assumptions to establish the authority of logic and mathematics.

The reason why the controversial proposal for axiom came out is that the "Bertrand Russell" in the early 1900s had argued about the fictitious of axiom. After Bertrand Russell argues that an axiom are fiction rather than true, people say "Let's assume there is an axiom." than saying "There is an axiom".

## Conventional view of Absoluteity and Relativity



There is absolute logic behind.

Relativity

???

Before Russell, mathematician and logician claimed that there was an axiom. That means there was an absolute logic. Absolute means that you do not need to know the process how it was made anymore because it is there. Russell proved that an axiom do not established because "Russell's paradox".

[Russell's paradox](#) Wikipedia

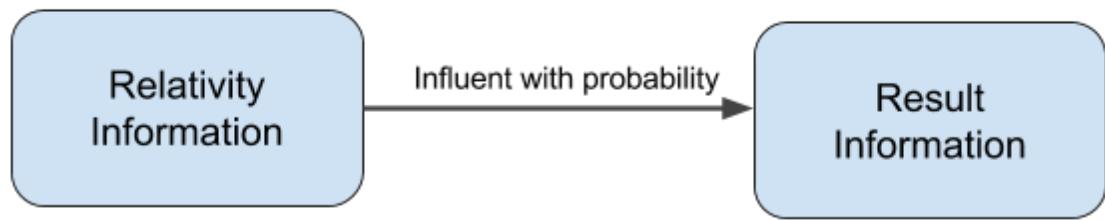
[Gödel's incompleteness theorems](#) Wikipedia

So what is the relative perspective of Deduction Theory? In Deduction Theory, we think there is no absolute logic, no absolute process of information processing, no absolute element, no absolute basis. Instead, Deduction Theory claims that if the relativity information as upper level rule is set, the result information can be probabilistically generated in a lower level and the information processing game can be operated.

## Deduction Theory's view of relativity



## Relativity Information and Result Information



The relativity in Deduction theory is different from the relativity that people claimed in history. So far, the relativity that people claimed in history, the relative viewpoint was the opposite of absoluteity and absolute viewpoint. But let's think. Absoluteity is a fiction. But how can the opposite idea can be established?

The relativity that old people thought was mainly disorder, chaos. But is that really so? Is relativity disorder, chaos, random? It's not like that.

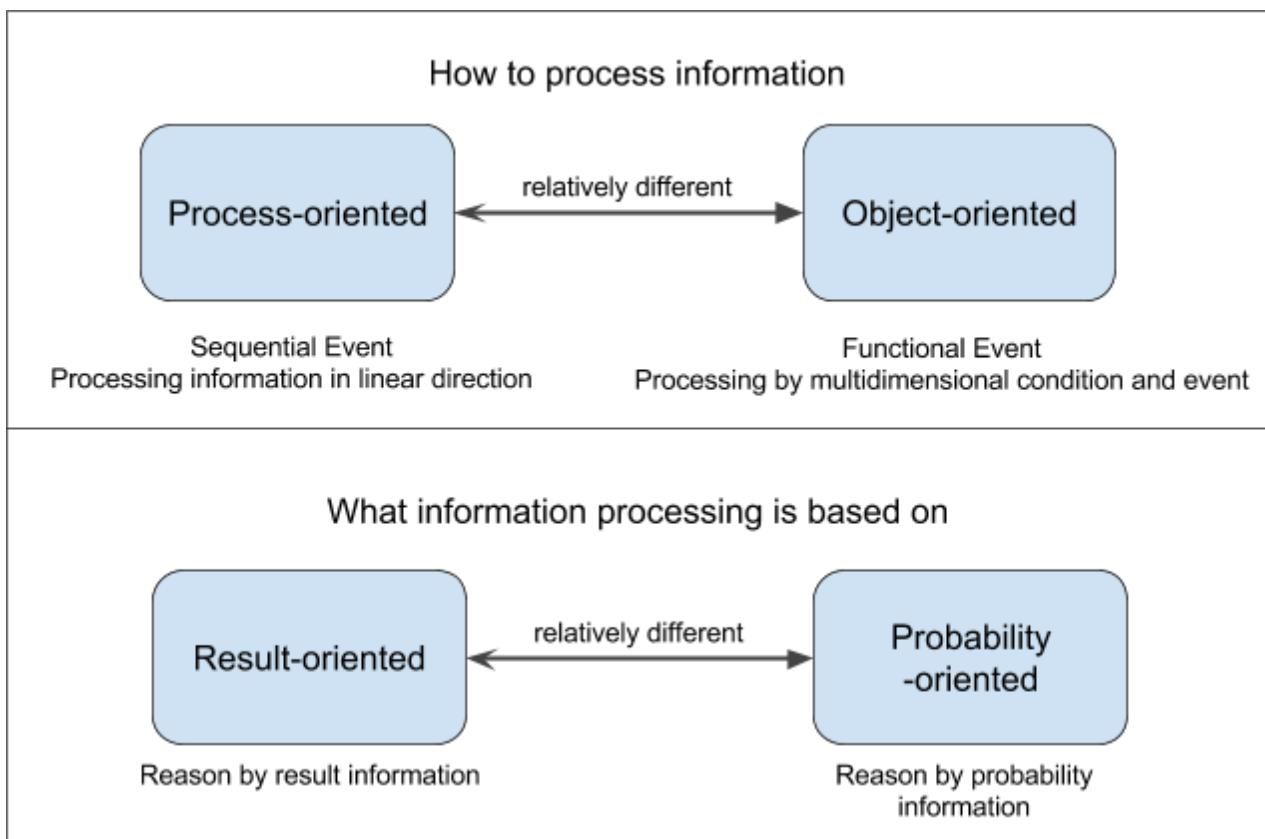
To understand deeper, we studied how process-oriented and object-oriented information processing are different, and how result-oriented and probability-oriented information processing are different.

Object-oriented information processing is relatively more relativistic than process-oriented.

This is because in process-oriented, an attempt is made to process information in an order that is absolutely certain, and when an input that does not correspond to the order comes in, or when a variable occurs, an error occurs and the program does not operate properly.

On the other hand, object-oriented program operate to response multidimensional conditions and events. Object-oriented programs are more likely to reduce errors. In fact, object-oriented programs are purposely developed to reduce the errors in procedural-oriented programs.

Let's Compare result-oriented and probability-oriented information processing. The probability-oriented information processing process is more relativistic. When performing result-oriented information processing, it is necessary to accurately specify the result information. If the condition is not correct, an error occurs. However, in the probability-oriented information processing process, probability patterns are deduced from a large amount of data and information processing decisions are made on the basis of it. For reference, result-oriented and probability-oriented are words made from Deductive Theory.



Object-oriented information processing and probability-oriented information processing are increasingly more "uncertain" because they are more relative? They are not. Rather, they increase the likelihood of reducing errors, increasing the accuracy of information processing. If you create a program with absolute perspective, the error is more likely to occur.

If you try to explain relativity as the opposite of absolutism, you try to interpret relativity in terms of absoluteity, and errors still occur. A novelist named Bernard Werber has asked this question in his book. "If there is no absolute truth in this world, and everything is a relative result that depends on the situation, is not it that you and me, and that we exist, are meaningless in this world? And if we think that this world is all relative, isn't it applying the absolute rule by the name of relativity?"

Let's separate it to two questions.

1. If there is no absolute truth in this world, and everything is a relative result that depends on the situation, is not it that you and me, and that we exist, are meaningless in this world?

2. If we think that this world is all relative, isn't it applying the absolute rule by the name of relativity?

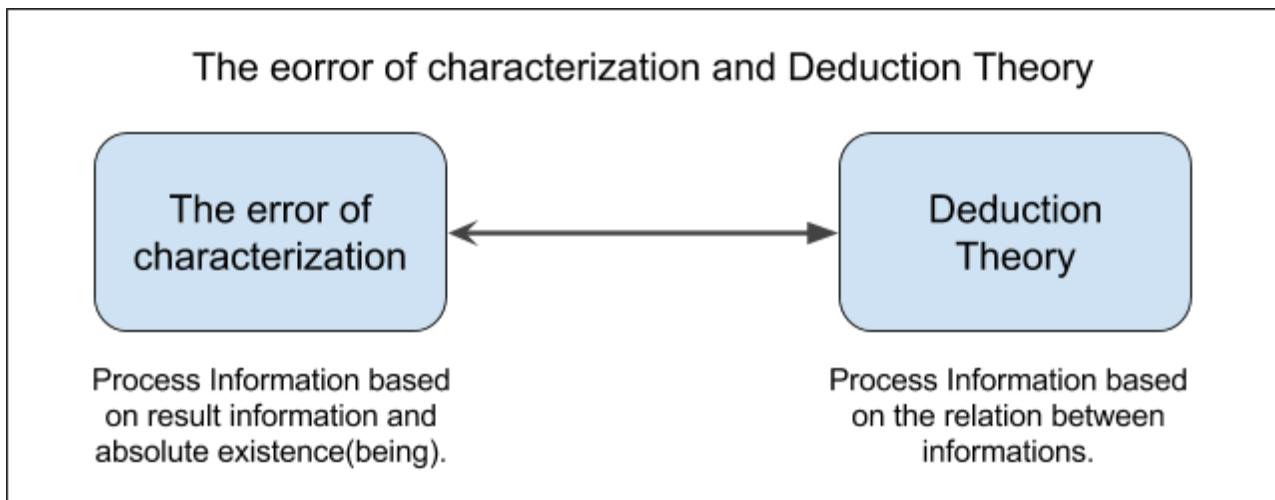
I'll answer these two questions.

**Werber's Question #1: If there is no absolute truth in this world, and everything is a relative result that depends on the situation, is not it that you and me, and that we exist, are meaningless in this world?**

This question assumes that there is no absolute point of view and no absolute truth, and at the rear end of the following sentence, he try to discuss relativity from absolute truth and absolute point of view. So it is an error.

In his discussing relativity, he try not to understand relativity as a relation but to understand it as a result. Here we can find something else besides the point of absolutism. It is an idea about result and being.

This is an interesting part. Assuming there is no absoluteness, even if you make a statement, the error continues because the person continues to think in terms of absoluteness. In Deduction Theory, this absoluteness view, the view to think on the basis of the result, and the view to think on the basis of existence are called the error of characterization.



There is no absoluteness in Deduction Theory. There is no absolute existence. The result information was created through the process.

Let's think about it. Have I been created through the process, or originally existed? We already know the answer. I was made through the process and I've keeping myself alive. My body created from my DNA information structure, and there is a conscious information structure that I created by thinking. My life have been processing in real time.

Was this universe created through the process, or did it exist in its original form? This universe was

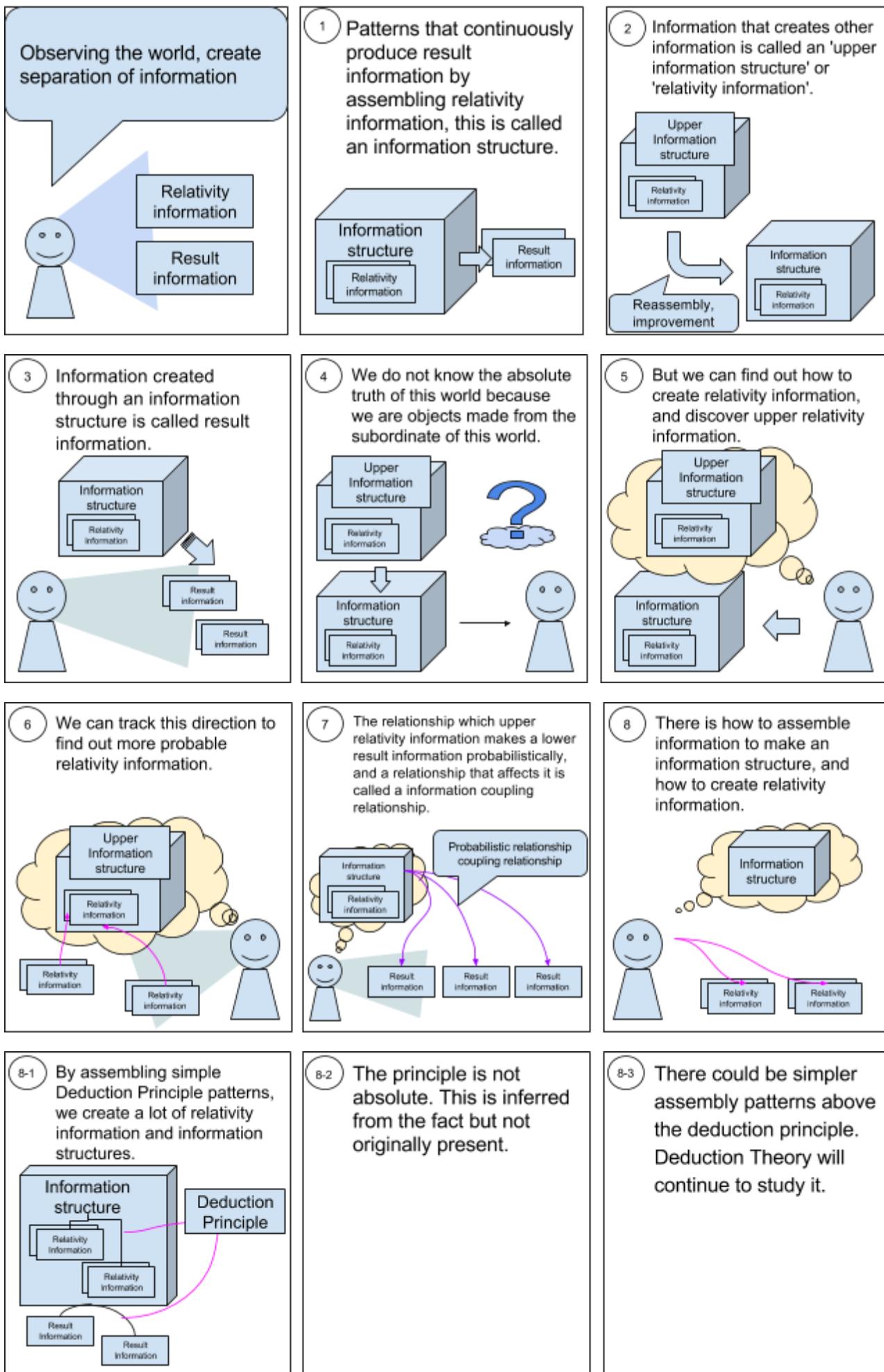
created through the process, and it is still continuing the process.

How Deduction Theory explain this relativity?

### The perspective of Deduction Theory

1. When information is assembled and able to continuously creates result information, it is called an information structure.
2. The information that creates other information is called an upper information structure or relativity information. The information structure is made by relativity information.
3. Information created through an information structure is called result information.
4. We do not know the absolute truth of this world. This is because we are objects created from the bottom of this world.
5. But we can find out how to create relativity information and discover upper relativity information.
6. So we can track this direction to find out higher, more probable relativity information.
7. The relationship which upper-level relativity information produces a lower-level result information probabilistically, and a relationship that affects it is called a information coupling relationship.
8. How to assemble information to make an information structure, and how to create relationship information. These are summarized in basic simple patterns as the Deduction Principle.
  - a. A lot of relativity information is made by assembling the Deduction Principle.
  - b. The Deduction Principle is not an absolute principle. This is inferred from the fact but it was not from original present.
  - c. There may be simpler assembly patterns above the deduction principle. Deduction Theory will continue to study it.
9. When you can construct the relativity information and create the result information, it is called an information structure.
10. If you change some result information in the information structure such as input and output, condition, etc. re-assembled and made available in other fields, it is called an opened information structure. An opened information structure is a module in a semi-assembled state. So this was called the deduction module in the past.
11. When you build a large information structure by connecting a lot of information structures, it is called a system. So information structure and system are synonymous essentially.

## The perspective of Deduction Theory



# Deduction Principle

## Separation

The Act to distinguish information such as positive and negative, existence and nothingness, 0 and 1.  
The act of making all the information is the addition of another action to the act of separating.

## Copy

The act of creating result information with the same appearance of identity

## Symmetry

The act of separation based on location difference with the act of copy

## Expand

The act of creating result information with different identity.  
Location can specified after expanding.

## Contrast

The act of separation based on location with the act of expand

## Balance

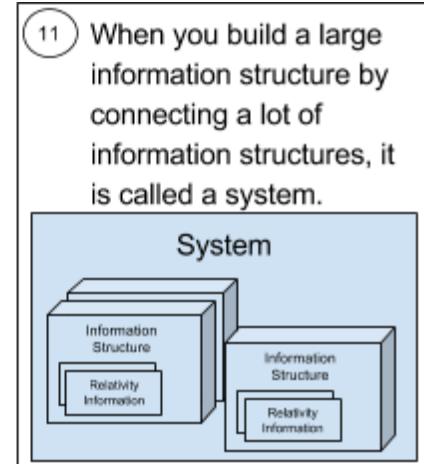
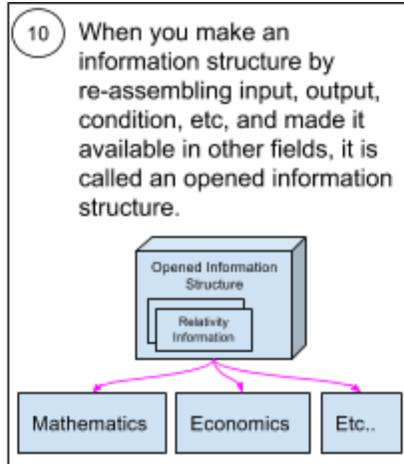
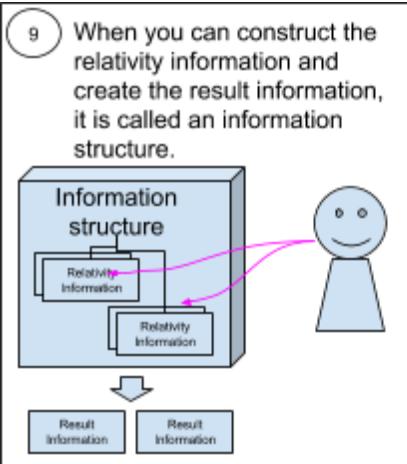
The act of separation based on copy and symmetry

## Causal Relation

A classification act that determines the information of processing begins and ends  
It creates time based on causality.

## Response

The act of separation based on expand and contrast



## Werber's Question #2: If we think that this world is all relative, isn't it applying the absolute rule by the name of relativity?

Question is about this. "Relative" means "there is no absolute rule", but isn't that "this world is all relative" means there is an absolute rule that governing the world with relativity?

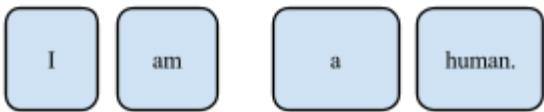
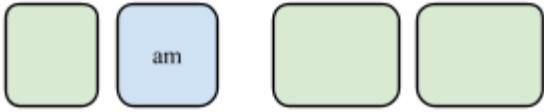
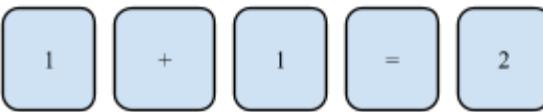
Deduction Theory says even in relativity, there is a stochastic relation that more or less influence with each others. It is not that absolute influence is exercised even in relativity.

The reason why this question arises because he have interpreted relative relations as absolute relationships again. No matter how logical the statement is, if the way of thinking does not change, the same type of error will continue to occur.

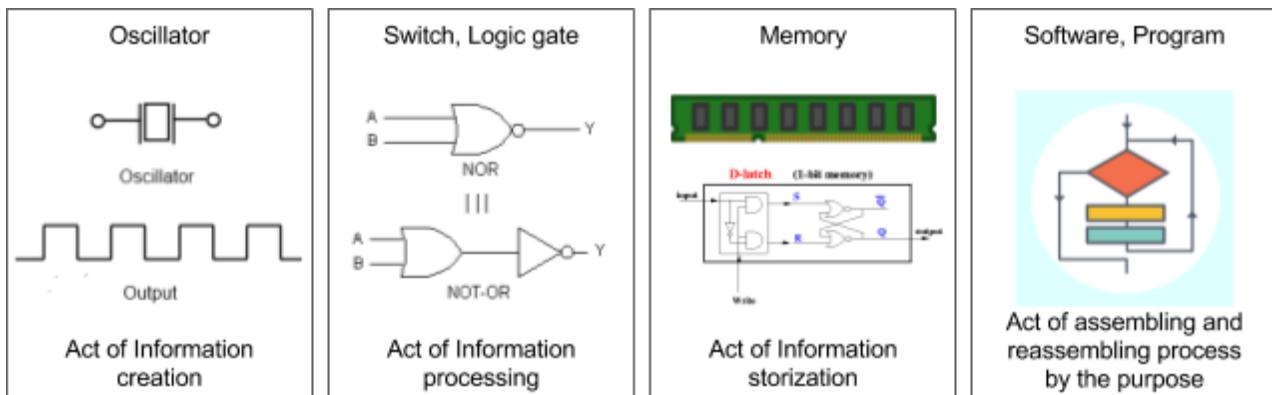
# Process, Pattern, Algorithm, Behavior, Action

The meanings of patterns and algorithms are the same. It means "the process of processing information". Patterns are relatively shorter and simpler, and algorithms are relatively longer and more complex.

In the Deduction Principle diagram above, "the act of doing something" comes out. "Process of processing information" is expressed by the act in language. Emptying nouns, names, and result information from sentences will leave the act and grammatical alignment. This sentence then becomes "an opened information structure" that can generate information by embedding other result information. In the language, verbs expressing action and behaviors, and grammar refining sentences serve as relativity information and information structures. Same as in mathematics, emptying a result information reveals the act of relativity information.

How to infer relativity information in sentences	How to infer relativity information in formula
<p>1) Example: I am a human.</p>  <p>2) In the example sentence, empty the noun corresponding to 'what'. Then 'how' will be left.</p>  <p>3) The remaining form acts as an opened information structure. You can create new information by inserting other result information.</p>	<p>1) Example: <math>1+1=2</math></p>  <p>2) In the example formula, empty the number corresponding to 'what'. Then 'how' will be left.</p>  <p>3) The remaining form acts as an opened information structure. You can create new information by inserting other result information.</p>

In the physical real world, the information processing process is represented by actions and behaviors. For example, in the case of a computer, information is replicated using a semiconductor that repeatedly replicates signal information called an oscillator, and signals are controlled to become specific pattern by using a switch and a logic gate, and the information is stored using a memory. Then, when you reassemble the pattern you saved by the purpose that you want, it becomes an algorithm and a program. This is how computers process information, and computer programs and software that the person decides how to assemble and use a certain pattern. For reference, relativity information and result information are made from Deduction Theory.



Let's point it out.

1. Patterns, algorithms, actions, and behaviors all mean "information processing".
  - a. A pattern is a relatively shorter and simpler process.
  - b. An algorithm is a relatively longer, more complex process.
2. Action and behavior are words used in the process of processing information.
  - a. Emptying nouns, names, and result information from sentences will leave the action and grammatical alignment.
  - b. This remained sentence then becomes "an opened information structure" that can generate information by embedding other result information.
  - c. Verbs and grammatical refinements that express behaviors in language act as relativity information and information structure.
  - d. In mathematics, emptying a result information reveals the act of relativity information.
3. In the material world, information processing processes are expressed as actions and behaviors.
  - a. For example, in the case of a computer, information is replicated using a semiconductor that repeatedly replicates signal information called an oscillator, and signals are controlled to become specific pattern by using a switch and a logic gate, and the information is stored using a memory.
  - b. Then, when you reassemble the pattern you saved by the purpose that you want, it becomes an algorithm and a program.
  - c. This is how computers process information, and computer programs and software that the person decides how to assemble and use a certain pattern.

## How to study in Deduction Theory

This is how to study in Deduction Theory.

Understand that all of this world is made up of information processing. Understand the relativity in Deduction Theory. Understand the errors of conventional absolute and relative, result information

error, characterization error.

Observing and analyzing this world from the viewpoint of Deduction Theory. Analyze the information processing process from we observed in my daily life, what we study, what we observed in my working.

Analyze long and complicated information processing processes in a shorter and simpler process, find out how they are assembled, and organize them. Make long, complex processes into assembly of short and simple processes. When analyzing the information processing process, we refer to simple patterns of Deduction Principles and apply them.

Then we can find out the difference by comparing the processing of information in the way that this pattern is duplication or expansion, what is symmetrical, or whether there is two or more overlaps. You can find out why this pattern is so created, how it works, and a more detailed process.

Analyze the information processing process and make it an opened information structure. We can make an opened information structure in language by removing result information from the sentence. The mathematical expressions are explained in a narrative and schematic form by removing result information and are made into an opened information structure. Analyze the computer program source to make it an opened information structure and apply it to other tasks. It can be applied to other fields and other tasks by identifying the information processing process and making it an opened information structure. It becomes one's creative ability and know-how.

We use the things we've learned so we create an information processing system for what we want to do. We create a system that has high probability of successful information processing and a massive automatic information processing.

Let's clear it.

1. Understand that this world is made up of information processing.
  - a. Understand the relativity of Deduction Theory.
  - b. Understand the errors of conventional absoluteity and relativity, result information error, characterization error.
2. Observing and analyzing the information processing process from the point of view of Deduction Theory.
  - a. Watch ourselves everyday what we do.
  - b. What we study
  - c. What we work for
3. Analyze long, complex information processing processes in a shorter and simpler process, find out how they are assembled and organize them.
  - a. Make long, complex processes into assembly of short and simple processes.

- b. When analyzing the information processing process, refer to simple patterns of Deduction Principles and apply them.
  - i. Then, we can find out the difference by comparing the information processing process in such a way that the pattern is a duplication of what, what is an extension, symmetry with what, or two or more overlaps.
  - ii. So we can find out why this pattern is so created, how it works, and a more detailed process.
- 4. Analyze the information processing process and make it an opened information structure.
  - a. We can make an opened information structure in language by removing result information from the sentence.
  - b. The mathematical expressions are explained in a narrative and schematic form by removing result information and are made into an opened information structure.
  - c. Analyze the computer program source to make it an opened information structure and apply it to other tasks.
  - d. It can be applied to other fields and other tasks by identifying the information processing process and making it an opened information structure.
  - e. It becomes one's creative ability and know-how.
- 5. We use the things we've learned so we create an information processing system for what we want to do.
  - a. Create a system that has high probability of successful information processing and a massive automatic information processing.

## Study about a view of Axiom

A view of Axiom is a way of thinking that claims 'an absolute relationship is established within the game called the axiom space.'

After mathematicians who encountered non-Euclidean geometry realized that mathematics does not match 1:1 with the real world, then they claim to 'make a game distinct from the real world and think that new game is absolute'.

### Criteria of axiom

1. All theorems are to be obtained from the axiom
2. If there is one proposition arbitrarily excluded from the axiomatic system, there should be some theorem that is impossible to prove.
3. It would be impossible to prove several theorems contradictory to each other.

Converts the view of Axiom to Deduction Theory words. The axiom system is called the upper rule.

1. All sub rules shall be assembled by assembling the upper rule without exception handling.
2. The upper rule shall not overlap other upper rules or refer to each other.
3. The lower rules made using the upper rule should not conflict with each other.

Gödel's incompleteness theorems use the axiomatic rule to determine the criterion of the axiom system. Contradictory relations between subordinate rules and proves that they can not make 'absolute games' that axiom claims.

What do we see when we look at axiom from the point of view of Deduction Theory?

The only way to find out if a game is work in the real world is by practicing the game directly in the real world.

If creates a hypothetical rule and expect the rule to be absolutely right, will be failed.

The latest scientific mindset is to constantly explore exceptions and higher rules that include exceptions if previously defined games and other results are found.

## How does Deduction Theory define the game?

Deduction Theory does not attempt to define an absolute game.

In deduction theory, this world is made up of information. The way this world continues is called information processing.

'An individual' is an information structure created by assembling information processing methods.

As the information structure 'an individual' explores better information processing because it tries to persist itself.

The information processing that it does inside is different and operated from the information processing that the outside world does.

It replicates the information processing that this world does to its inside.

It build the top rule by assembling the information processing method. This rule is relatively more certain within it, but it is separate from the world's information processing.

The only way it can see if it can sustain itself using its inner information processing method is to practice the way it does it.

We using computer and programming to create massive automatic information processing system.

In a view of information processing, computer program can separate between process-oriented and object-oriented.

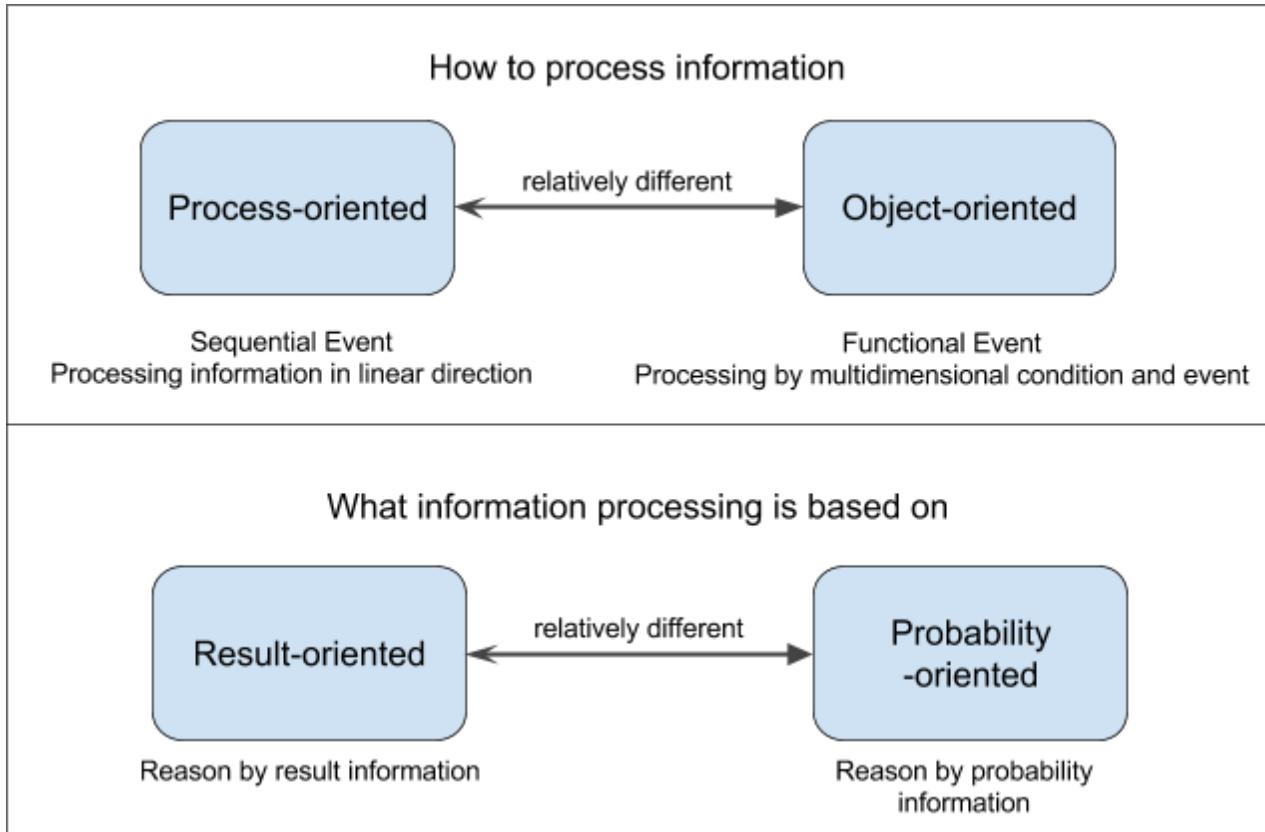
In a view of reasoning, computer program can separate between result-oriented and probability-oriented.

The process-oriented treats the behaviors defined in a closed sequence using data.

The object-oriented handles data using relatively opened sequential objects.

The result-oriented uses the result data as a basis for information processing and makes an information processing decision.

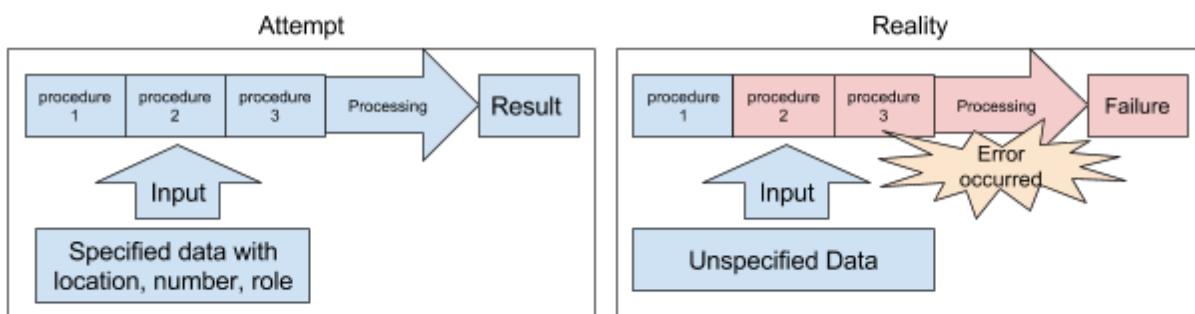
The probability-oriented generates a probability pattern from the result information data and makes an information processing decision by the probability pattern.



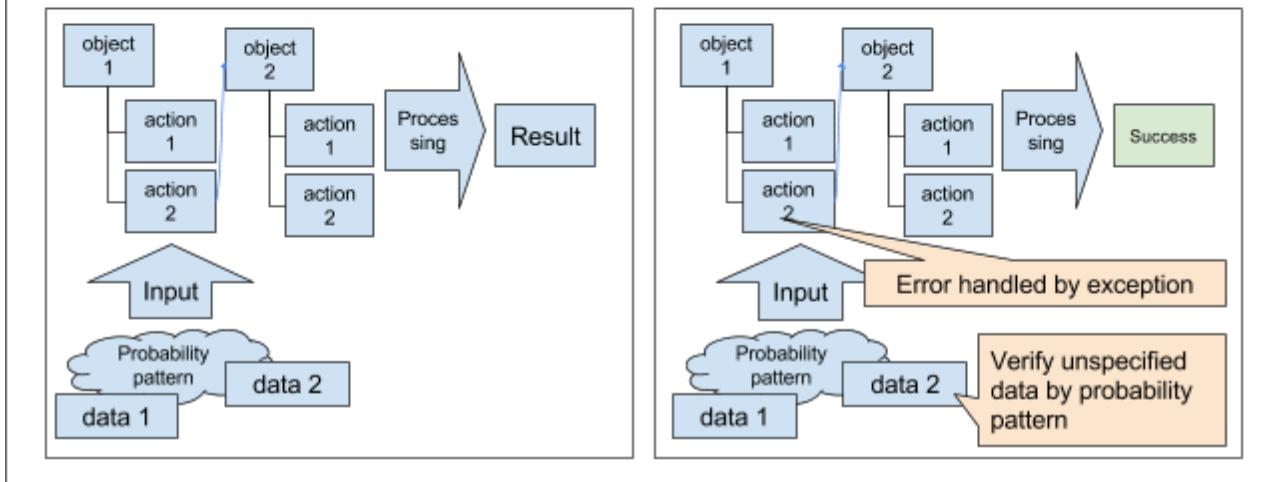
Process-oriented and result-oriented methods cause errors when empty space is created during the process and the basis of the decision is not correct result data.

The object-oriented method and the probability-oriented method can process the empty space by flexibly adjusting the process and perform the information processing while maintaining the information processing accuracy by assuming a probability estimate even if the data based on the judgment is not correct. Therefore, massive automated information processing can be performed effectively.

## Case of process-oriented and result-oriented



## Case of object-oriented and probability-oriented

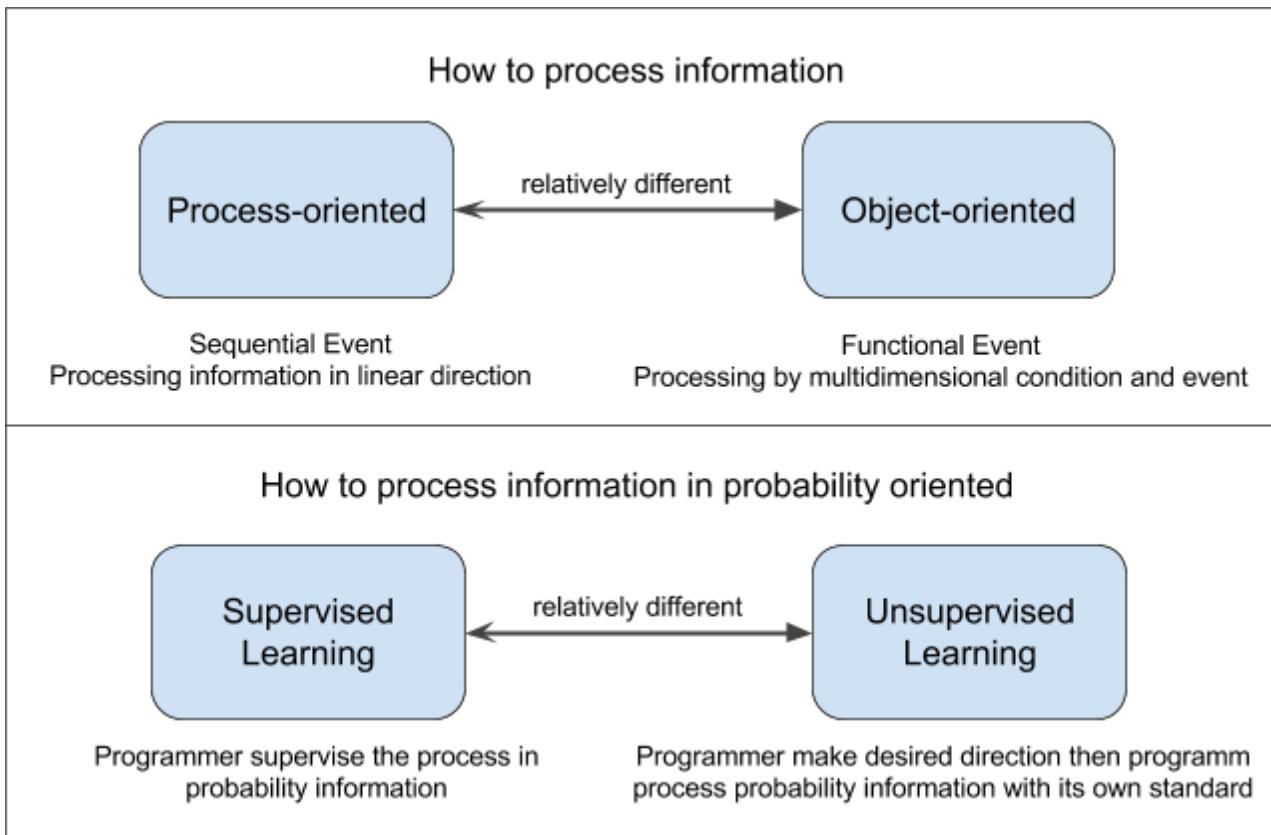


Big Data refers to an massive automated information processing system using a large amount of data. Machine learning and deep learning use object oriented, probability oriented methods while using large amounts of data. Since these two views can be applied in a multidimensional manner, there are cases of big data plus machine learning, and big data but not machine learning. For example, while processing a large amount of information automatically, you can still work with the conventional data processing like Excel, VBA, and SQL.

However, most big data systems use both fixed and unstructured data and perform probability-oriented information processing by machine learning. This is because if you try to process massive data as a result oriented program, it takes too much time and computing resources to process and specify the data as result information, and if you get an error during the process, you have to check again from the beginning.

Unsupervised learning is the process by which a program deduces information coupling and asymmetry by itself, even if the programmer decides the underlying algorithm less. Supervised learning, on the other hand, is more of a programmer offering basic algorithms, and the program enhances it. Deep Learning is named after "Deep Neural Network, DNN". Deep learning takes some supervised learning in the process of solving a given task, but it increases the proportion of unsupervised learning relatively and makes the program perform probability-oriented information

processing by itself.



Unsupervised learning allows a programmer to present only what he wants to output, and to process the information up to that point, so that the program can infer the information coupling and asymmetry in the data set itself. Information coupling and asymmetry are similarities and differences between patterns of information, which is relativity information.

Compared to this, supervised learning, the programmer gives more steps in the process of going to the output. This is a relative difference. There are cases where the learning is relatively unsupervised, and there are cases where the learning is less unsupervised.

The term of artificial intelligence is used to refer to information structures, including machine learning and deep learning, that act as physical results. But in Deduction Theory we do not use a word artificial intelligence, we use the term Relativity Information Intelligence.

Because in word artificial intelligence has two meanings.

1. Human-generated information processing system(intelligence)
2. Human-like information processing system(intelligence)

But let's think about it once. First, are there any non-artificial tools created by humans? When human beings touch their hands, they are artificial. So this classification is not appropriate. Second, if it is human-like and artificial intelligence, what if the information processing system later

evolves to a much higher level than human ability or evolves to another dimension? So this is not suitable name for the system either.

Therefore, we stud Deduction Theory, call it Relativity Information Intelligence which means that this information processing system is suitable for processing relativity information.

The reason we study machine learning and deep learning research is because we are trying to do probablitic information processing, relativity information processing, so massive automated information processing effectively in our work.

Prior to Deep Learning, there was a way to think about probability information as an object that can be controlled and predicted absolutely.

For example, it is assuming that there are movie genres and movie viewers, hypothesize that 'There will be a equation that male viewers watch more of action genres then drama', and calculate the hypothesis using a machine learning model or a similar mathematical method, putting actual data on the computer and checking whether the expected result is generated within the acceptable error range.

The reason we say that existing machine learning researchers would have this way of thinking is that because there is a definition "Machine learning is an act of finding an approximate function 'F' that solves the same type of problem by using the presented data and the pairs of decision result.".

The way of thinking above is used in science and industry field until now. We do not intend to disparage it.

But we have a different mindset as Deduction Theory.

We think the term 'find a generalized function that solves a problem' is going towards finding more absolute criteria. However, there is no absoluteness in the world view of Deduction Theory. From the point of view of Deduction Theory, the error increases further in that direction.

We will research and develop programs that are based on 'Finding a way to solve a problem relatively constantly and improving through it'.

## Relativity Information Intelligence

This is what Relativity Information Intelligence do.

1. Relatively more object-oriented information processing
2. Relatively more probability-oriented information processing
3. Relatively more way of unsupervised learning
4. Relatively more relativity information processing
5. So massive automated information processing effectively in the task.

Relativity Information Intelligence is not an absolutely defined result.

It refers to a system that performs relativity information processing from relatively to more object oriented, more probability oriented, and more unsupervised learning.

By raising this level, you can create a bot system that processes relativity information in a uniform view of any subject, which is called Relativity Information Intelligence in Deduction Theory.

## 20161127 study Julia Lang

We are trying to port Perceptron code that we studied in C++ to Julia.

The Julia language is a programming language that allows mathematical grammar to take precedence over conflicting or confusing parts of mathematical notation in the grammar of existing programming languages in order to facilitate programming from the viewpoint of mathematicians and data engineers.

There are some features in Julia such as the starting number of the array is 1, the square can be computed by  $x^2$ , there is no object class but in other object-oriented programming languages have them, the function is used centered, the dynamic variable such as Python or Javascript is automatically determined by the interpreter.

In Julia language, instead of having an object class, there is a Type, which is similar to the C language Construct, which is a collection of variables and puts them into a single variable Type. It differs from object class in that there is no inheritance concept, and it is necessary to put it into a variable in order to put another function in place of the constructor function. Julia uses the Type as an object.

So if you force it, you can create a structure similar to the object-oriented language of the previous method.

Conventional object-oriented languages have objects with built-in properties and functions. The type used by Julia has built-in properties. Since the function is written outside type, it can be used by sharing other types and functions.

What is more opened information processing?

It is more opened to be able to dynamically bundle several methods, use them as objects, and reassemble them. But there is no programming language created yet from this perspective perfectly. The Julia language is a bit closer to this concept in that function can be shared with other types.

How do we program in this more opened perspective?

In Deduction Theory, there is a sentence “Create what, which and why by using how”

Functions are also a block of information processing.

We can treat the object as a bag of connection and puts the information processing method into connection.

How can we handle probability information processing in a programming language?

Do not declare variables beforehand. It is possible to calculate information by distinguishing between similar information and other information.

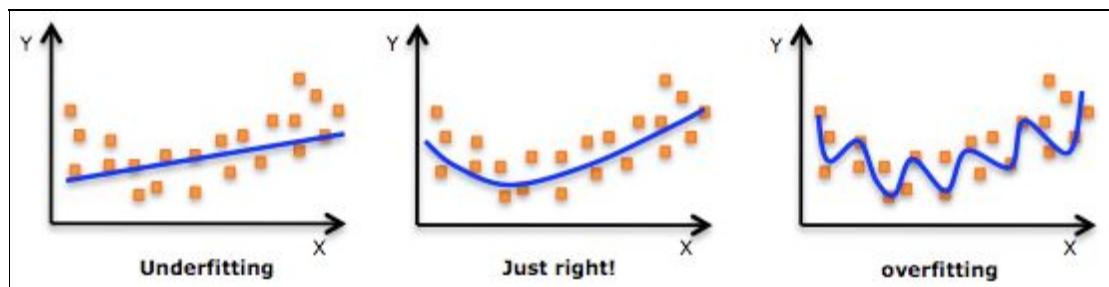
If you need a name, you can name it afterward, or let the computer give a proper name itself.

Even if you use a similar name without precisely typing the variable name, it is possible to process the information by inputting a stochastic numerical value without inputting a fixed numerical value.

When the external environment changes, the internal information processing method is reassembled. When the developer changes the information processing structure, the details are automatically adjusted.

## 20161128 Study Convolutional Neural Network

[Convolutional Neural Network - Wikipedia](#)



Overfitting is a phenomenon in which the Neural Network is excessively obsessed with certain events when the number of layers of the Neural Network is greatly increased and the amount of sample data is input.

CNN has its own optimization method to avoid overfitting as much as possible.

CNN solves overfitting issues by reducing model complexity.

The Convolutional Neural Network, Convnet or CNN firstly, imitates the human optic nerve structure and classifies the input data as a group of small zones(Convolutional layer). Secondly, it combines the data to easily find the data synchronization such as information coupling and

asymmetry(Pooling layer , SubSampling).

[S231n: Convolutional Neural Networks for Visual Recognition](#) - Stanford CS class github  
[UNDERSTANDING CONVOLUTIONAL NEURAL NETWORKS FOR NLP](#) wildml blog [gihub](#)  
[UFLDL Tutorial](#) - Stanford Deep Learning Tutorial

In the pre-training phase, CNN sets unsupervised information for the convolution layer and subsampling, but afterward supervise the actual information by using the labeled data.

## 20161203 Study Perceptron and Back-Propagation

The Gradient Descent Method is used as the internal information processing of Back-Propagation method which distributes the weight so that the deep learning neurons output the target value.

The process of receiving and outputting a neuron is as follows.

$\text{output} = \text{activation function}(\sigma(\text{input})) = \text{weight} * \text{input} + \text{bias}$

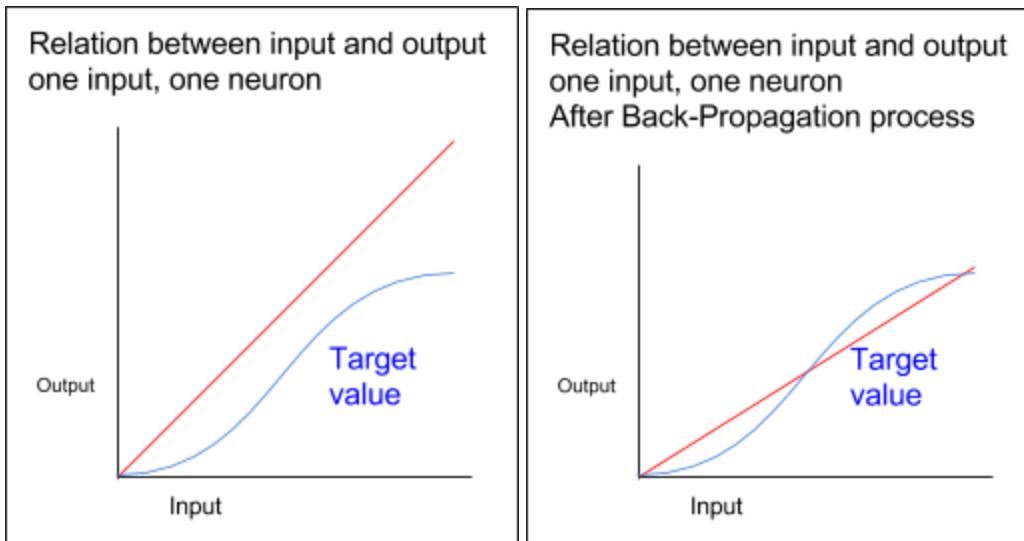
The weight is multiplied by the input and the bias is added. This adjust them with the activation function.

The formula to multiply the weight by the input value is the same as the dot product in Vector mathematics.

In Vector mathematics, the constant obtained by dot producting between two vectors is called a scaler, which is a feature that represents the distance relation between two vectors as a constant. In neuron programming, it look at the weighted array and the input value array as a pair of vectors.

Below is the process of Back-Propagation.

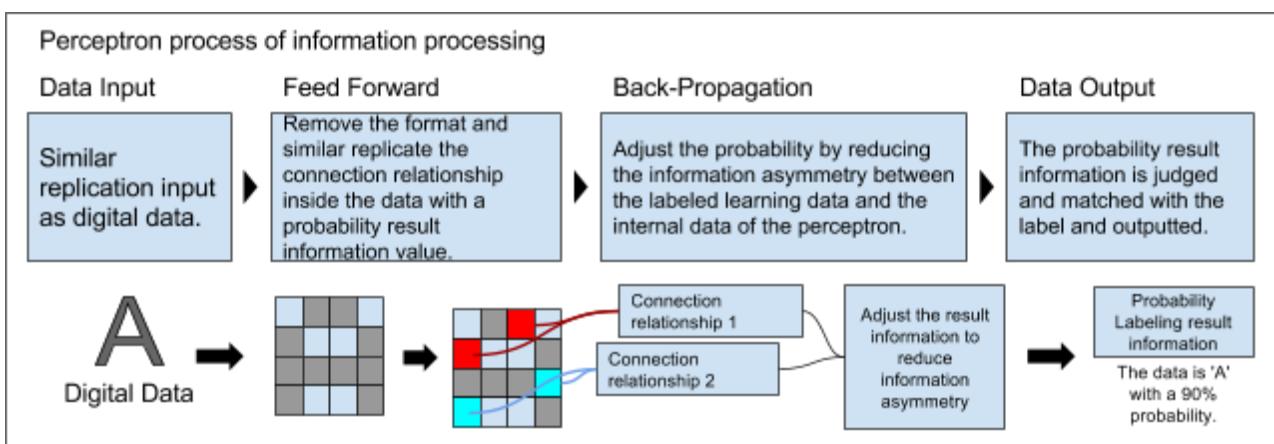
$\text{weight} = \text{old weight} - \text{learning rate} * (\text{currunt output} - \text{disired ouput}) * \text{gradient desent function on activation function}(\text{current output}) * \text{old input}$

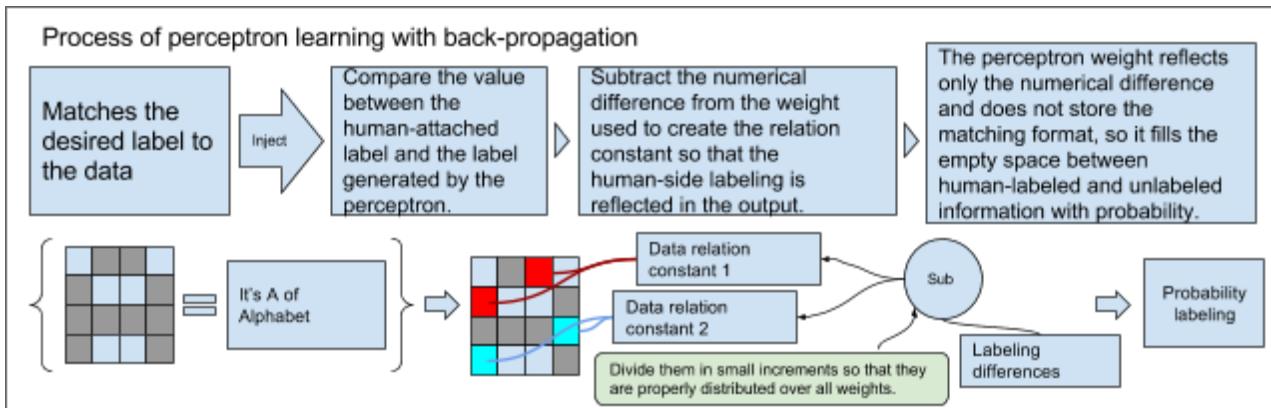
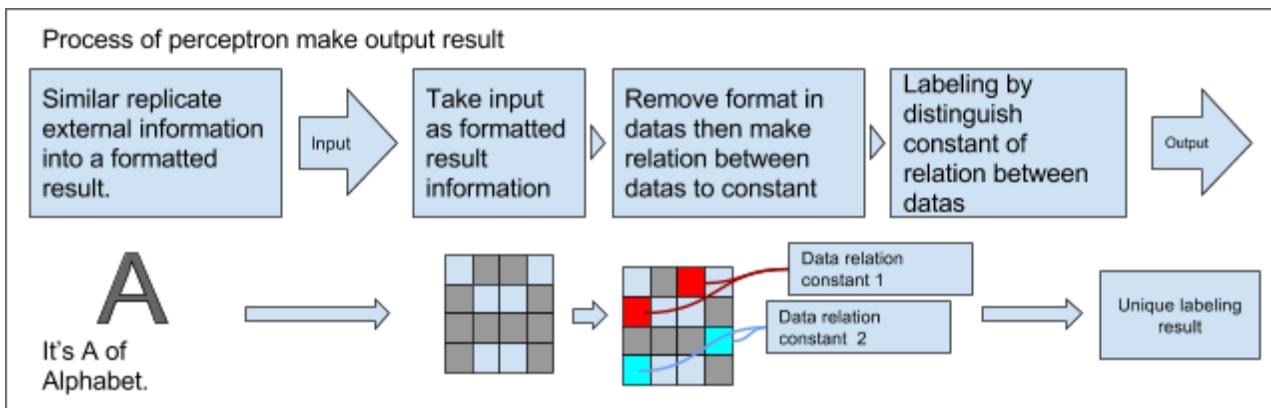


The reason for doing the gradient descent method is that if the neuron uses a certain activation function internally, the activation function tries to use it to track how much weight is to be input to get the target output value.

How can I explain this process more simply?

1. In process of making output, when we look at the activation process of the neural net, it removes the format from the connected data and makes the relation between the data into the probabilistic result information.
  - a. Because the format is removed, the input value can not be restored with the output value and the weight value. Inverse operation of mathematics is not established.
2. Back-propagation is a learning process performed by a neural network. It is the act of assigning probability scores to the degree that humans see the relationship between the data (or by putting the rules of scoring by humans) into consideration.



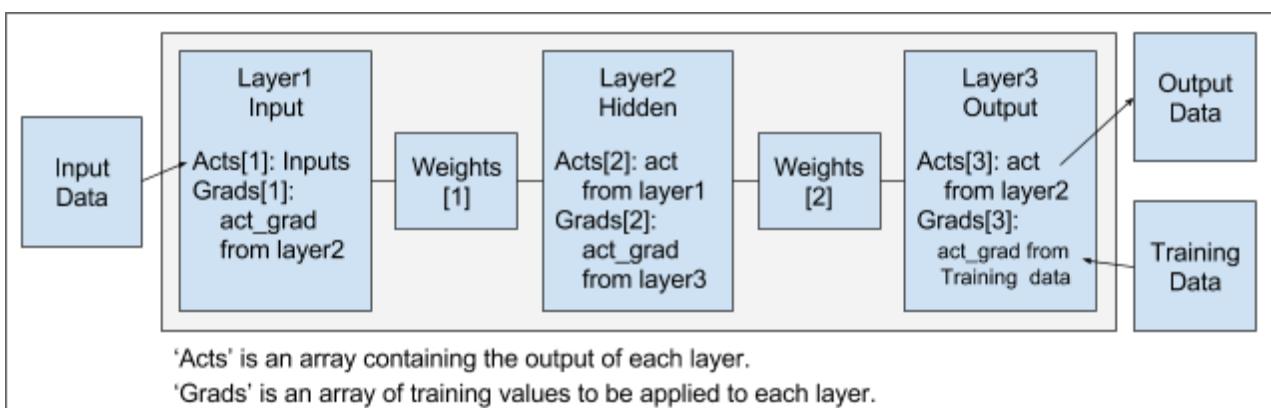


What can we see from this?

The output is in response to the input. The output information is more opened by indicating the relationship without the original format, but the process of receiving and outputting the input is closed and fixed.

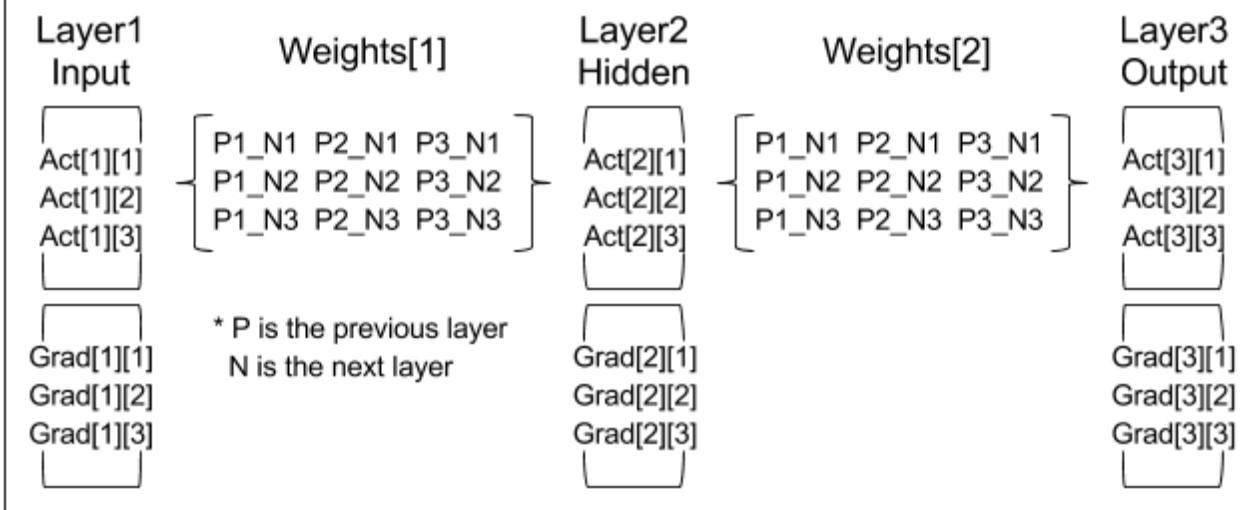
FNN, CNN, and DQN are similar to the reptilian brain in comparison to human information processing.

## 20161217 The processing of multi layer perceptron example code

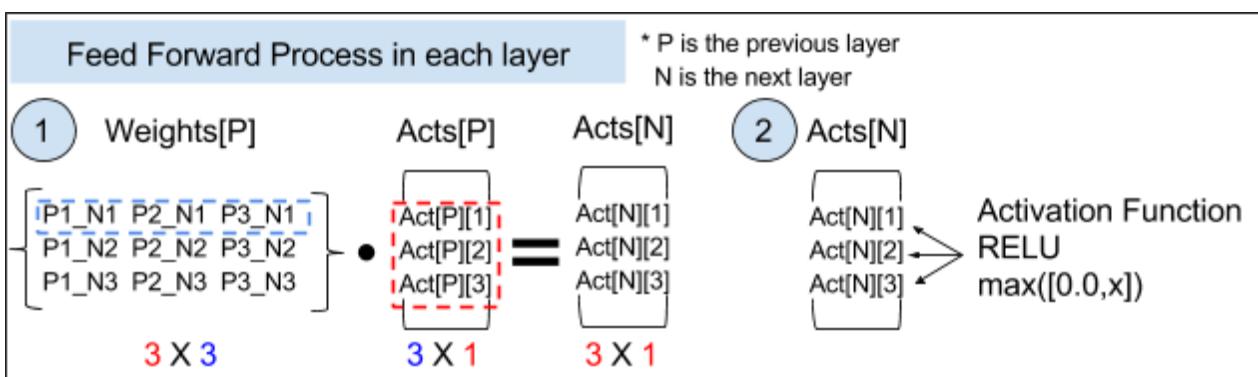


Acts, Grads are vector variables and Weights are matrix variables.

## Structure of Neural Network with 3 input, 3 output and 1 hidden layer

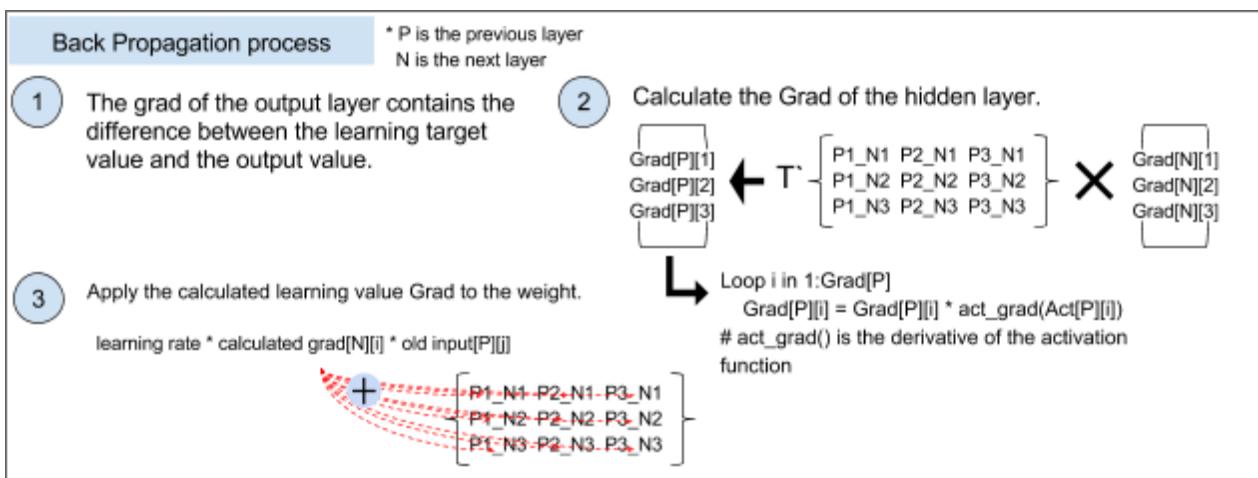


'Acts' is an internal component of the Grads vector and the Weights matrix variable.



The formula for the above procedure is as follows. I omitted the bias in the diagram. \* From the neuron layer view, Acts [P] is input and Acts [N] is output.

$$\text{output} = \text{activation function}(\sigma(\text{input})) \Leftarrow \sigma(\text{input}) = \text{weight} * \text{input} + \text{bias}$$



The formula for the above process:

$$\text{weight} = \text{old weight} - \text{learning rate} * (\text{currunt output} - \text{disired ouput}) * \text{gradient descent function on activation function}(\text{current output}) * \text{old input}$$

In the example, we modified the (currunt output - disrupted ouput) to (disused ouput - currunt output) and applied it to the weight instead of subtracting the learned value from the old weight. The elements of the Grad array are (currunt output - disused ouput) \* gradient descent function on activation function (current output).

The result of the implementation is that the learning is done normally, and the output is normal. In the future, we will look for other examples, examples created in other languages, compare them.

## Study reinforcement Learning

### 1. We read DeepMind DQN paper and made a memo.

[Human-Level Control through Deep Reinforcement Learning](#) Deepmind reinforcement Learning paper

In the text, the process of DQN is summarized as follows. Below is an excerpt from the article.

.....

we use a deep convolutional neural network to approximate the optimal action-value function

$$Q^*(s, a) = \max_{\pi} \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots | s_t = s, a_t = a, \pi],$$

which is the maximum sum of rewards  $r_t$  discounted by  $\gamma$  at each timestep  $t$ , achievable by a behaviour policy  $\pi = P(a|s)$ , after making an observation ( $s$ ) and taking an action ( $a$ ).

.....

The symbols and terms in the above equation are unfamiliar to me so, I search to wikipedia. [List of mathematical symbols](#)

Max prints the maximum value in the given set. Below max is a space which the condition is entered. Pi is used in this paper to mean a policy function.

E is the expected value and is used in the probability theory. Meaning is the average value expected in the random process.

| Is used to mean "such that." The information after | is the information that the preceding information refers to.

### 2. Install and run the source code published in the Deep Mind Atari article.

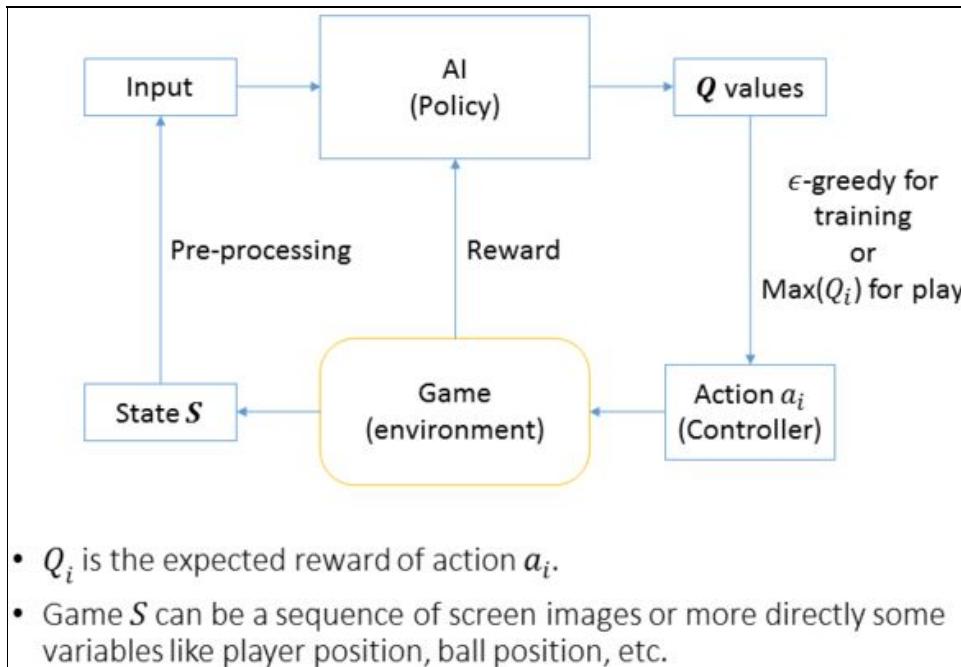
To see how DQN works, I installed and run the published source code.

[Human-level control through deep reinforcement learning](#) - Source code

The reinforcement learning method is DQN and the other is Policy gradient (PG).

### 3. listen another class content

I've listened another class content to understand better about it.



In the above diagram, AI refers to the neural net.

Inputs may receive predefined game state values (location, score, character state, etc.) or CNN output values processed with screen data.

$Q$  values is an array of estimated reward values (expected reward values) for each action that can be taken.

E-greedy is a brute force rule that randomly selects an action that is likely to receive a large reward even if the action gives less reward currently.

Max ( $Q_i$ ) is the sum of the present and the future, estimating the total value of the rewards, and executing the corresponding action.

$$Q(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \left( \underbrace{r_{t+1}}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \right)$$

The output function  $Q$  is generated by executing old values + learning rate \* (reward + discount factor \* estimate of optimal future value).

1.  $Q(s_t, a_t)$  outputs a  $Q$  value array obtained by input numerical values of 'agent state' and 'action to execute'. (Array of expected reward values)
2. The old value is the reward that can be obtained when the state is changed by executing

- the previous action.
3. Estimate of optimal future value: Lists the actions that can be taken on the next turn in the current state and finds the most likely action to get reward from.
  4. Discount factor: Reduces compensation every time an action is taken to avoid repeating meaningless behavior in the absence of compensation.
  5. Perform probability information processing by back-propagation with applying learning rate from target learning value.

Let's try again the question I had before I learned reinforcement learning.

How does the neural net adjust the weights on reinforcement learning?

It adjust the weights in response to the rules for receiving rewards in the external environment of the game.

How did people adjust the weight before reinforcement learning? The learning value corresponded to the input value was input by the developer or the matching learning value was pre-inputted.

#### **4. Re-analyze the formula in the Atari paper.**

:::::

we use a deep convolutional neural network to approximate the optimal action-value function

$$Q^*(s, a) = \max_{\pi} \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots | s_t = s, a_t = a, \pi],$$

which is the maximum sum of rewards  $r_t$  discounted by  $\gamma$  at each timestep  $t$ , achievable by a behaviour policy  $\pi = P(a|s)$ , after making an observation ( $s$ ) and taking an action ( $a$ ).

:::::

Is the information process closed and fixed? Or is the information process opened and flexible?

The way of making a rule ( $P_i$ ) is opened. In the actual process, a random function called "E-greedy" finds a "brute force" to get more rewards.

Will the rules of the system be updated in real time? Or is it fixed?

It updates instantly while experiencing in real time.

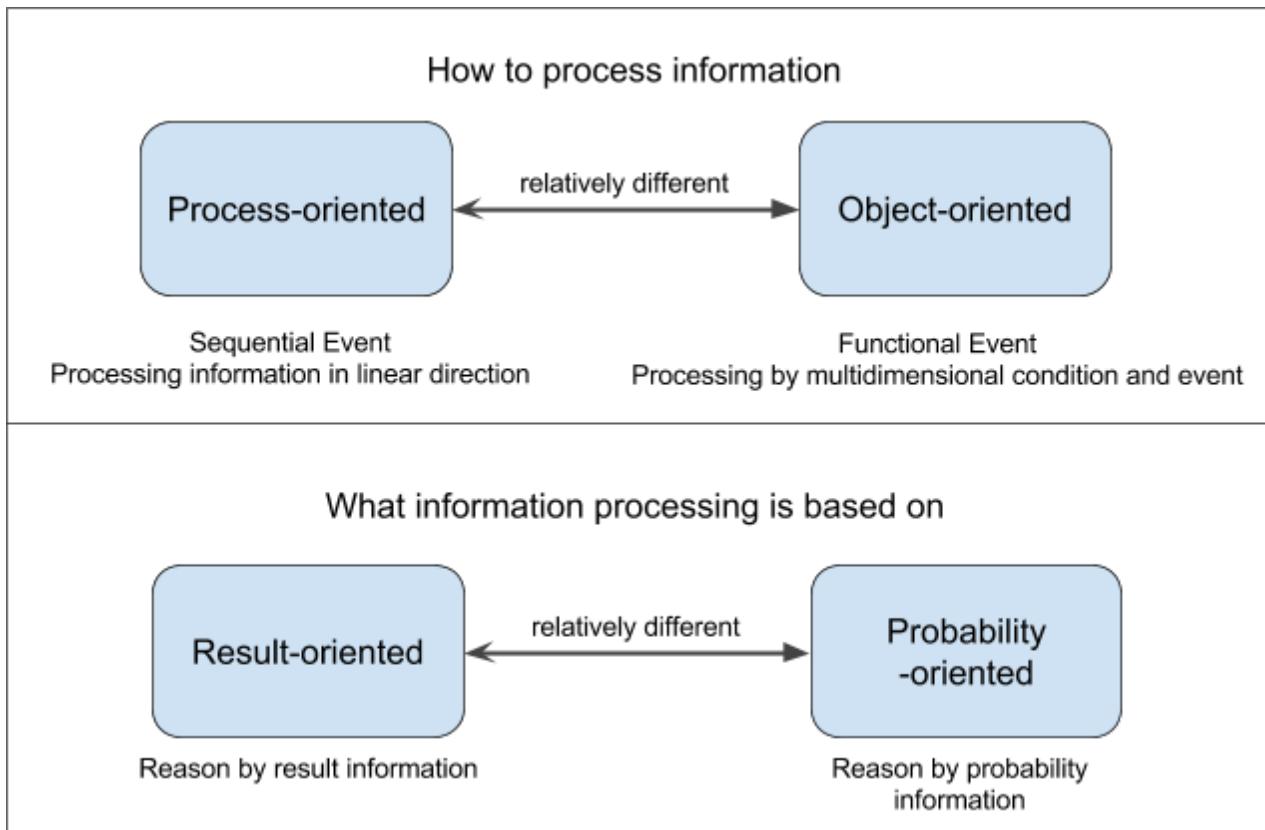
DQN suggests relatively absolute game rules and induces neural networks to create information structures that effective and follows game rule while iterating over attempts.

Instant output is performed in response to the input without comparing the information processing process inside of the individual. This is similar to the information processing of the reptilian brain.

# Explaining Deep Learning in Deduction Theory

In Deduction Theory, deep learning and machine learning are considered as probability oriented information processing types.

We see deep learning and machine learning as an attempt to make all types of information processing relatively close to probability information processing.



An important concept of probability information processing is to produce information that is relatively close to the relativity information than result information.

To say it an easy way, make an output information as "how" not "what".

So why do we have to process probability information? Why do we need it?

Even if you use the result information, you can process with it. So far, the computer has done it. The result information is data. However, in this case, if the input does not match the result, the information processing error occurs. Anyone who has used the console command screen should have seen a lot of the words "Invalid command or file name" when typing commands.

```
chase@jaguar:~$ mkdir 'my documents'  
chase@jaguar:~$ ls './my documents'  
chase@jaguar:~$ ls ./my documents  
ls: cannot access './my': No such file or directory  
ls: cannot access 'documents': No such file or directory  
chase@jaguar:~$ █
```

The more information the computer processes based on the result information, that is, the more information processing in absoluteness, the less the ability of the computer to solve problems in the real world. Problem solving ability becomes limited.

This is because, if the reality is dataized into the result information from an absolute point of view that is not existent, the data is out of the reality. The more information processing is done from the point of view of absoluteity, the more the relation with reality becomes relatively far away. In other words, the ability to solve problems in reality becomes falling short.



This verse comes out in a book called 'The Book of The Way' written by Lao-Tzu, "The Way that can be told of is not an unvarying way, The names that can be named are not unvarying names.". This book and Lao-Tzu's concept called as Taoism.

道可道 非常道, The Way that can be told of is not an unvarying way.

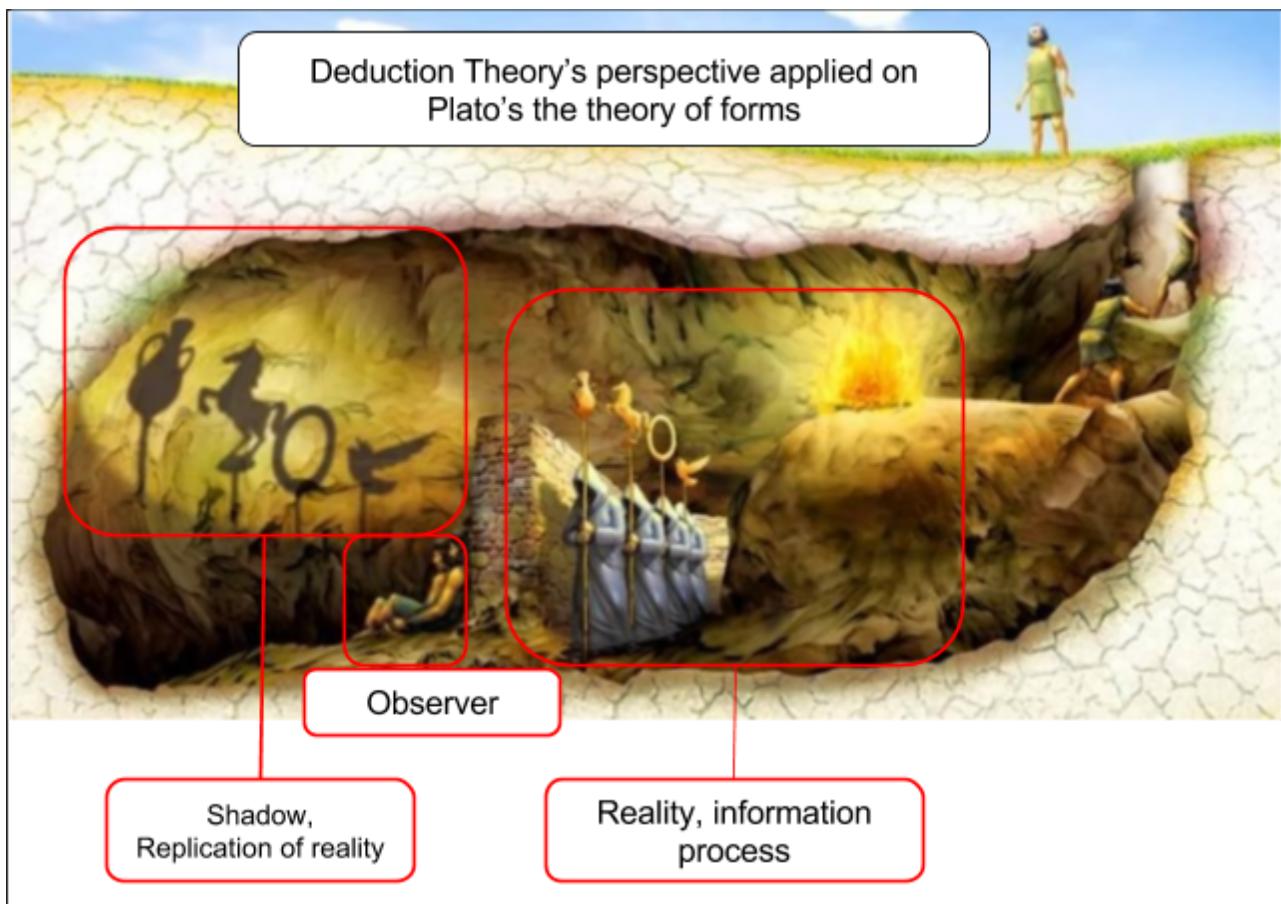
名可名 非常名, The names that can be named are not unvarying names.

[Tao\\_Te\\_Ching](#) The book of the way - Wikipedia

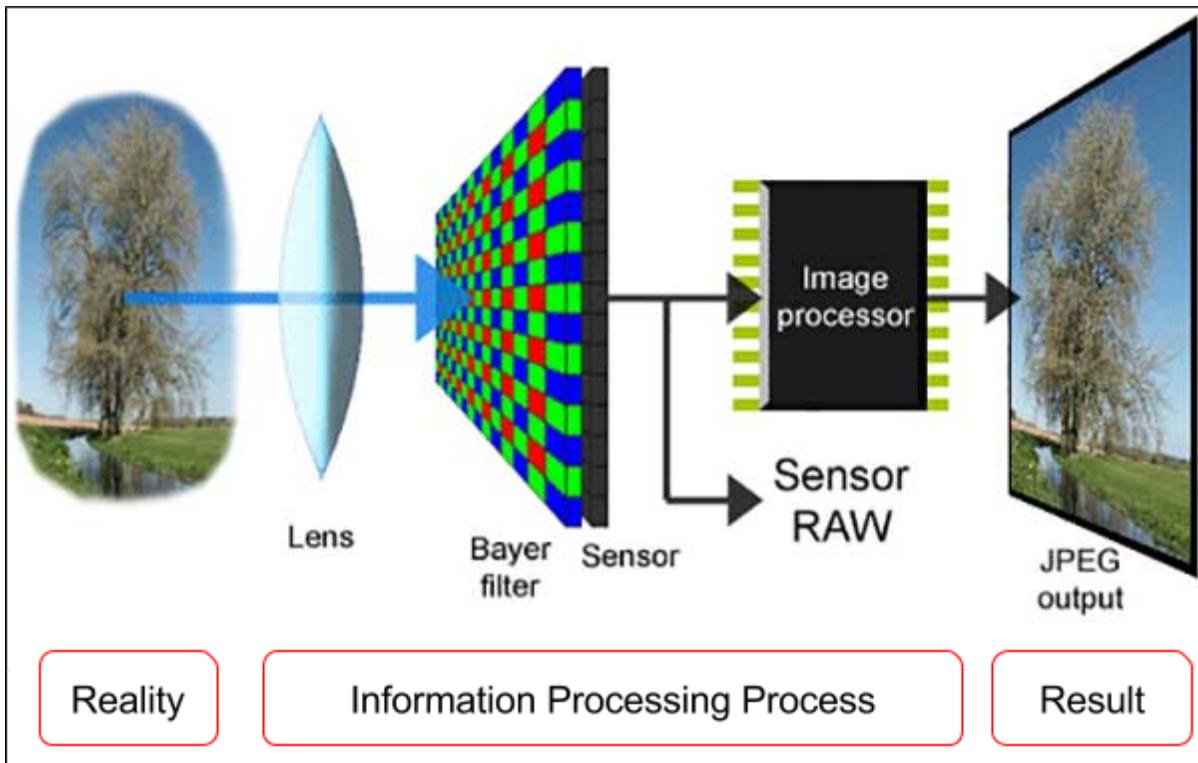
For reference, Gottfried W. Leibniz read a principle of Yin-Yang in 'I Ching(Classic of Changes)' and used it as a reference for creating a concept of binary number which is used on computer nowadays. Yin and Yang are concepts derived from Taoism.

The relationship between "the reality corresponding to the essence" and "imitation of the reality" which is our perception of processing information is also a famous theme from ancient times in the West. It is known as Plato's the theory of forms. Plato regarded Idea as the essence of all things, and as we perceive and reason, we see its shadow imitation. I will explain relativity information processing by modifying Plato's opinion slightly.

When Plato makes the theory of forms, he said there is perfect forms called 'Idea'. But in Deduction Theory, we think there is no perfect forms. We will use only pictures that illustrate the theory of forms as a reference.



Let's say you have a computer program that creates the data by using various input devices. For example, there is a digital camera. The CCD sensor(or CMOS) of a digital camera converts light information obtained through a lens into digital data. This data is the result information in Deduction Theory.

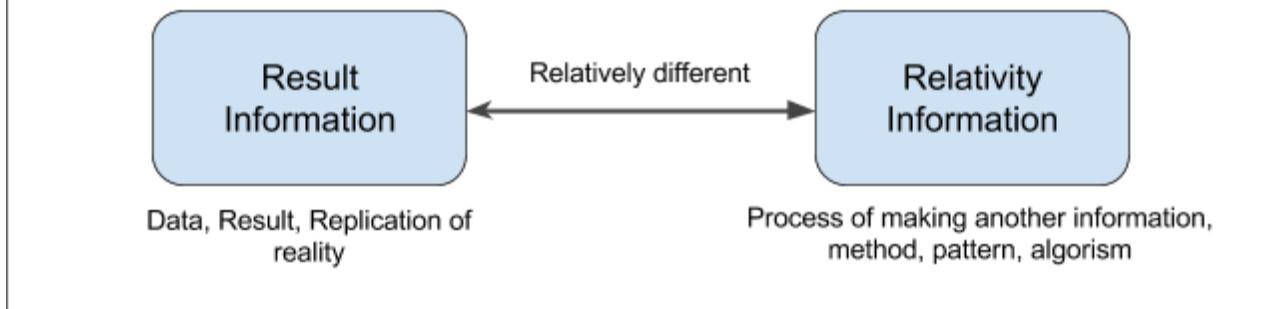


The data created by the digital camera is the result information that a person organizes and processes the information processing process. But is this photo data taken with a digital camera real? No, It is imitation of reality. But if you consider this data as reality and reasoning from it, likely error will be accumulated.

How then can we escape the accumulation of this error? Let's think about this once.

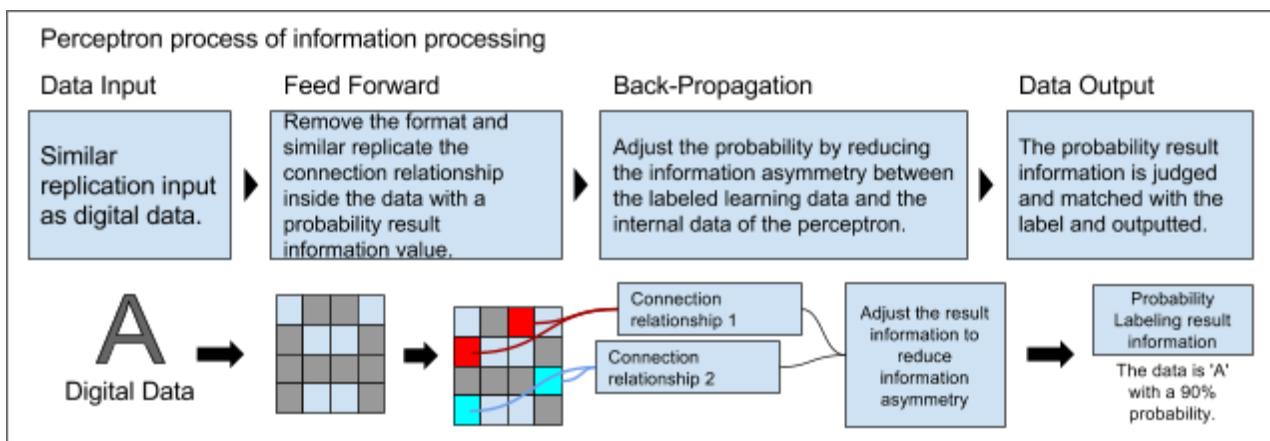
1. Admit that the result information we produce is not all of reality. We see this world in terms of relativity in Deduction Theory, not absolute.
  - a. All data, all statements are regarded as limited result information and not believed absolutely.
2. Realize that there is a way of producing the result information, "information processing".
  - a. This is called an information structure using relativity information in deduction theory.
  - b. Relativity information refers to processes, principles, patterns, algorithms, and so on.
  - c. An information structure is a structure that produces information consisting of relativity information.
3. Analyze the result information data to infer the "information processing process".
4. Creates information processing process. This is an intelligence.
  - a. If the computer can create the information processing process by itself, it becomes computer intelligence.
    - i. This is currently called artificial intelligence in the industry.

## Difference between Result information and Relativity information



In fact, the science and discovery that we speak is the act of collecting data in reality and deducing the underlying principles of information processing behind it. Now we try to make this special information processing activity possible on a computer.

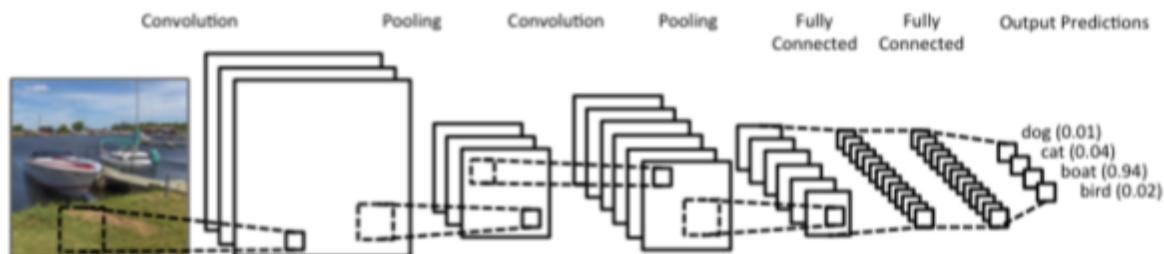
Conventional programs prior to machine learning and deep learning received a result information as input and generated output a result information. However, deep learning program generates output "a probability result information" instead a result information. The diagram below is showing the information processing process of Perceptron in deep learning.



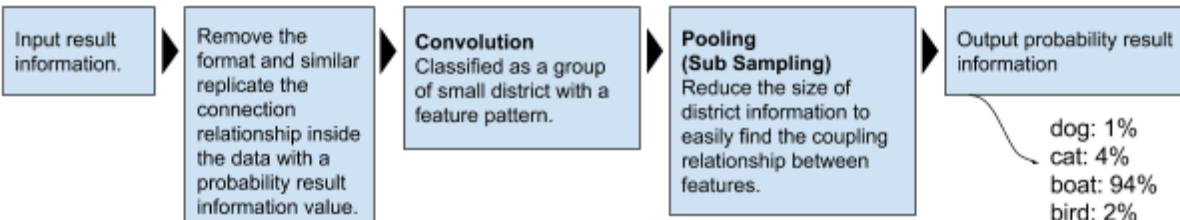
What is probability result information? This is the result information that is used as a basis for probabilistic judgment of other result information. In deep learning, probability result information is a constant form. This probability result information, for simple example is "a feature that means a cat with 80% probability".

In conventional computer programs, a cat is a cat, and a dog is a dog as fixed result information. And an error occurred when the result information that did not match the predetermined absolute result information was input. However, in deep learning, output data is something like "a pattern data(feature) resembling a cat with a probability of 70%" or "a pattern data resembling a dog with 80% probability". Then it overlaps this pattern(feature) to increase the probability.

### Process of Convolution Neural Network



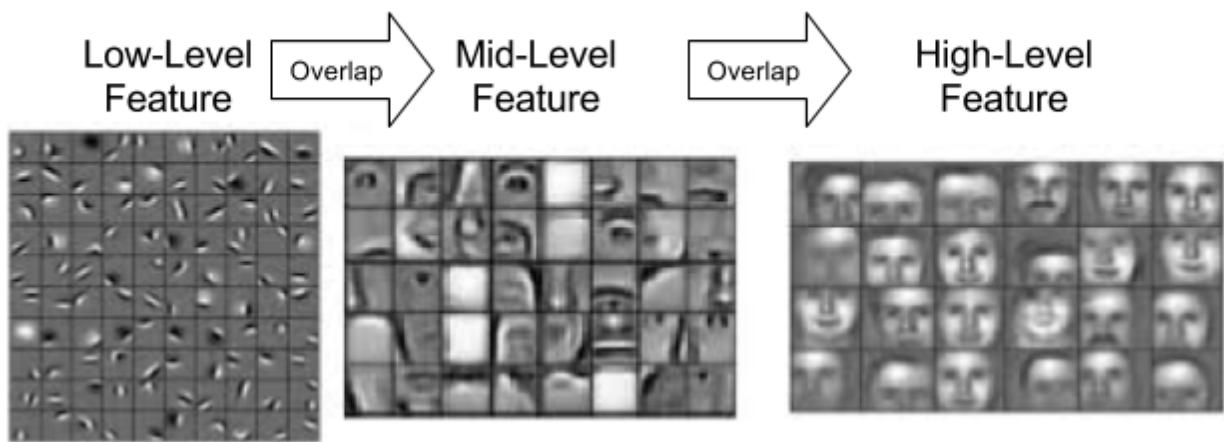
Explain the process with Deduction Theory



The image referred from: [wildml](#) blog

Convolution Neural Network is one of the method in Deep Learning branch. Convolution divides large data into small group of zones and creates a feature or pattern by replicating the connected relationships between the data as probability result information.

### Feature Map in Convolutional Neural Networks (CNN)



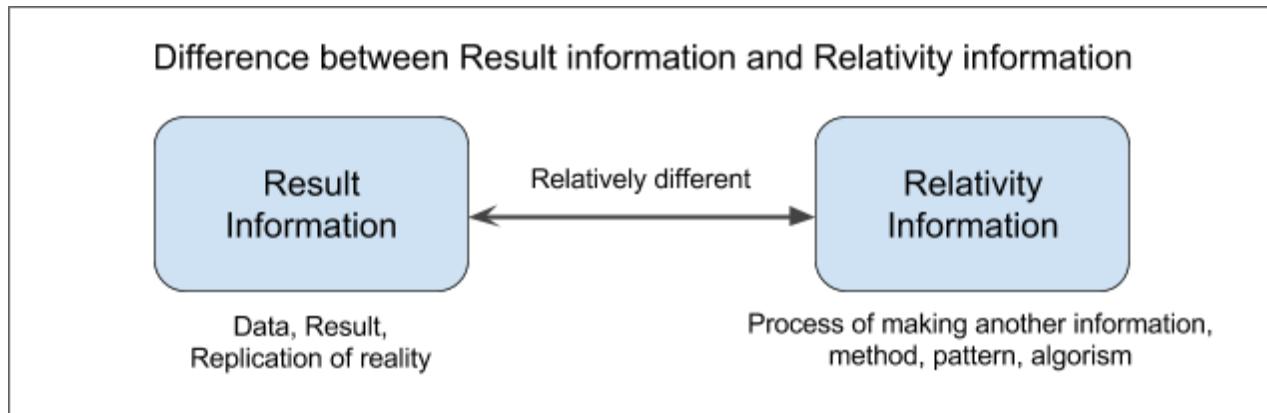
Features are created by each Convolution step. CNN passes through several Convolution stages and overlaps probabilistic result information every time it goes through each step to determine a wide range of information.

Connection relativity result information, Convolution data, and pooling data are all probability result information.

probability result information is generated by overlapping the probability result information

generated at each step.

The reason for naming this feature and pattern as "probability result information" is that it is not a function or algorithm type but a constant type. This "probability result information" is relatively closer to the relativity information between the result information and the relativity information.



Let's compare machine learning and deep learning. In the process of outputting the result information in conventional machine learning, a programmer manually inputs the algorithm from the outside into the program. This is relatively more restricted guidance. In this term, deep learning guide less. In deep learning, programmer let computer make algorithm relatively more by itself. The guidance by the programmer in deep learning is mainly labeling.

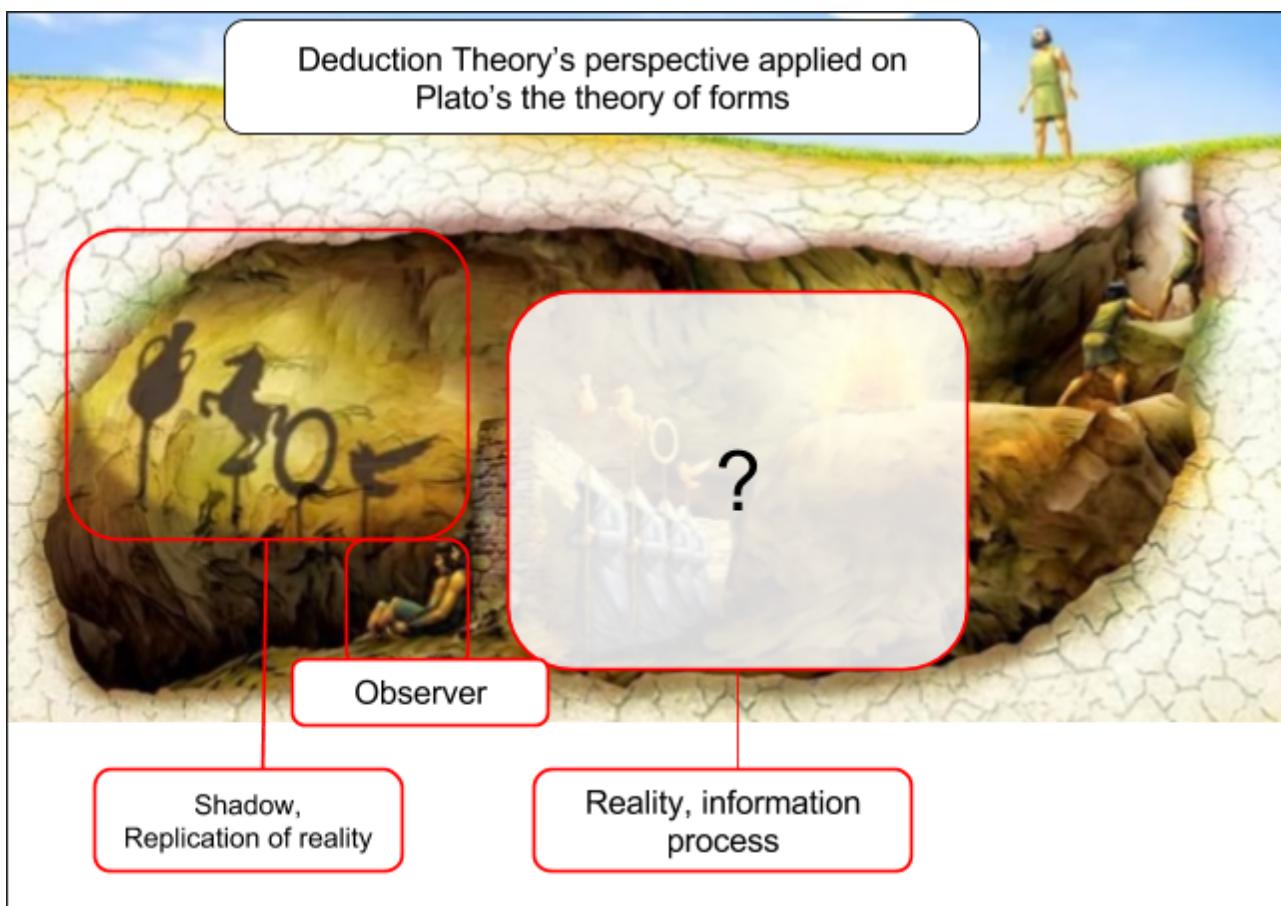
So, what benefits do we get when the program produces output that is closer to relativity information than to result information?

1. Since the program does not make decisions based on absolute result information data, the associated errors will be reduced.
2. In the world of absolute information processing, there are only two options, right or wrong. In the world of relativity information processing, even if it is wrong, learning can be continued to increase the probability of successful information processing.
  - a. Of course, errors also appear stochastic.
  - b. However, if the probabilities are continuously increased by overlapping, the information processing of meaningful trust level can be performed.
3. When a program creates relativity information, it can perform creative information processing activities that had been done only by humans until now.

## Benefits of similar replication of information

I will explain concept of similar replication in Deduction Theory how this process work in deep learning (multi layered perceptron, CNN, DQN). And talk about benefits of similar replication of

information.



Let's take a look at Plato's cave once again. In this picture, we do not know the exact reality, the truth, the information processing process. And we collect datas, result informations which is shadow of reality as observers.

About there, the conventional statistics had this view.

1. Collecting a sufficient enough number of data and classifying it as a specific criterion will lead to a meaningful trend and correlation.
2. It is the concept of "big data" if that transforms "a sufficient number of data" into a terabyte, petabyte "huge amount of data".

Thus, statistics have achieved many results. Many trends and correlations were found. However, there was a limit which the existing statistics could not exceed.

1. Trends and correlations could be relative indicators of decision, but there were always exceptions that did not fit.
  - a. It was impossible to get rid of this exception.
  - b. Also it was very difficult to reduce this exception to a negligible level.
2. Sometime the specific criteria used for classification were wrong.
  - a. It seems the trends and correlations of the statistics fit reasonably well, but later it turned out doesn't fit with reality.

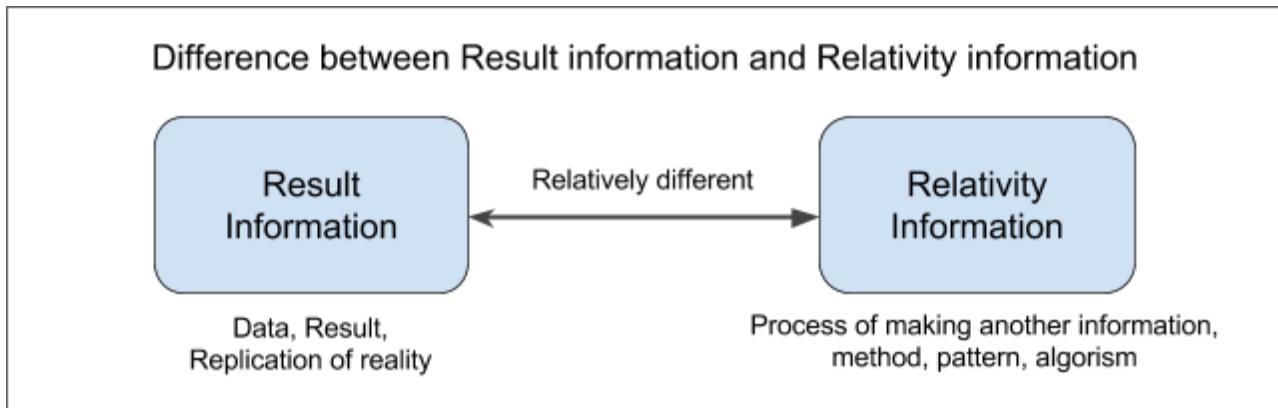
- b. In statistics, therefore, the knowledge obtained through statistics is called correlation meaning relativity, not causality which means absoluteness. Statisticians basically know that they do not deal with the absolute truth while they work on it. This is a good attitude.
- 3. Sometimes adding or adjusting specific criteria to increase the accuracy of the statistics may more of lead to inaccurate trends and poor correlation.

Hmm, what was the problem? Deduction Theory explains:

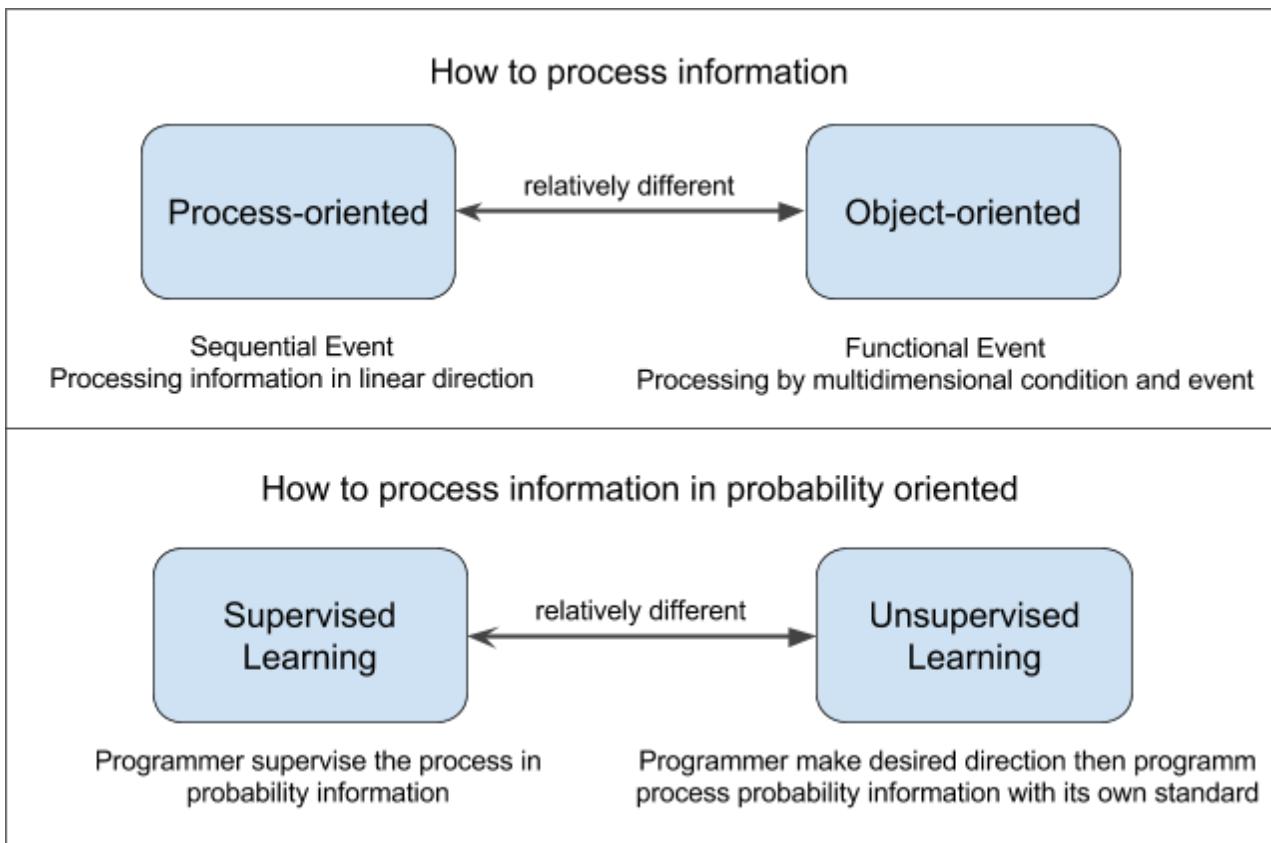
The cause of the problem is the thought that there was a "specific standard." It was the cause that the statistician entered a certain standard.

Conventional statistics generate a specific criterion by a statistician and then classify all data according to that criterion. But where is this particular criterion closer to absolute or relative? It is closer to absolute.

Is this particular criterion relatively close to a criterion based on result information or close to a criterion based on relativity information? It is closer to the result information.



Then, the concept of unstructured data analysis emerged in the big data field. Unstructured data analysis handles unstructured data such as pictures, video, audio, and documents. The specific criteria for classifying data are not absolute. Unstructured data analysis does not classify data by applying absolute criteria but collects data and finds and processes relative criteria such as tag, node, and the other data mining. However, it is the same as before in that statistician set these standards outside of the program.

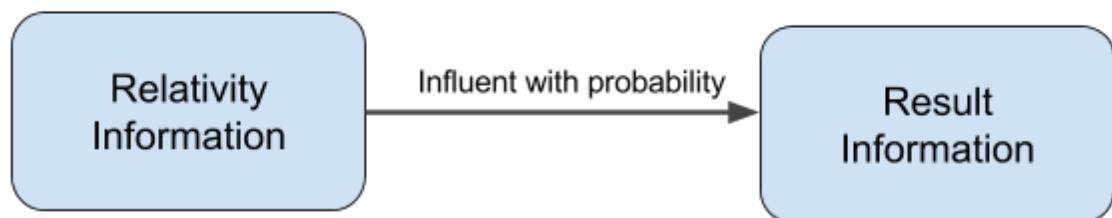


Let's see the machine learning. There are many ways in machine learning. But comparing them to deep learning, we can find out the most noticeable relative differences. During some machine learning, the data is vectorized in the same way as deep learning, and probability information processing is performed. But there is a big difference from deep learning. Machine learning is the process of vectorizing data and then determining how the information is different from the similarity of the information by a specific criterion is given from the outside. Deep learning, on the other hand, leave that part to the Deep Learning program for doing it by itself.

## Deduction Theory's view of relativity



## Relativity Information and Result Information



Relation between Informations that can create the other lower level information

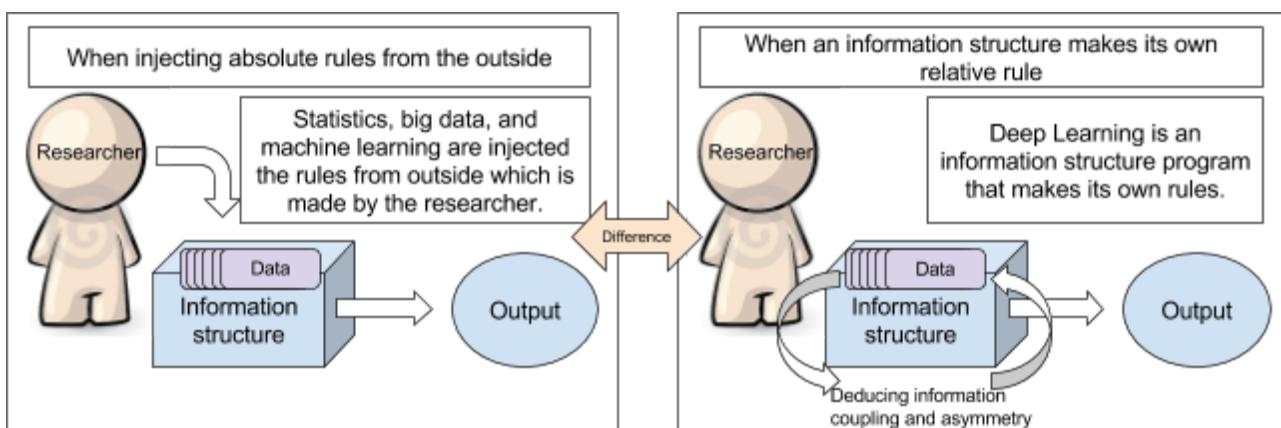
Created by Relativity Information

How is deep learning different? Deep learning does not classify all data into one or two specific criteria that a person has set out. Deep Learning analyzes all the data on the basis of the program itself. Deep Learning finds similarities and differences in each data and collects them.

1. In conventional statistics, a statistician creates a specific criterion and classifies all data by applying the criterion absolutely.
  - a. It forces the program to a criterion based on the result information.
  - b. This criterion is relatively close to absoluteity in the distinction of absoluteity and relativity.
  - c. Thus, errors arise from absoluteity. Why? Because absoluteity does not exist, but because of certain criteria, assumed absoluteity.
2. Unstructured data analytics in the Big Data is more closer to probability information processing than structured data analysis.
  - a. It uses relative criteria such as tags, nodes, and the other data mining that are more loose and multidimensional.
  - b. However, it has the same problem as the conventional statistics that a statistician creates a specific criterion for processing information and inject it to the program unidirectionally.
3. Some machine learning technique uses data vectorization similar like deep learning to infer

similarities and differences between data.

- a. However, it has the same problem as the conventional statistics that a statistician creates a specific criterion for processing information and inject it to the program unidirectionally.
4. Deep Learning analyzes the data and lets the program find out the relative similarities and differences between the data and collect such relative criteria.
  - a. Deep Learning deduces similarities and differences between data within a program. And the programmer does not force a specific criteria from the outside.
  - b. This is a big difference in logical process.
    - i. Does programmer set a specific criterion from the outside and make the program unidirectional(as absoluteity)?
    - ii. Or does the program creates relative standards from within and by itself?



In Deduction Theory, the activity of analyzing information and finding similarities and differences between the information is called "Inference of information coupling relationship and information asymmetry". The similarities and differences between the information itself are relatively close to the relativity information. Thus, the information coupling relationship and the information asymmetry are relatively closer to the relativity information.

Information coupling relationship: Similarities between informations

Information asymmetry: Differences between informations

What about if we reproduce the data in a similar replication form when we do this and then find similar and different point to the replicated data? Then this effect will occur.

1. Similarities and differences become exaggerated while the data is replicated. Thus, possible to find more information coupling relationship and information asymmetry that didn't appear before because they were too small and narrow. The more the relativity standard is obtained.
2. While the data is replicated, the total number and size of the data becomes larger, and the effect of analyzing the mass data is obtained.

Technologies such as Perceptron, CNN, and RNN in Deep Learning play a role in helping to analyze relative criteria by similar replicating information. Deduction Theory is the first to explain the technique of deep learning as inference with similar replication, information coupling and information asymmetry.

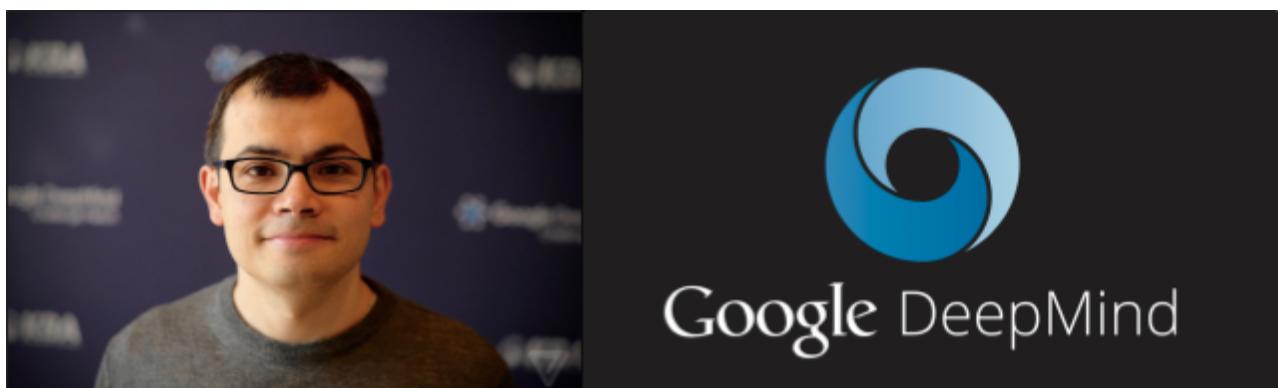
Some Korean industry researchers have commented on our journal about over-fitting and CNN. They said “Overfitting is a concept derived from Machine Learning Regression and is not related to CNN. Over fitting part should removed from CNN section.”. Those who have studied machine learning for a long time seem to have that view.

But it was pointless to us. We analyze how we deal with probability information processing and look at things only from the viewpoint of how we can do probability information processing better. What we wrote is that the method we used in CNN is similar in that it replicates information and creates relative standards, which helps reduce overfitting, which is a problem that occurs in conventional machine learning methods. And we did not claim that CNN was a technology aimed to reduce overfitting. These Korean industry researchers have wasting their time to pointing out what they don't need to.

Now, you can obtain relative criteria by finding similarities and differences in similar replication of informations. Relative criteria serve as relativity information. However, this activity was also happening in the human brain.

## The view of Deduction Theory on cognition, memory, and imagination

We read Denis Hassabis's “The Future of Memory” paper. He is the founder of DeepMind. [10]



We extract some part from the original paper.

.....

The tight linkage between remembering the past and imagining the future has led several investigators to propose that a key function of memory is to provide a basis for predicting the future via imagined scenarios and that the ability to flexibly recombine elements of past experience into simulations of novel future events is therefore an adaptive process.

...

It has been established that the ability to generate specific and detailed simulations of future events is associated with effective coping by enhancing the ability of individuals to engage in emotional regulation and appropriate problem-solving activities. Numerous studies have also established that views of the future are associated with a prevalent positivity or optimism bias. Another promising domain centers on the phenomenon of temporal discounting: people typically devalue a future reward according to the extent of delay before the reward is delivered.-

...

Studies of remembering the past and imagining the future should benefit from establishing closer connections with work  
on narrative processing and the representation of nonpersonal fictional information.

.....

The main contents of the paper are summarized as follows.

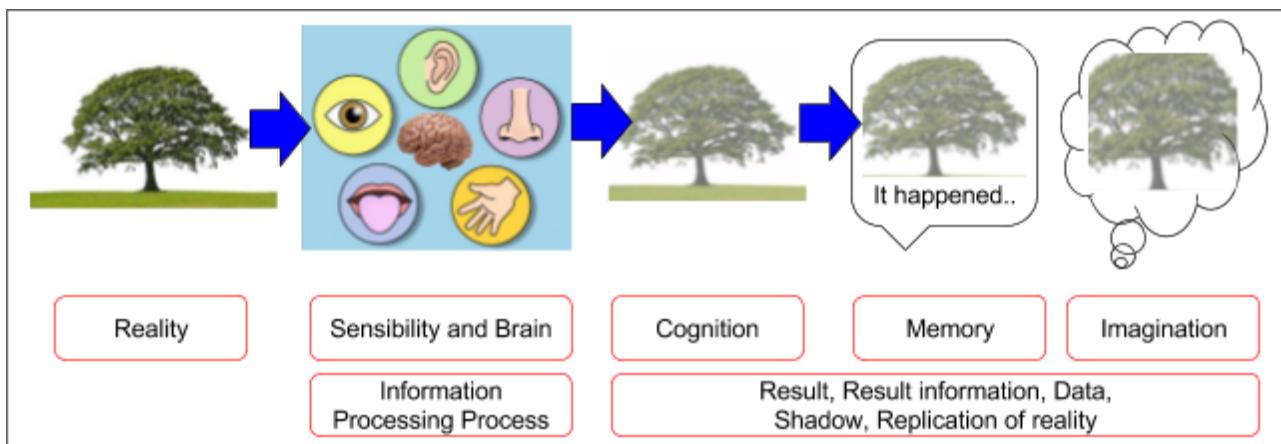
1. Remembering the past and imagining the future have a similar relationship in information processing.
2. The main function of memory is to resemble elements of past experience to simulate fictional events and to create future scenarios.
3. It appears that these evaluation methods are used when reassembling past experience elements.
  - a. Emotional regulation
  - b. Possibility of problem solving
  - c. Reward devaluation when reward point becomes far, and etc.

Deduction Theory has this view on cognition, memory, and imagination.

1. Memory and imagination are generated by the same information processing method in the brain. Both are similar replication of information.
2. The memory stores similar replication of information that is recognized by observing reality.
  - a. This memory is not reality. Because it is a similar replication of reality, not an exact replication.
    - i. It's the same as the data created by a digital camera is like a replica of reality.
  - b. What people perceive is result information, not the reality, and it is the information

that is made by similar replication activity.

- i. Cognition is a similar replication of the reality by transforming the information of reality using the senses of the human.
3. Imagination is another similar replication of information that is recognized and remembered.
  - a. Imagination is not consistent with memory. So it is a similar replication, not an exact replication.
4. Memory and imagination is not always happen in order. There are many cases that imagining while people perceiving reality. This is called daydream and fancy.



The information structure we call intelligence similarly replicates information of reality as cognition, and again similarly replicates information as memory and imagination.

1. Cognition as similar replication
2. Memory as similar replication
3. Imagination as similar replication

The reason for making similar replication is using these as the basis of decision through finding similarities and differences in the informations.

Cognition is not consistent with reality. Memory is not consistent with reality. Imagination is also inconsistent with reality. Therefore, cognition, memory and imagination are not much different from the essence of similar replication. Cognition and memory are considered as that it happened in reality because we believe that it happened in reality. For this reason, many mistakenly think imagination as memory and cognition. When it comes to information processing in the intelligence information structure, it seems to be labeled as "it happened in reality." Thus, memory and imagination can be understood as relative differences by this labeling.

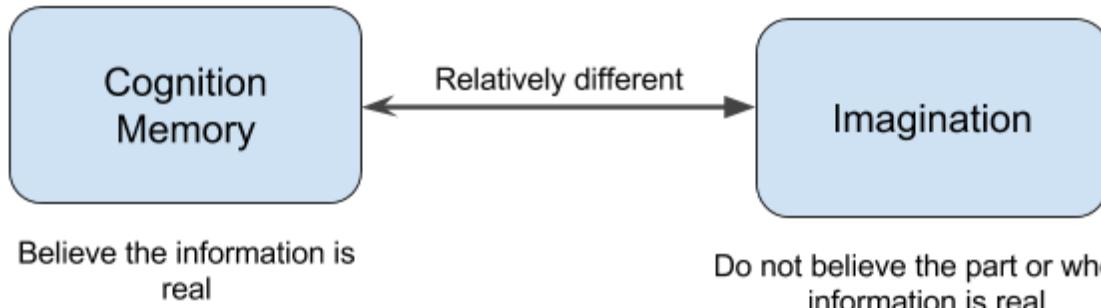
1. Both memory and imagination are similar replication of what they are perceived by observing real-world information.
2. Cognition, memory and imagination are the same activities in nature. This corresponds to a

- similar replication of information.
3. The difference between cognition, memory and imagination is that the intelligence information structure is labeling that information is "happened in reality" about cognition and memory.
    - a. This corresponds to a person's belief that the information about cognition and memory has happened in reality.
  4. For imagination, doing another labeling. Imagination is labeled as "it did not happen in reality."
  5. So, because of this labeling difference, it's possible that memory is regarded as imagination, and imagination is regarded as memory, also cognition is regarded as imagination, and imagination is regarded as cognition by changing the label.
    - a. The characteristics of this information processing are the basis of hypnosis and self suggestion activity.
    - b. It is the basis of advertising, persuasion, agitation and propaganda.
    - c. It is the basis of cognitive bias, cognitive dissonance and stereotypes.



We often see trickery happens in people's cognitive abilities. Magic is an activity that uses human cognitive bias, that is, distortion of cognitive ability. The reason why trick happens is that because people consider it as reality. Most people think that what they perceive is very close to reality. And also most people are very generous about their memories. People think that what they remember is almost always objectively truth. Thus, cognitive dissonance, cognitive bias and stereotypes occur.

## Relationship between cognition, memory and imagination



However, once we understand the concept of similar replication of information, we can also find out about its benefits. As we have already repeatedly explained above, there is an advantage when we accept the concept of relativity and similar replication of information as they are. In other words, it can be used as a relative basis for making decisions and finding out similarities and differences by comparing similar replicated information.

### How to get beyond the limits of cognition, memory, and imagination

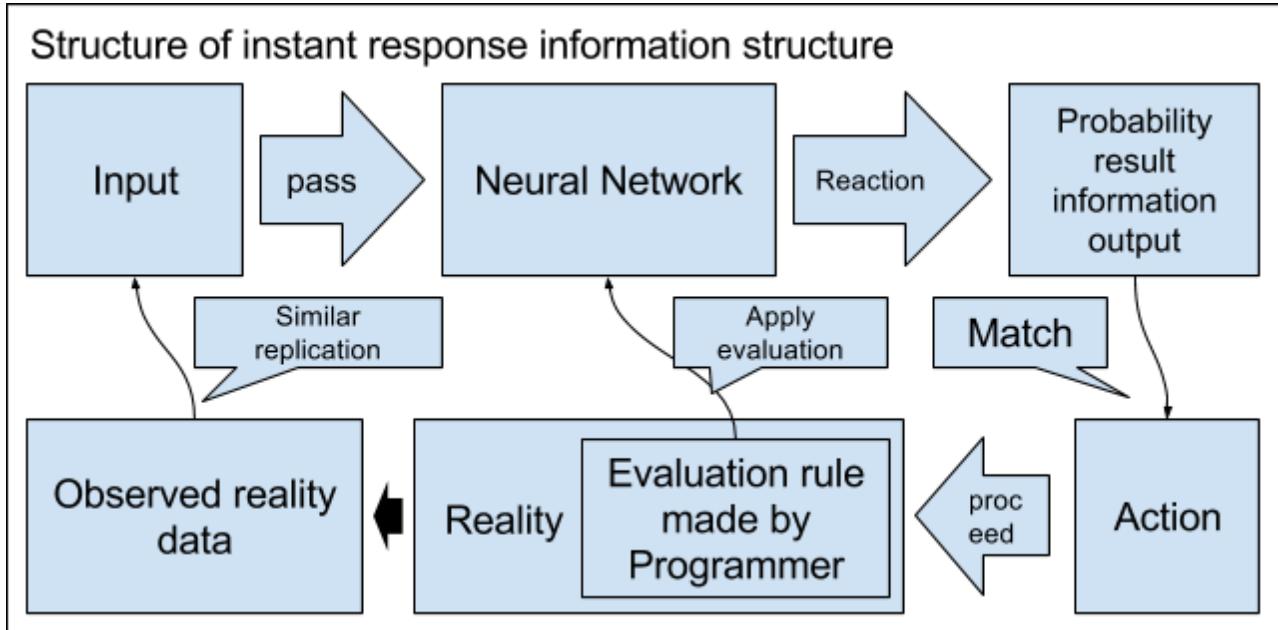
1. Realize that cognition, memory and imagination are similar replication of information.
  - a. Do not believe any information absolutely.
2. Observe the cognition, memory, and imagined information to find differences and similarities.
  - a. Make the decision with using the relative criteria obtained there.
  - b. This is the insight which intelligence has.
3. If you thinking about cognition, memory, and imagination is absolute and believe it blindly, you will have an error. The error is weighted.
  - a. These errors are called the mistake of absoluteity, the error of characterization in Deduction Theory.

After this analysis, I can explain why the artificial neural network studies have grown so fast after using multilayer perceptron and deep layers. This is because it is an implementation of the information processing method of humans that infer similarities and differences between similar replication of information on a computer.

## Deep Learning's Future, Scenario Information Processing

We will call the information processing process that we have researched while studying the multilayer perceptron as "instant response information structure". The reason why we called this an

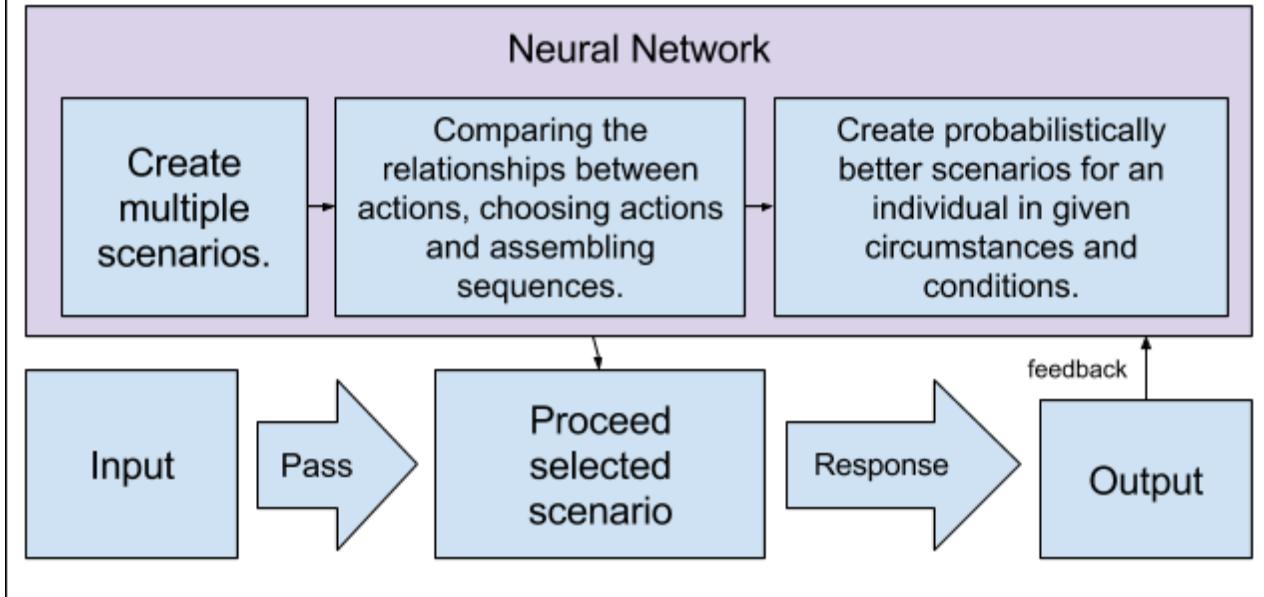
instant response information structure is because the information structure is designed to instant response based on probability result information when it processes information. This is quite similar to the reptilian brain in the human brain's information processing functions.



In contrast to this, there is information processing based on a decision of "Sequence between action and action, information processing process", that is a scenario. There are information processing methods that imagine multiple scenarios and compare them to create better scenarios and execute them. This differs from an instant response to probability result information. We call this scenario information structure.

Scenario information processing is information processing that compares processes and creates and selects better processes. A scenario information structure is an information structure that performs the same processing.

## Design of scenario information structure



Deep learning technology developed to date is relatively close to the instant response information structure. It has not yet reached the "scenario information processing" and "scenario information structure". But this direction will be the future of deep learning.

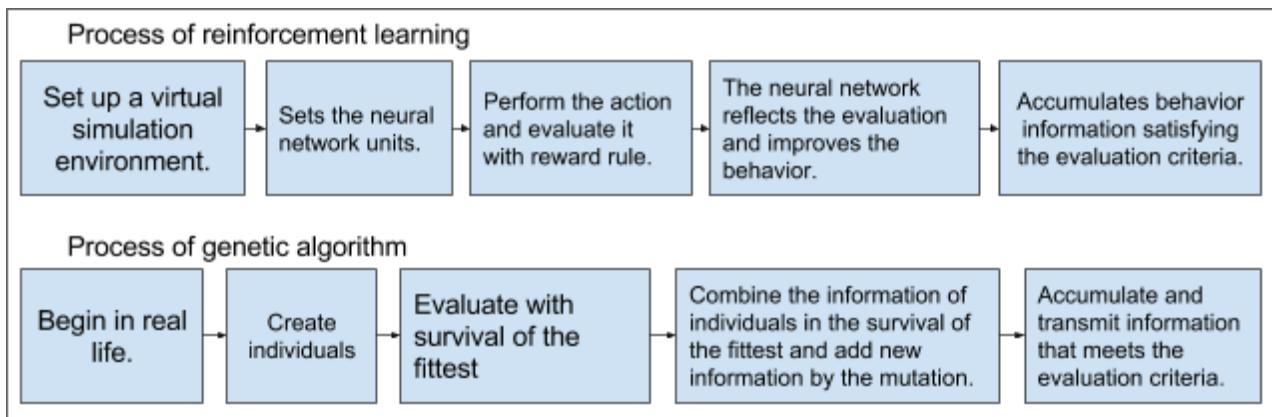
## Comparing DQN and Scenario Information Structure

There is a saying, "Hitting one's head to the ground." in Korean adage. This means that you don't have an experience or a knowledge about the task, but you just do it and get an experiences from it. DQN makes millions of challenge in a virtual simulation space. It builds experience and learns better decision making in the form of reward.

This method also occurs in living organism by DNA. A living organism update DNA information when new life borns from the challenge called "survival of the fittest". So DQN and DNA genetic algorithms have similar parts. For DNA, survival of the fittest is a reward.

The difference between genetic algorithms and DQN is that DQN is a multidimensional experience in simulation space rather than in real world space.

Let's compare the genetic algorithm with the DQN by plotting different points.



Let's look at the movie "Edge of Tomorrow". The concept of this movie is the same as DQN.

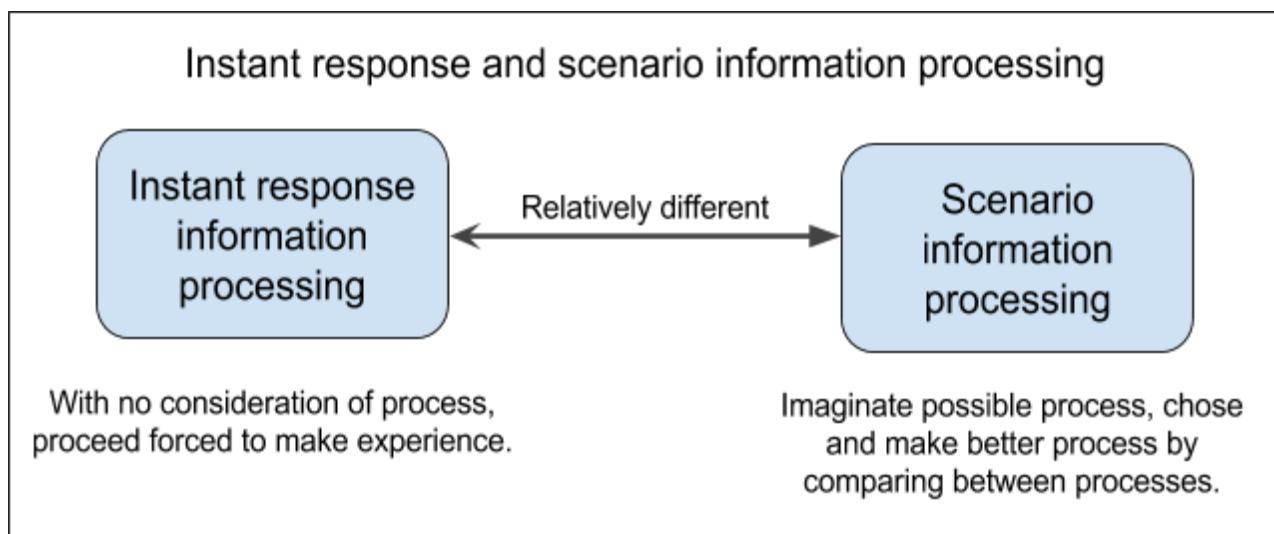
Because the main character(Tom Cruise) made the time becomes multidimensional because of the supernatural power of the alien, he made the reality to a simulation, and then he learned it through millions of times at the ground with head-butting and improved his problem solving ability. Here, let's focus on the multidimensional part of time and the simulation part of reality, except "the superpowers of aliens". This is the core of the DQN.



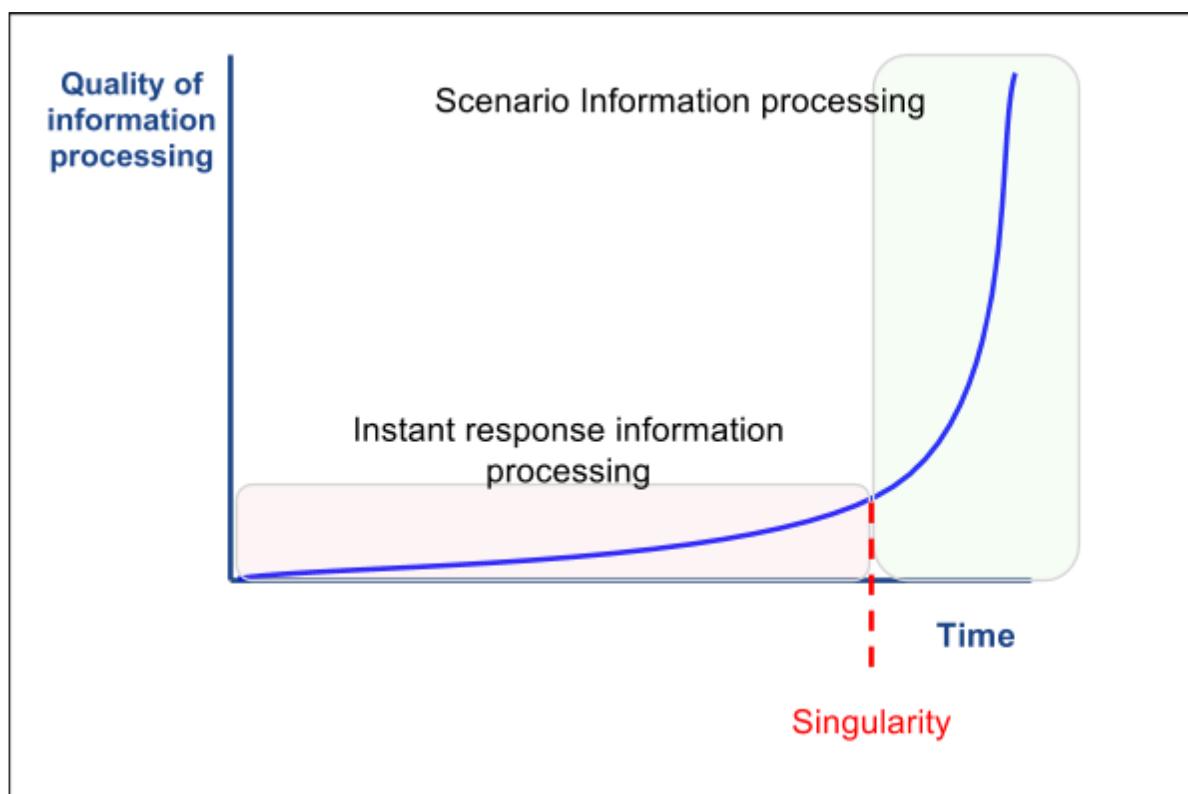
1. DQN makes the experiences in the multidimensional simulation of space and time, then superimposes them on statistical learning.
2. The direction of decision making is determined in the form of reward. This is learning through rewards.
3. DQN is similar to the genetic algorithm in which DNA leaves offspring.
  - a. The difference is that the DQN learns by using the multi dimension of time in the virtual simulation space.
  - b. On the other hand, the genetic algorithm behavior of DNA occurs not in the simulation space but in the real world space which is only one way. Genetic algorithms make a lot of objects and learn by considering individual's experiences as multidimensional experiences.
    - i. For DNA, survival of the fittest is a reward.

The DQN and the genetic algorithm in DNA are considered as the instant response information structure. This is because there is no procedure for comparing the processes inside the information

structure and selecting a better process. This is different from the scenario information structure.



The reality is only unidirectional. Life is only once. Of course, DNA can combine the experience of many individuals with the survival of the fittest to perform probabilistic multidimensional information processing. But from the standpoint of an object, if it's dead, it has done. It can not be repeated many times like a simulation game. Thus, a higher organism developed new information processing methods in order to enhance the survival of the individual in addition to species-level evolution. That is the scenario information processing method that compares the process. This is the singularity of intelligence that occurred in the evolution of life.

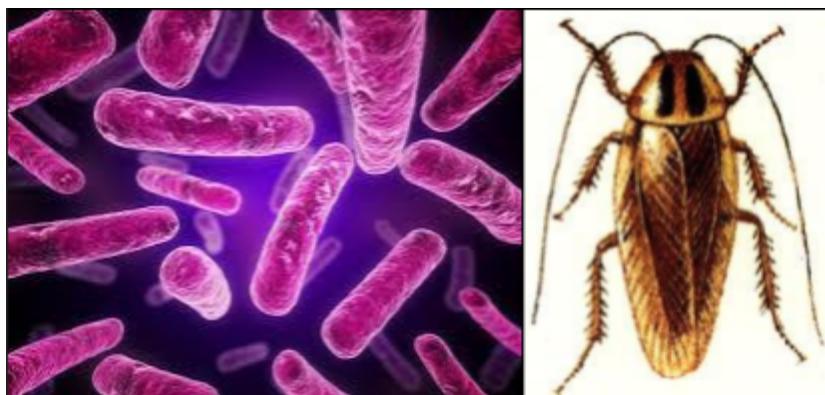


### Features of scenario information processing

1. Based on the experiences that the individual has accumulated so far, it imagines the information processing processes or scenarios, that can be practiced in reality.
  - a. Here we use the mechanisms of memory and imagination.
2. This scenario imagination plays the same role as the simulation space did in the DQN. In other words, it creates the multidimensional nature of time and experience.
  - a. The difference from the DQN is that the simulation space of the DQN is absolutely created by the researcher from the outside of the individual.
  - b. However, scenario information processing is not a simulation space injected from the outside, but a simulation space created by an individual itself.
  - c. Here, the part where an individual creates a simulation space by itself becomes an information processing method that is closer to the relativity information described in Deduction Theory.
3. The individual imagines a scenario and finds similarities and differences between the scenarios.
  - a. It is not known what the scenario will lead to in the real world.
    - i. Because in this world, absoluteness does not exist. Assuming absoluteness and blindly believing in any scenario, we will encounter unexpected errors. The error increases with time.
      1. In terms of hardware, human beings who have already been able to process the scenario information also have repeatedly made this mistake.
      2. This mistake is called the mistake of absoluteity, the error of characterization in Deduction Theory.
  - b. However, it is still possible to find similarities and differences between scenarios, even if we cannot be surely known what the consequences will be in reality.
  - c. Furthermore, it is possible to accelerate this information coupling relationship and information asymmetry inference by similar replicating scenarios.
4. Then we can find the probabilistic characteristics of each scenario.
  - a. And we can stack up the probabilities.
5. So we create better scenarios and implement them to improve its problem-solving skills.
  - a. This is a way to improve the problem-solving ability by creating a virtual simulation space on its own by making the time multi-dimensionally.
  - b. From the point of view of ancient religions and philosophy, scenario information processing has the same effect as living a life multiple times from the perspective of an individual who lives only once. In other words, it is realized eternity.
    - i. But this is not a physical eternity. It is eternal in the concept of metaphorical logic information processing.

The lower level organism does more closer to the instant reaction information processing. This is a relative comparison. The closer to the lower level organism, the more instant response information

processing will be achieved. Lower level organism often show "evolution patterns after endangered species"; once most of the species die and decline dramatically, a large-scale evolution takes place in an attempt to adapt to the environment at the DNA level. The external environment includes competition with other species and other entities besides the physical environment.



How does DNA recognize the extinction crisis, the rapid decline of the entities? DNA perceives a rapid decline in entities as a lack of diversity in gene pools in reproduction. It is easy to detect that the genetic diversity is reduced and inbred mating. It then accelerates evolution in that direction. It also adds new information in the form of mutations. The genetic algorithm of this DNA is applied to improve breeds of livestock and crops.



Lifes evolves in response to the external environment. Conversely, if they do not have any environmental challenges and threatening by survival, they won't evolve. This is the basic principle of evolution so far known. So, there are creatures that have continued without evolutional changing for millions of years.

## Lifes which are called a living fossil

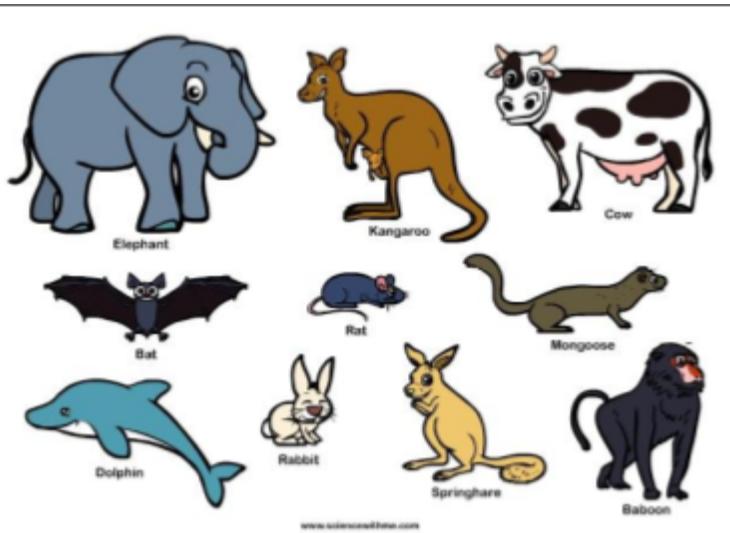
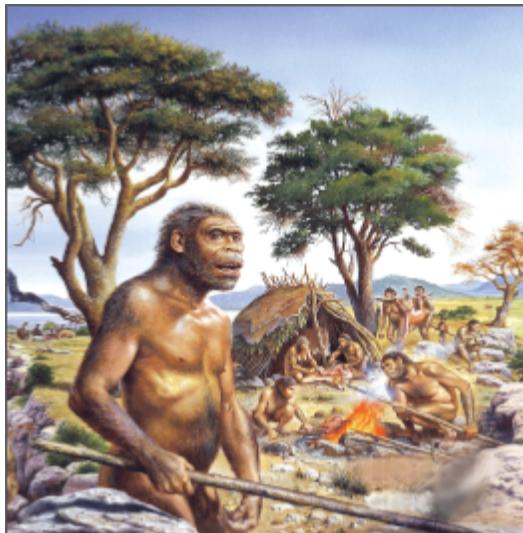


Tadpole shrimp



Ginkgo biloba

The lower level organism does more closer to the instant response information processing and the evolution of DNA is also relatively close to instant response information processing. Higher level organism does relatively more scenario information processing with comparing process. Humans are the most developed of these abilities in life forms.



Using this way increases the ability of each individual to stand up against the dangers of the outside environment. It is the ability to change the external environment to its advantage. It is the ability to create information structures and use them as tools. It is called "meme" in the evolutionary anthropology that the ability to organize information and use it as a tool is learned and imitated.

Why does scenario information processing become more resistant to external environmental hazards? The reason is simple. Let's make two comparisons.

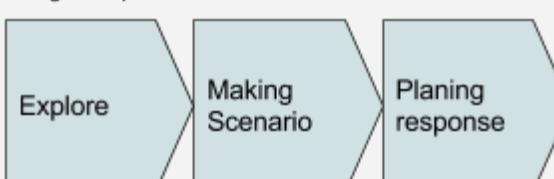
1. When the environment changes as risks, the individual responds or reacts to it.
2. Regardless of environmental changes, thinking about multi scenarios of possible good and bad situations, comparing the scenarios, and creating a scenario that responds to all possible risks and opportunities in a multidimensional manner.

- a. Think about how to prevent going to bad situation.
- b. Think about how to make a better situation.
- c. Find out the similarities and differences between the scenarios and identify the probabilistic characteristics of each scenarios.
- d. Build the long-term goals and long-term roadmap by assembling the scenarios.
- e. Run the scenario and feedback the information coupling relationship and information asymmetry of how closely it matches the real world.

This comparison shows that scenario information processing is relatively much more capable of standing on environmental changes. We can also see a similar phenomenon when we look at a company where people work together. Some companies does not respond until after they have suffered substantial losses due to changes in market conditions. But a wiser company prepares itself for a scenario that is relative risk and relative opportunity in advance. It is called scenario planning in business administration. And this was derived from the game theory of mathematics.

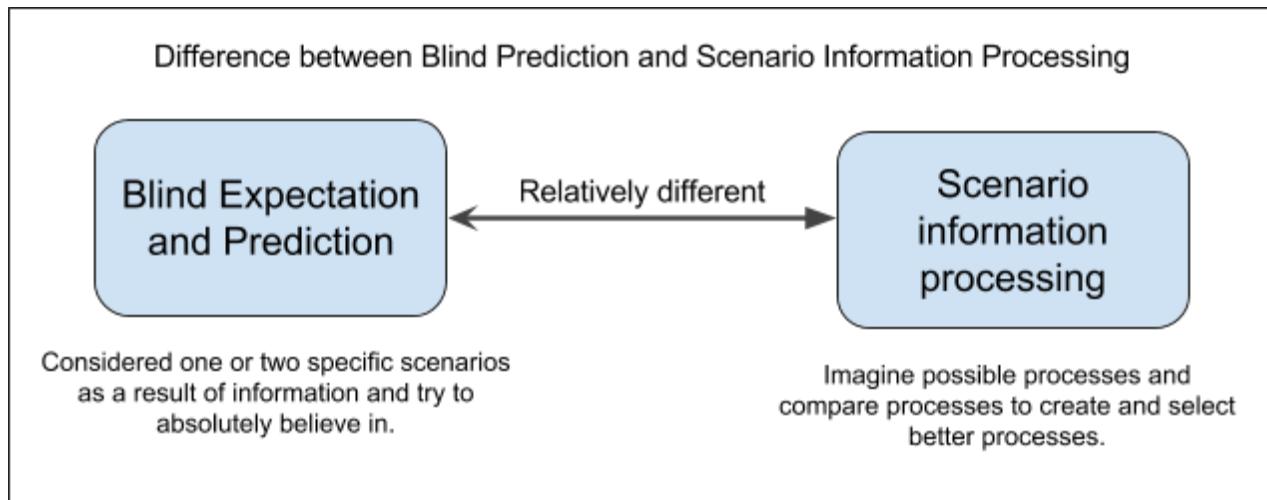
<b>NOKIA</b>	<b>SAMSUNG</b>	<b>Google</b>	<b>Apple</b>
Responds late so failed	Responds late but survived	Prepare for the future and create a new market	

Interestingly, when teaching scenario planning in the class room, there is an emphasis by the leader to the students: "The scenario is not a tool for blind future predictions, but a probabilistic tool for managing possible risks and opportunities.". What does this mean? It means not to believe one or two specific scenarios are said to mean absolutely and consequently that it will happen in the future. If you believe in a scenario like that, you will not be able to respond properly when things that does not fit in the scenario happen in real life. More easily, it means you will fail.

What is Scenario Planning?	Prediction VS Scenario Planning				
<p>A strategic planning method for making meaningful long-term plans</p>  <p>Explore      Making Scenario      Planing response</p> <p><b>! Scenario planning is not a tool of prediction</b></p>	<table border="1"> <thead> <tr> <th>Prediction</th><th>Scenario</th></tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> <li>- Future can be predicted</li> <li>- Uncertainty can be removed</li> <li>- Compress future as numbers</li> <li>- Make only one future</li> <li>- Wanting to know the future</li> <li>- Concentrate on near future</li> </ul> </td><td> <ul style="list-style-type: none"> <li>- Future is not predictable</li> <li>- acknowledge uncertainty</li> <li>- Describes future as a story</li> <li>- Make multi future scenarios</li> <li>- Want to prepare for future</li> <li>- Looking forward to further future</li> </ul> </td></tr> </tbody> </table>	Prediction	Scenario	<ul style="list-style-type: none"> <li>- Future can be predicted</li> <li>- Uncertainty can be removed</li> <li>- Compress future as numbers</li> <li>- Make only one future</li> <li>- Wanting to know the future</li> <li>- Concentrate on near future</li> </ul>	<ul style="list-style-type: none"> <li>- Future is not predictable</li> <li>- acknowledge uncertainty</li> <li>- Describes future as a story</li> <li>- Make multi future scenarios</li> <li>- Want to prepare for future</li> <li>- Looking forward to further future</li> </ul>
Prediction	Scenario				
<ul style="list-style-type: none"> <li>- Future can be predicted</li> <li>- Uncertainty can be removed</li> <li>- Compress future as numbers</li> <li>- Make only one future</li> <li>- Wanting to know the future</li> <li>- Concentrate on near future</li> </ul>	<ul style="list-style-type: none"> <li>- Future is not predictable</li> <li>- acknowledge uncertainty</li> <li>- Describes future as a story</li> <li>- Make multi future scenarios</li> <li>- Want to prepare for future</li> <li>- Looking forward to further future</li> </ul>				

However, by making the worst case and possible risk scenarios that can happen realistically and

by preparing a way to prepare for them, the damage can be greatly reduced when the risk actually occurs. Likewise, by creating opportunities for realistic scenarios and making effective use of those opportunities, you can realize opportunities when the opportunity occurs. This is a big difference.



1. Scenario information processing does not blindly believe that certain scenarios will happen in the future.
  - a. If you consider one or two of specific scenarios as a result information and absolutely believe them, you will have blind expectations and predictions.
  - b. It is impossible to know everything about the future from the standpoint of the individual. This is because of the relativity of information in Deduction Theory.
  - c. By blindly believing in a particular scenario, when something that does not fit in the scenario is happening in real life, it becomes impossible to respond properly.
  - d. In absolute worldview and result information processing worldview, If you try to process scenario information by using these worldviews, you will fall into those mistakes above.
  - e. This mistake corresponds to the mistake of absoluteity, the error of characterization in Deduction Theory.
2. Scenario information processing should be used as a probabilistic tool to manage possible risks and possible opportunities.
  - a. In the worst cases that can occur in reality, making the possible risks a scenario and preparing them ahead of time can greatly reduce the damage when the risk actually occurs.
    - i. Furthermore, the risk can be prevented from occurring.
  - b. Making a realistic opportunity for a scenario and creating a way to take advantage of that opportunity in advance can realize them when you have the opportunity.
3. So, scenario information processing is to create a multi-dimensional probabilistic nesting scenario that compares the processes to prepare for various events that may occur in the future.
  - a. Probabilistic tools, and multidimensional probabilistic overlapping scenarios

correspond to relativity information processing in Deduction Theory.

The characteristics of scenario management have already been developed for decades as a sub-theme of game theory in mathematics and business management . However, it is Deduction Theory that the first time to explains why and how those are.

Let's organize.

1. When it closer to the lower level organism, it does more instant response information processing.
2. The DNA of living organisms respond instantly to the environmental challenge.
3. When it closer to the higher level organism, it does more scenario information processing by comparing process relatively.
  - a. Humans are the most developed of these abilities.
4. Using scenario information processing capability increases the ability of each individual to stand up against the risks of the external environment.
  - a. It changes the external environment to better on itself.
  - b. Create an information structure itself and use it as a tool.
    - i. An information structure refers to how to do things, what process to do things and know-how.
    - ii. It is called meme in evolutionary anthropology to transfer this ability by learning, imitation and etc.
5. Scenario information processing is the process of comparing multiple scenarios to create a multi-dimensional probabilistic nesting scenario to prepare for various events that may occur in the future. It does not blindly believe in choosing a particular scenario.

## Future research tasks

We will analyze the existing deep-learning technology in the future and organize it as Deduction Theory and to create intelligence information structure in a further developed direction. As we have written, Deduction Theory explains outcome of the deep learning that has been done so far theoretically. And Deduction Theory provides theoretical ideas on which parts of the future can be improved to create better intelligence information structures. This attempt is the first by us.

Let's repeat this verse comes out in a book called 'The Book of The Way' written by Lao-Tzu: "The Way that can be told of is not an unvarying way; The names that can be named are not unvarying names." .

道可道 非常道, The Way that can be told of is not an unvarying way.

名可名 非常名, The names that can be named are not unvarying names.

[Tao\\_Te\\_Ching](#) The book of the way - Wikipedia

This scripture, presumed to have been written more than 2500 years ago, now may be rewritten as Deduction Theory.

If you name the reality and put it into the result information, it is no longer reality. This is the relativity of the world we live in. However, we can find similarities and differences in the result information and access the truth through relativity information. This way, we can find out the origin of all things and create all things.

## Conclusion

In this paper, we explained concept of Deduction Theory and show how it applies on Deep Learning. We proved theoretically why deep learning is better than the other methods. We study Deduction Theory long before Deep Learning gets popular, we see the success of Deep Learning as proof of rightness on our direction. We presented a better intelligence information structure than Deep Learning now. Our goal is to create better decision making system, better information structure and Relativity Information Intelligence. Relativity Information Intelligence is an information structure that relatively close to object-oriented, probability-oriented, unsupervised learning, relativity information processing, scenario information processing and massive automated information processing. We made a new definition of relativity in Deduction Theory. We logically clarify a difference between absoluteity and relativity. And it shows what is more close to relativity information and what is not. This method gives an ability to choose relatively close direction to achieve Relativity Information Intelligence.

## References

- [1] Bertrand Russell (1919) "Introduction to Mathematical Philosophy". London: George Allen & Unwin. (ISBN 0-415-09604-9 for Routledge paperback)
- [2] Gottlob Frege (1879) Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Halle.
- [3] Gödel, Kurt (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". 《Monatshefte für Mathematik und Physik》
- [4] Aizerman, M. A. and Braverman, E. M. and Lev I. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control,

25:821–837, 1964.

[5] Frank Rosenblatt (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386–408. doi 10.1037/h0042519.

[6] David Chalmers, Robert M. French & Douglas R. Hofstadter (1990) [High-Level Perception, Representation, and Analogy: A Critique of Artificial Intelligence Methodology](#). Journal of Experimental and Theoretical Artificial Intelligence 4:185-211, 1992. Reprinted in (D. R. Hofstadter) Fluid Concepts and Creative Analogies. Basic Books.

[7] Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, Jürgen Schmidhuber (2015) "[LSTM: A Search Space Odyssey](#)"

[8] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), 1746–1751.

[9] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg & Demis Hassabis (2015) "[Human-Level Control through Deep Reinforcement Learning](#)". Nature. 529 – 533

[10] Schacter D, Addis D, Hassabis D, Martin V, Spreng N, Szpunar K (2012) "[The Future of Memory: Remembering, Imagining, and the Brain](#)". Neuron. 76, 677 – 694