Brain Science of Creativity

Beyond Complex Systems Theory of Biological Systems

Kazuhiro Sakamoto

Preface

Our minds are filled with a variety of thoughts and feelings, such as "Oh, let's do this!" or, "I would die for this person." How should we understand these things that are created and generated in our own minds?

Many readers may think that such issues are just a matter of sensibility and will forever be a mystery wrapped in an enigma, just as the inside of genius inventors or artists cannot be understood.

On the other hand, I'm sure many of you have wished that you could have a robot like Doraemon, famous Japanese manga robot character, that would stay with you and solve all your problems. If we wish for such a robot, we have to figure out how to process information and come up with something on the spot, even when an unexpected situation arises, as a science and embody it as a technology. In this sense, the goal of the "brain science of creativity" described in this book is not a pie in the sky, as many researchers and engineers arhttps://www.deepl.com/ja/translatore now working hard to develop robots.

No, artificial intelligence, which is making great strides, will make it happen in the near future, won't it? Many readers may think so. Unfortunately, current artificial intelligence and machine learning can only do what it has learned to do. It can learn something we don't expect, which may surprise us and seem creative at first glance. However, it is not explicitly equipped with the principles of creating something flexibly and on the fly.

Some people may say, "Well, even if that were true, brain science is advancing rapidly, and in time, we will discover new principles, won't we?" Again, unfortunately, just piling up detailed research will not get us to what we are looking for. We need to think carefully about the questions, "What is the problem?" or "How should we understand how the brain works?"

In this book, I will mainly discuss how, if it's possible, science can capture what is created and produced in living creatures and the brain based on my research and other related studies. Unfortunately, I cannot go so far as to say, "If you do this, you can make Doraemon!" However, I tried my best to answer the questions, "What are the problems that need to be solved?" and "What do we know about the problem?

Specifically, we will approach from one field of science that can be called the complex systems theory of life or complex systems biology. This theory aims to elucidate the basic principles that enable living systems (brains, robots, etc.) to adapt and work well in the real world, and has developed from the academic field of nonlinear nonequilibrium thermal and statistical mechanics. I believe that having such a field as a foundation will pave the way to a scientific understanding of molecules and atoms and, in a sense, to a unified understanding of life and the creative aspects of the mind.

It would be my great pleasure if this book could bring a calm perspective to the current excessive expectations of artificial intelligence and brain science, and help readers to deepen their understanding of life and mind.

Index

Part 1. Viewing the Brain as a Complex System

Chapter 1. Thinking about the Brain in Terms of the Complex Systems Theory of Life

This book discusses the creative aspects of the brain from the perspective of the complex systems theory of life.

The complex systems theory of life is an attempt to discuss living systems from the perspective of complex systems theory. This theory deals with the generation of temporally and spatially ordered structures (this is called emergent phenomena of complex systems). Particularly in complex systems of living systems, the interaction between part and whole, micro and macro, is an important issue.

The question of how to build harmonious relationships between parts and the whole, between micro and macro, also leads to a fundamental social issue: what kind of relationship is desirable between individuals and society, between human society and the global environment? Therefore, I believe that studying the brain as an organ responsible for observing and interacting with the world from the perspective of the complex systems theory of life will not only reveal the principles behind the improvisational and creative information processing of the brain seen in various situations, but will also be useful in solving these serious social problems.

In this chapter, after discussing the reasons for the current great social interest in brain science, I will delve one step further into why we need to consider the brain from the perspective of the complex systems theory of life, and give an overview of the theory.

I. Reasons for the Interest in Brain Science

Aging Population, High Stress, and Information Technology

People's interest in the brain continues to grow. In bookstores, there are many books on the subject, such as “Train Your Brain with XXX.” People from various fields participate in the Japan Neuroscience Society, and it makes me feel that brain research is active. It is even more active in the U.S. The annual meeting of the Society for Neuroscience Society is attended by about 30,000 people, which is about 10 times that of the participants in the Japan Neuroscience Society. In his 2013 State of the Union address, President Obama mentioned the brain as one of the areas of science and technology in which the U.S. should invest heavily, indicating that the U.S. is considering the brain and neuroscience as one of the keys to lead the world in the future.

So, why is brain and neuroscience research so active today?

The first reason is probably the aging of our society. As the number of elderly people increases, so does the number of patients with dementia (see Fig. 1.1). When you have dementia, it is difficult not only for you but also for your surroundings. This increases the need for dementia-related research. The number of patients suffering from stroke (cerebrovascular accident) is not small, although many of them survive, and according to the "Summary of the 2014 Patient Survey" by the Ministry of Health, Labor and Welfare of Japan, the number of patients is approximately 1.2 million in Japan. Naturally, there is a growing need for radical treatments, efficient and inexpensive rehabilitation methods, and devices that support our lives.

Secondly, there is a growing desire for computers and machines that can perform advanced information processing and operations in place of humans. This is exactly what the recent interest in robots, artificial intelligence and deep learning is all about. It is understandable that research on information processing and motor control mechanisms in the brain is active.

The third reason would be the increase in the number of people with mental problems. Of course, problems of the mind are a universal problem. Since ancient times, philosophers and religious people have been interested in none other than the problem of the mind. However, the hustle and bustle of modern society has increased the number of patients with depression (see Fig. 1.1) and the need for solutions.

However, I think that the growing interest in brain research is not only due to the state of modern society, but also due to human activities on a larger time scale, and therefore there is a reason that many researchers are not aware of. In my opinion, it is about worldview, or how we perceive this world.

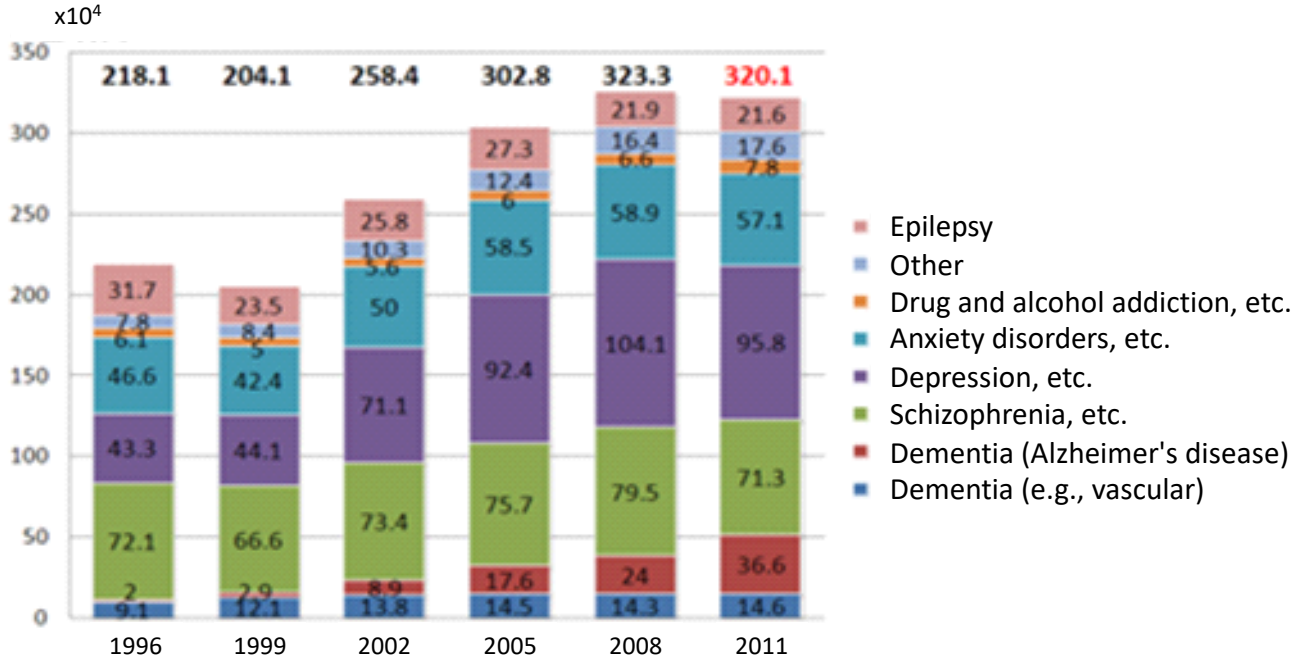


Figure 1.1. Number of patients by mental disorder in Japan (from Ref. 1).

Unidirectional World Images Underlie the Modern Times

The image of the world, that is, the image of what this world is like, tends to vary not only from person to person, but also from time to time and place to place. The image created by the West in the 18th and 19th centuries seems to have a distinctive feature. In my opinion, it is a kind of unidirectional world image.

The best example of the unidirectional world image is the one symbolized by “Laplace's Demon.” It is a deterministic, causal, and deductive image of the world proposed by the French mathematician Laplace. The idea is that any future can be predicted, if there is a demonic intelligence that can know the position and momentum of all matter in the world at a given moment, and can completely calculate its changes using the laws of classical physics. It is as if things are determined in one direction, from the past to the future.

Science involves measurement and observation. In the world image created by modern times, observation is also “unidirectional.” Observations in early science did not affect the results of the observations. I don't think that Tycho Brahe's observations of the planets, which led to the dawn of modern science, were influenced by the fact that he looked through a telescope at the stars. Early science began with such things, that is, those that did not require consideration of the interaction between the act of observation and the object to be observed.

In the world of mathematics, too, something “unidirectional” was about to be sought. Hilbert's program, which called for a restructuring of mathematics so that all propositions could be proved “unidirectionally” from a few axioms, is a good example.

Modern society also seems to have been “unidirectional” in a sense. People's images of the world are often influenced by social conditions, often in ways they are unconscious of. The above-mentioned image of the world created by modern science should have been influenced by the mood of Western Europe at the time from the Age of Discovery to the Age of Imperialism. I haven't lived at that time, so I didn't experience the atmosphere, but it might be like, “The world is big enough. Let's expand more and more! Let's make a lot of money by mining natural resources and throwing garbage away!!” Here, too, I feel “unidirectional” views of things in terms of advancement, exploitation, and disposal. The reason why such an atmosphere was not so problematic is because the scale of human society's activities was still sufficiently small compared to the global environment. Even if humans had a negative impact on the environment, it was still somehow within the resilience of the environment.

This is not the case today. The establishment of quantum mechanics, the discovery of chaos, Gödel's incompleteness theorem, pollution, and global environmental problems that have emerged in the 20th century seem to stand directly against maintaining a unidirectional image of the world.

The Contemporary Times Highlight Bidirectional World Images

In contrast, what the contemporary times highlight, and what we need to deal with, is what we can call bidirectional world images. In other words, they are images of the world in which the subject (observer/actor) and the object (environment/other) interact so strongly that they cannot be separable. However, the fact that such a world image is becoming apparent is different from whether people have such a world view. When I talk to top scientists, I often secretly feel that they have an unidirectional view of the world. Therefore, in this discussion, I will use the words "world image" and "world view" differently.

Aside from various philosophical discussions here, the subject refers to something that observes the world and acts on the world in some way. In this discussion, the world, which can be called the object in contrast to the subject, refers to the environment surrounding the subject and the others the subject faces.

You may wonder what I am referring to, but it is something very familiar.

For example, your own family. It may be possible to sneak in a hidden camera and observe how your family behaves when you are not around. However, as a husband and father, you are faced with the questions of how to guide your family in the right direction when your words and actions affect them and their words and actions affect you.

The same goes for environmental issues. What human society spits out is beginning to torment humans themselves through the global environment. We are in a situation where the subject of human society and the object of the global environment are strongly interacting with each other, and we have to somehow bring them together in a positive direction.

Another example that is more limited is the power transmission system when buying and selling is significantly liberalized. As liberalization progresses and the number of participants increases, there is a possibility that various businesses will speculatively buy and sell electricity, just as if they were buying and selling stocks. These companies will carefully watch the current power companies, which will remain be major players. However, if the power companies continue to be responsible for the transmission system as they are now, they will need to control the system so that it does not go down, anticipating how their behavior will cause speculators to behave. In the sense that they have to internally observe the state of the whole transmission system and achieve integrity of the entire system, there is something similar to slime molds that will be outlined later.

In order to understand the whole world in a unified manner in the contemporary times, it is becoming more and more necessary to understand the world, including subjects who observe the world and work on it. This situation is probably the unseen historical pressure that is pushing brain research forward. This pressure may be the reason why brain science is so popular right now. I believe it is an academic field that has the potential to bring some perspective to the overly fragmented science and confused society.

Life Cannot be Understood in Molecular Terms Alone

So, if we simply work hard on brain research, will we naturally gain a unified understanding of the world? If we study the molecular level in detail, will we be able to find a way to achieve a harmonious and consistent relationship between the subject and the world constituting a strongly interacting system? If I am asked these questions, I would have to answer that you won't get it so easily.

The idea that we can understand the whole thing and its behavior by dividing it into its components, such as molecules and atoms, and clarifying the components is called (element) reductionism. There are more researchers than I imagined who implicitly assume (and sometimes unabashedly stated!) reductionism.

It is true that when you have a problem, it is very useful to identify the key factors or elements. When debugging a program that is not working well, the first thing to do is to identify the statement that is causing the error. If you have a thorn stuck in the sole of your foot, you should first remove the thorn.

Similarly, when trying to understand life or, conversely, to solve problems that threaten the survival of life, there are many cases where specific components and substances are crucially important. Many diseases that were once feared as incurable were caused by infection with specific bacteria or viruses. Eliminating such specific causes of diseases would greatly reduce the number of patients suffering from such diseases. It is well known that the number of cases of tuberculosis, caused by Mycobacterium tuberculosis, has been greatly reduced with the advent of antibiotics.

However, is it possible to truly understand life without considering the "organic relationships" among these components and substances?

For example, an electric circuit works only when the parts that make it up are correctly wired according to the purpose of the circuit. Of course, some parts will be indispensable. However, understanding them is not necessarily the same as understanding how and why the electrical circuit works.

The same is true when you truly understand life. Today, it is possible to examine the DNA of ancient mummies. However, no one would think that a mummy is alive just because DNA was obtained from it. When the person of the mummy was alive, the information in the DNA might have been expressed at the right time and place to perform some function, but that cannot be expected in the mummy state.

In the first place, the reason why I aspired to the study and research described in this book was because I had great doubts about the element reductionist way of thinking when I was young. My younger sister lost her mental balance and stopped going to school when I left home for higher education. Eventually, she was taken to the hospital by our mother and prescribed antidepressants. I felt a sense of remorse that I had triggered the problem, but I also thought, "This is not a problem that can be cured just by taking antidepressants. It's a problem with her living environment, human relationships, and the system surrounding her. It's like giving cold medicine to a person living naked in the middle of winter when he catches a cold!" I couldn't help but be furious at the stupidity of the doctor.

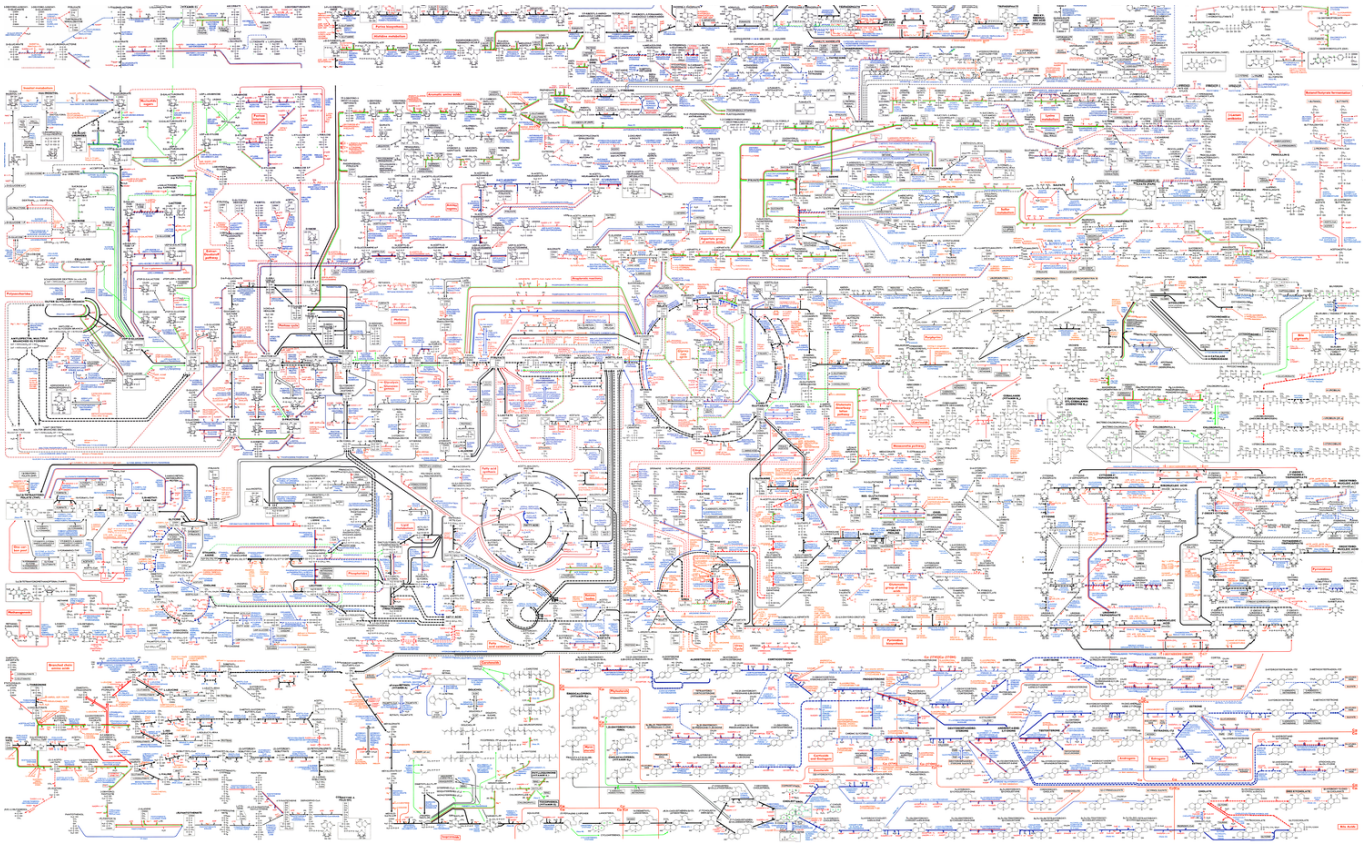


Figure 1.2. metabolic map from Roche (from Ref. 2).

With a little thought, anyone can see the merits and demerits of extreme element reductionism, which holds that life can be understood by dividing it into its components. Even when you are to pursue such a policy, such as identifying important proteins or genes, your insight and conscience will try to make the "merits" of the research larger and the "demerits" smaller.

Then, what about Figure 1.2? This figure covers the various reaction pathways in the body. In other words, it shows not only the components but also the relationships among them. Many researchers in life science are working hard day and night to identify the reaction pathways in living organisms of their interest, and of course, this is important for understanding life. Figure 1 is powerful enough to inspire a sense of awe at the cumulative efforts of so many researchers. However, if you can memorize all the reactions shown here, will you feel that you have understand the essence of life? In the first place, does understanding life as a system mean examining every part of it?

The circuit shown in Fig. 1.3A is a series connection of resistors, coils, capacitors, and power supplies. The behavior of the circuit cannot be understood only from its wiring. The change in the current flowing through the circuit from the moment the switch is turned on (transient response) may be as shown in Fig. 1.3B, or it may show the damped oscillation shown in Fig. 1.3C. Depending on the purpose of the circuit, you may not want to have the behavior shown in Fig. 1.3C. In order to avoid the damping oscillation, it is necessary to have a magnitude relationship of $R^2C > 4L$ between the resistance value R, the inductance L, and the capacitance C.

In the same way, in order for an organic relationship to be established between the components of the metabolic map shown in Fig. 1.2, and for the metabolic circuit to function or not behave in an undesirable manner, it is not enough that a metabolic relationship exists for each component, but some conditions must be satisfied.

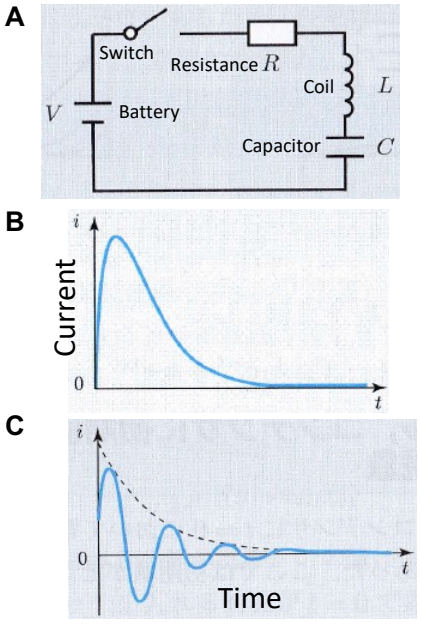


Figure 1.3. RLC series circuit (A) and its transient response (B, C).

Complex Systems Theory Brings Perspective to the Brain

This is where the need to understand life as a system comes in. Apart from elucidating the existence of the important components of life and the relationships among them, it is necessary to clarify what conditions must be met for organic relationships to be established and for necessary functions to be realized.

However, in order to achieve the principle of creativity, we need to go one step further. Of course, it is very meaningful to compare life as a system with human-made systems such as machines, and to ask how the relationships between the elements differ, and how the functions expressed by the elements differ, such as how a camera differs from a retina. Nevertheless, I still don't feel like I've captured the essence of life. The camera is made by humans, but the retina is made by itself. Isn't it possible to understand life as a system that creates its own structure and function?

This is the reason why I try to think about the brain from the perspective of complex systems theory of life, which is a theory of living systems based on complex systems theory. This theory, which has its origins in non-linear and non-equilibrium thermal and statistical mechanics, deals with how temporally and spatially ordered patterns, seemingly defying the natural tendency to “it is no use crying over spilt milk,” are naturally generated. This perspective gives us a somewhat better perspective on life and being “created” by the brain. In the next section, I will give a very brief introduction to the complex systems theory of life.

II. What is Complex Systems Theory of Life?

A cold bowl of miso soup or a chemical reaction that has passed for a long time after completion is said to be in equilibrium in the sense of thermal and statistical mechanics. The overall state of such an equilibrium system can be understood as a collection and addition of individual states (i.e., a linear sum). On the other hand, this is not the case for a hot bowl of miso soup where heat is rapidly escaping, or a chemical reaction of some kind. Such a system is called a nonlinear nonequilibrium system. A living system in which heat and energy flow exist through eating and excreting food is a typical nonlinear nonequilibrium system.

In non-linear non-equilibrium systems, ordered structures, regular spatial patterns and temporal rhythms are generated autonomously when conditions are right. Such phenomena are sometimes referred to as emergent or self-organizing phenomena in complex systems. One of the main arguments of this book is that emergent phenomena in the nervous system (sometimes including the environment) are behind certain creative functions found at various levels of the brain.

What is particularly important in complex systems such as living systems is the interaction between part and whole, micro and macro. From this perspective, I will devote much of the following to nonlinear oscillators and synchronization phenomena between them. This section briefly introduces the concepts that are necessary for the subsequent discussion of the brain.

For a more detailed discussion, please refer to the Appendix.

Pattern Generation through Bifurcation

In thermodynamics and statistical mechanics, the law of increasing entropy is a universal phenomenon that we all experience, "it is no use crying over spilt milk." Things go in the direction of disorder. However, there are some phenomena that seem to be exceptions at first glance. The best example is living organisms, which naturally produce spatiotemporal patterns such as beautiful patterns and rhythmic oscillations. Such patterns are called dissipative structures. The phenomena that give rise to such spatiotemporal patterns is called self-organization or emergent phenomena of complex systems. Such phenomena occur when chemical reactions and heat diffusion are far from equilibrium, i.e., in nonlinear nonequilibrium systems.

Time developments of things, such as chemical reactions, are described by differential equations (see Appendix for details). If the system is far from equilibrium, the differential equations cannot be approximated as linear equations, and nonlinearity becomes a problem. A point in a differential equation where the parameter does not change with time is called a singular point. Nonlinear differential equations can have multiple singular points, which is not the case in linear differential equations. Depending on the conditions, the equilibrium point becomes unstable and a transition to a stable state with a temporal or spatial structure can occur. Such a transition is called a bifurcation. The stable state is also called an attractor because it returns to that state even if there is some disturbance.

As an example, consider the case where the time development of the deviation X from the equilibrium point of a certain quantity is expressed as $\mu X - X^3$ (μ is a constant); when μ is -1, the time change of X is expressed as $-X - X^3 = -X(1 + X^2)$, and the singularity is $X = 0$, which is the equilibrium point. Even if there is a deviation of 0.1, for example, from the singularity, the change is -0.101, which cancels out the deviation, so the singularity point is stable. On the other hand, when μ becomes 1, the time development of X is expressed as $X(1 - X^2)$, and the singularity increases to three, 0 ± 1 (μ is referred to as a bifurcation parameter). If X shifts by 0.1 from singularity 0, the change is 0.099 in the direction of amplifying the shift, so we can say that singularity 0 is unstable. By the same argument, we can see that the newly created singularities ± 1 are stable. Also, when the deviation from singularity 0 is as small as 0.1, linear approximation of equation $X - X^3$ with X results in a change of 0.1, which is not much different from the case without linear approximation, On the other hand, when the deviation from singularity 0 is as large as 10, the change with the linear approximation is 10, while the change without approximation is $10 - 10^3 = -990$, which is a large difference. That is, linear approximation is not valid in the latter case. This example shows why nonlinearity becomes a problem for systems with far from equilibrium.

In nature, there are always turbulences such as thermal noise. As a system approaches a critical point where a bifurcation occurs, the stability of the system decreases, i.e., the effects of disturbances cannot be easily counteracted. At this moment, the fluctuations become larger when measuring a certain quantity of the system. The increase in fluctuations as a precursor to a bifurcation is called critical fluctuations.

Most complex systems are composed of a large number of elements and factors. However, the contribution of each element/factor is not equal. Some factors quickly cancel disturbances while others do not, and the overall state of the system can be better understood by focusing on the latter. The quantity that represents the behavior of the entire system is called the order parameter. The order parameter is not necessarily a microscopic parameter. Sometimes a macroscopic factor can take control of the entire system, including microscopic factors. The fact that such a case exists clearly shows that element reductionism is not a panacea.

Synchronizing Nonlinear Oscillators

There are various oscillatory phenomena in the natural world. In particular, living things have many such phenomena, such as sleep-wake cycles, heartbeats, walking rhythms, and cell division cycles. These are all nonlinear oscillations. The differential equation of nonlinear oscillation describing the time development of the quantity of interest X does not contain only the linear term of X, aX (a is a constant) but also a nonlinear term such as bX^2 (b is also a constant), and exhibits oscillation. Those that show oscillation are often called oscillators.

The spring pendulum, which you learn about in high school differential equations or physics classes, is a linear oscillator. The time development of displacement from a steady state X (acceleration in this case) is expressed, in the simplest case, as $-X$. In the absence of friction, if we assume an initial value of X_0, that is, if we start the oscillation at a value of X_0, the oscillation will continue with an amplitude of the absolute value of $|X_0|$. This is called the preservation of the initial value. If a disturbance is added and the amplitude is disturbed to $|X_0| + \Delta X$, it will still oscillate with that amplitude. If there is friction, the amplitude will gradually diminish and converge to $X = 0$.

Nonlinear oscillators exhibit different properties. In the following, I explain them using the van der Pol (VdP) oscillator as a typical nonlinear oscillator. The differential equation describing the VdP oscillator includes a friction term. However, unlike a linear oscillator, it contains a non-linear term for the displacement X from zero singularity rather than a constant. The nonlinear term provides a “negative friction” region around zero. Negative friction is hard to imagine, but X receives a force in the changing direction, in contrast to the case of positive friction where X receives a force in the opposite direction of change. Therefore, X does not converge to zero. Once X goes beyond the negative friction region, it receives positive friction, the velocity diminishes and is pulled back to the zero point. Such a nonlinear mechanism does not allow preservation of initial values. No matter where the nonlinear oscillation starts from, it will settle down to a fixed amplitude. Any perturbations disappear. Such constancy in oscillation underlies the stability and homeostasis of biological systems.

When multiple nonlinear oscillators interact with each other, they may spontaneously synchronize under certain conditions. This phenomenon is called mutual entrainment or entrainment. Entrainment can be observed among large numbers of nonlinear oscillators. There are a lot of examples of entrainment in biological systems. As for artificial things, the entrainment among many metronomes is spectacular, for example. Search on YouTube using the key word ”metronome synchronization.” Generally speaking, entrainment is more likely to occur when the fundamental frequencies of the oscillators are close or when the interaction is strong, while, for example, entrainment between oscillators with a certain frequency and the double frequency can occur. For the latter case, there seems to be much room for establishing theories. A system consisting of a large number of nonlinear oscillators may exhibit a phase transition from a state in which each oscillates independently. That is, they suddenly become synchronized when macroscopic parameters such as overall connectivity are properly adjusted.

Currently, there are researches that utilize the mutual entrainment between nonlinear oscillators for information processing or control. Entrainment is expected to be useful in achieving a globally good or consistent relationship between elements of the system of interest. Oscillation is accompanied by a phase (an angle when one cycle is represented by a circle) which is a neutrally stable value (the boundary between stable and unstable states). Namely, the circularity and neutrality of the phase are considered to be useful.

Biological Systems in which Parts and the Whole Interact

Biological systems are nonlinear nonequilibrium systems in which flows of energy, substances *etc.* are generated by eating and excreting. In such systems, fine structures or oscillatory activities emerge. In this sense, they can be regarded as typical complex systems. However, in contrast to other complex systems, the interaction between parts and the whole is the central issue in biological systems. Here, I will illustrate this problem by using an artificial muscle motor “the stream cell” and the information processing in true slime molds or the plasmodium of *Physarum polycephalum* as examples.

Actin and myosin, which are string-like molecules, are the main components of muscle. Muscle contraction is caused by the sliding of these molecules each other using the energy of ATPs (adenosine triphosphates). To fabricate a stream cell, first, actin and myosin molecules are obtained by decomposing muscles. The cell (Fig. 1.4 top) has a ring-shaped slit structure, and actin molecules are attached to its walls in aligned directions. When myosin molecules and ATPs are added into the slit, the liquid in the slit starts to move and rotate.

In the stream cell, the flow of liquid and the rate of ATP degradation, which reflects the reaction rate between actin and myosin, correlate with each other. When the flow is obstructed, the reaction rate slows down. In other words, there is an interaction between the macroscopic phenomenon of flow and the microscopic biochemical reactions. A micro-macro interaction exists in the sense that as the actin and myosin molecules react to each other, while the flow enhances the coherence between each reaction of the molecules. This demonstration suggests that parts and the whole have to be in consistent relationships for biological systems to survive as a whole, in contrast to ordinary self-organizing systems in which macroscopic spatiotemporal patterns emerge as a mere consequence of microscopic interactions between elements.

The importance of the consistent relationship between parts and the whole is even more apparent in the slime mold or eubacteria (Fig. 1.4, bottom). It is an amoeba-like multinucleated unicellular organism that has no nervous system or other organs for information processing or control. However, they approach their food and escape from unpleasant things as a whole. It has a mesh-like structure called *plasmodium*. The outer part of *plasmodium* is called the exoplasm or gel, and the inner part is called the endoplasm or sol. In the slime mold, various things such as local thickness and calcium concentration oscillate. The interaction of these local oscillations allows it to process information so that it behaves in an consistent manner as a whole. the molds to proceed information to achieve consistent behavior as a whole. In other words, a true slime mold can be regarded as a system of collective nonlinear oscillators.

When a true slime mold encounters a food, the frequency at the site facing the food increases, whereas the frequency decreases when facing an aversive stimulus. The influence of the increased frequency spreads to other part of the body, causing entrainment. As a result, the frequencies in other parts of the body also increase, while a phase gradient, or phase delay, occurs. The eubacteria as a whole moves in the direction of the phase gradient. If one part is facing a certain food, but the opposite part is encountering a more preferred food, that part will retract. In other words, even though the slime mold does not have a nervous system, each part behaves as if it "knows" where the best food is in the whole body. The information that each part needs to know its place in the whole is called "positional information". Although I will not go into it further in this book, the development of theories on the "part-whole" problem in complex systems is essential now that research on regenerative medicine is flourishing.

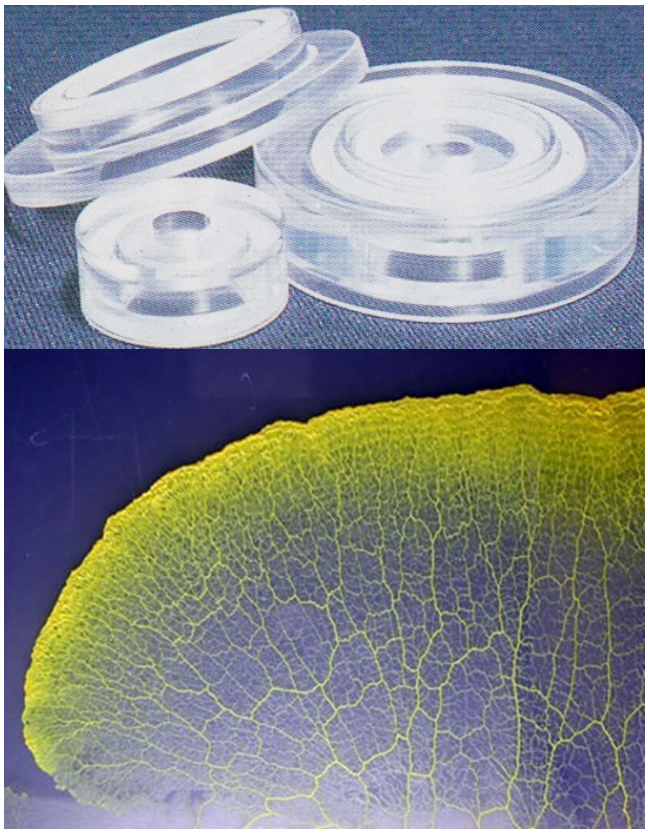


Figure 1.4. Stream cell (from Ref. 3) and true slime mold (from Ref. 4).

Based on the concepts and items of the theory of complex biological systems briefly introduced above, the discussion in Chapters 2, 3, and 4 will be as follows.

In Chapter 2, I describe neurons as a fundamental constituent in the nervous system. In particular, I will emphasize their properties as nonlinear oscillators. In other words, Namely, I show that spiking activities of a neuron is understood as nonlinear oscillation generated when a certain threshold is exceeded.

As a representative example of information processing in the cortex, Chapter 3 focuses on the primary visual cortex (area V1), which is the main entry point for visual input in the cortex. I will discuss not only the classical results there, but also how the nature of nonlinear oscillations can be involved in the information processing in V1, in relation to the part-whole problem of figure-ground separation.

On the other hand, in Chapter 4, I talk about the prefrontal cortex which is anatomically far from perceptual inputs and motor outputs. The prefrontal cortex is considered to be involved in creativity and I would like to introduce the results of our physiological research there. Particularly, the generation of specific behavioral goals for achieving a relatively abstract purpose corresponds to emergent phenomena in the brain such as bifurcation, synchronization and critical fluctuation found in the neuronal activities.

References

1) http://www.mhlw.go.jp/kokoro/speciality/data.html

2) http://www.expasy.ch/cgi-bin/show_thumbnails.pl

3) Shimizu H. *Biological Systems and Information*. Japan Broadcasting Corporation Press, Tokyo (1987) in Japanese

4) https://matome.naver.jp/odai/2147288592419813401/2147289215925741903

Chapter 2. Neuron as a Nonlinear Unit

This chapter provides an overview of the characteristics of neurons, the major components of the nervous system. Section I illustrates the basic structure and behavior of a neuron. In particular, it explains the membrane potential and how it responds to signals from other cells in an impulse-like manner. Specifically, we explain the Hodgkin-Huxley model, which describes the electrical behavior of a neuron. Section II introduces a nonlinear oscillator model, called the KYS oscillator, which qualitatively reproduces the neuronal behavior.

The Hodgkin-Huxley model is an excellent model that accurately reproduces the behavior of a neuron. However, due to the complexity of the equations, it is difficult to understand the nature of the impulse-like response that a neuron exhibits once a threshold is exceeded, from the perspective of nonlinear dynamics. The KYS oscillator (see Appendix A2 for details), derived from the Van der Pol oscillator, a typical nonlinear oscillator, can help us understand the nature of the neuronal response.

As you know, there are many ways to model the response of a neuron. In artificial neural networks, simple units that represent a "1" when the sum of the inputs exceeds a threshold and a "0" otherwise are widely used. However, to understand the neural system as a complex system, it would be desirable for the unit to equip the properties of a nonlinear oscillator. In this sense, the KYS oscillator is an excellent theoretical model.

I. Neuron and its Impulse

The Basic Structure of a Neuron

A neuron has a structure consisting of dendrites, soma (cell body), and axon (Fig. 2.1). The terminal of axion is “connected” to the dendrites or soma of other neuron, or effectors such as muscles. Although they are connected, they are not attached to each other, but rather have a narrow gap between them. This connection is called a synapse. When excitation (more on excitation later) is conducted through the axon, neurotransmitters are released from the terminal. By binding the transmitters to the receptors, the postsynaptic cell is excited or inhibited. This allows the presynaptic cell to transmit signals to the postsynaptic cell.

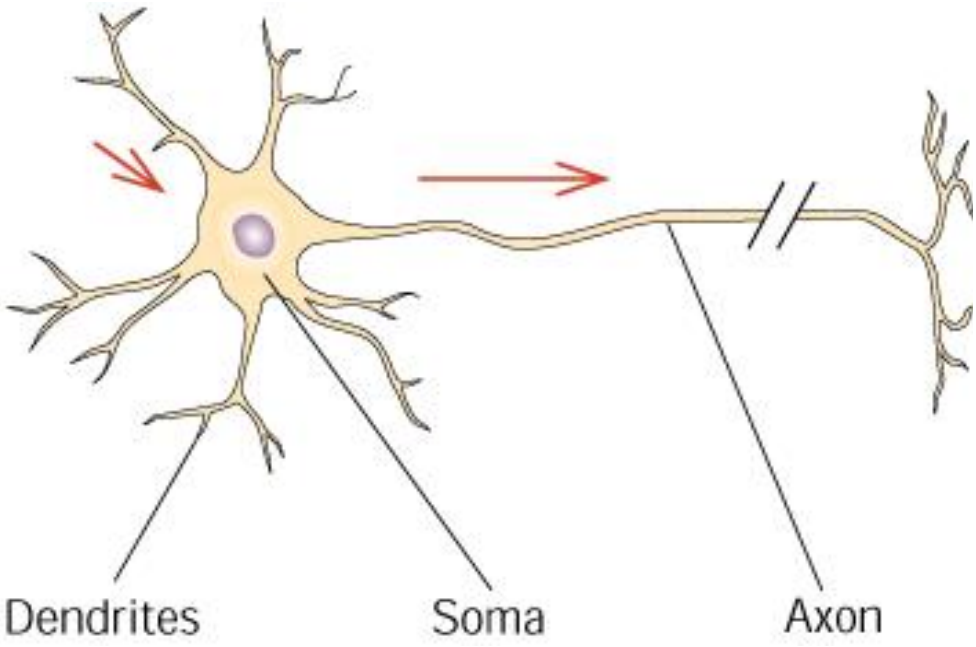


Figure 2.1. Schematic diagram of the basic structure of a neuron (from Ref. 1).

It is easy to describe a neuron as described above based on our current knowledges. However, in the past, when there was no crucial techniques, it was not easy to obtain such an image.

No one would argue that the Italian anatomist Golge and the Spaniard Ramon y Cajal made major contributions to the elucidation of the structure of neurons (Fig. 2.2). For their achievements, the two were awarded the Novel Prize in Physiology together in 1906. However, the two anatomists were in direct conflict over the structure of neurons, and even gave opposite theories in their Nobel Prize acceptance speeches.

Novel scientific discoveries are often brought about from new experiment techniques. A breakthrough in elucidating the structure of the nervous system was the development of a staining method, later called Golgi staining, in which nerve cells were stained with chromite silver. Staining a portion of the nervous system with this method revealed the structure of dendrites, somas and axons. However, at the time, it was not possible to observe the fine structure of synapses, so Golgi proposed a "reticular" theory, in which axons of different nerve cells are merged at the synapses (like slime mold?).

Cajal was also using Golgi staining, but he had a “neuron” theory that cells are not merged but separated. Their controversy was not resolved until many years after their deaths, when the synaptic gap was discovered by the invention of electron microscope.

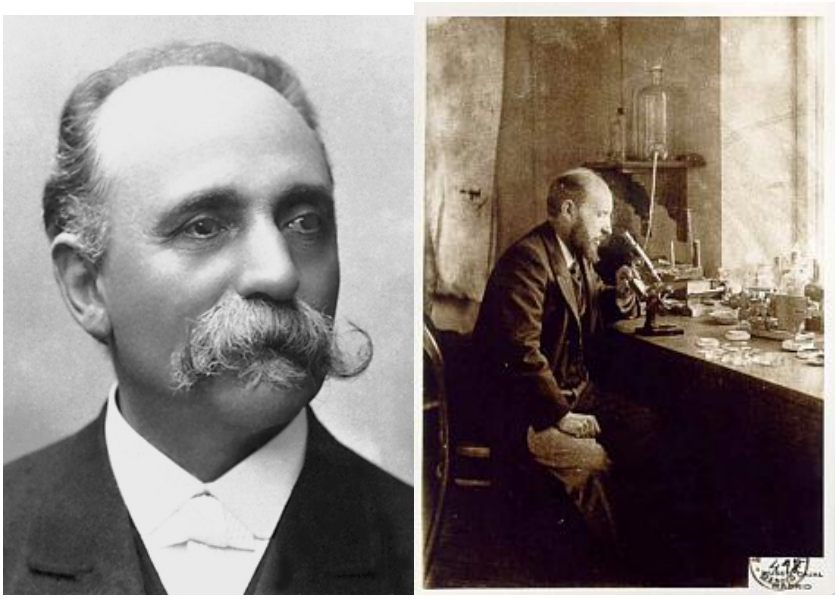


Figure 2.2. Golgi (left) and Ramon y Cajal (right).

Both the two men from the countries of art left beautiful anatomical sketches. Nevertheless, one’s theory was revealed to be correct and the other’s was not. What brought them different theories? This is a fascinating question for me who often think about the sensibilities of scientists.

Neuronal Activity is a Bioelectric Phenomenon[2)]

At the end of the 18th century, the Italians Galvani and Volta discovered the existence of bioelectrical phenomena through their experiments on frog muscles. Since then, for about a hundred years, the function and mechanism of nerve cells have been studied mainly as bioelectrical phenomena in frog muscles with nerves (nerve-muscle preparation).

Applying an electric current to a nerve-muscle preparation causes the muscle to contract. This suggests that the electric current induces some kind of change in the nerve, which is then transmitted to the muscle, resulting in the muscle contraction. This change is called excitation, and clarifying its properties and substances was a major challenge at the time.

In 19th century, they do not have instruments to measure minute currents or voltages, and had no ways to evaluate the effects of electrical stimulation except by using muscle contractions. Even in such a situation, devices that generate various patterns of electrical stimulation were invented by Helmholtz of Germany and other scientists, which elucidated the nature of excitation step by step.

First of all, it became known that a minimum intensity of electric current, that is, a threshold, is required to induce excitation. In addition, it has become clear that electrical stimulation immediately after the onset of excitation does not produce excitation. This period of non-response to stimulation is called the refractory period. On the other hand, at the time, there was a tendency to think that the excitation transmitted through nerve fibers was transmitted at the speed of light, like an electric current through a metal conductor, but it turned out that it was actually much slower than that. In addition, it was revealed that the excitation is only in two states: either it occurs or it does not occur (the so-called all-or-nothing law).

It was known by the end of the nineteenth century that the electric potential at the cut surface of a nerve (i.e. inside) is lower than that at the nerve surface. By developing a highly sensitive ammeter, the German Du Bois-Reymond discovered that the difference between these potentials becomes smaller when excitation is transmitted (this is called an action potential). With this, he is credited as the discoverer of the action potential in neurons. However, it was not until the development of more sophisticated measurement devices that the details of action potentials were revealed, as described below.

Based on these discoveries, Bernstein (also from Germany) proposed a hypothesis in the early 20th century, later called Bernstein's membrane hypothesis. He hypothesized that the membranes of neurons have the property of allowing certain types of ions with a certain charge to pass through freely, but not others (this property is called selective permeability). Specifically, it was assumed that the potassium ion concentration inside the cell is higher than that outside due to selective permeability when the cell is not excited, which causes a potential difference between the inside and outside of the cell membrane (see BOX). This potential is called the resting membrane potential. Bernstein thought that the resting membrane potential is the equilibrium potential of potassium ions. He also hypothesized that excitation, or the generation of an action potential, is caused by a transient loss of this resting membrane potential, that is, by a temporary loss of selective permeability.

Bernstein’s membrane hypothesis was no correct eventually. However, through testing his hypothesis, the nature of action potential has been elucidated. A good hypothesis has great significance even if it is not correct.

BOX *Voltage Difference and Selective Permeability*

Not all substances can move freely in and out of the lipid bilayer of the cell membrane. In particular, the movement of charged particles, i.e., ions requires special passages in the cell membrane. Let's consider a case where there is a special passage for a certain ion (selective permeability).

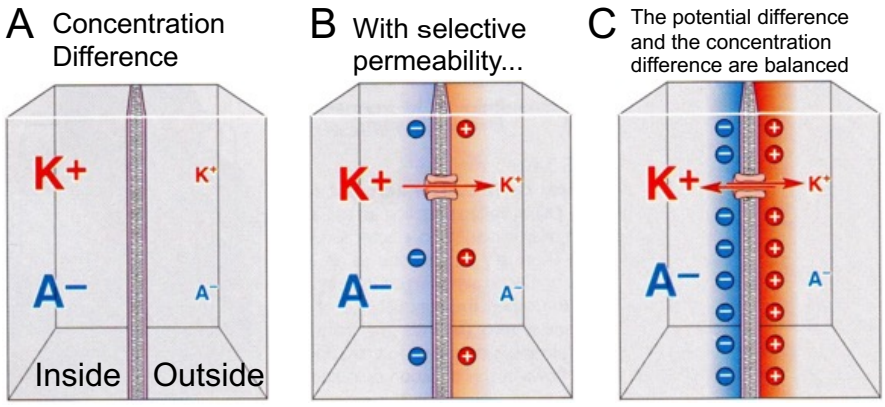


Figure. Schematic diagram of the mechanism by which selective permeability produces a potential difference. Modified from Ref. 4.

Consider the case where a positively charged ion (cation, in this case the potassium ion K^+) and its negatively charged counterpart (anion, in this case the hypothetical ion A^-) exist on the inside and outside of a membrane that mimics a cell membrane, and their concentrations are higher on the inside of the membrane (Figure A). Next, consider the case where a K^+-only pathway is added to this membrane (Figure B). Naturally, K^+ will flow outward according to the concentration gradient. The positive and negative ions with different concentrations inside and outside are electrically attracted to each other across the thin membrane. As K^+ flows out, the number of A^- ions lined up inside the membrane increases, and as a result, the potential inside the membrane becomes negative. This potential difference in turn pulls the K^+ on the outside of the membrane to the inside, and the K+ traffic apparently stops when the outflow due to the concentration difference is balanced by the inflow due to the potential difference, while the K^+ concentration on the inside remains higher (Figure C). The potential difference at this point is called the equilibrium potential of K^+. The equilibrium potential occurs when there is a difference in the concentration of an ion between the inside and outside of the membrane and when there is selective permeability of that ion.

There are multiple types of cations and anions inside and outside cells. However as shown in the table, there are significant differences in concentration inside and outside the cells depending on the ions, while the concentration pattern is similar across species. The equation for obtaining the equilibrium potential is called the Nernst equation.

$$E_{ion} = \frac{RT}{FZ} \ln \frac{[ion]_o}{[ion]_i}$$

E_{ion} is the equilibrium potential of the ion, R is the gas constant of 8.31, T is the absolute temperature (273 + ° C), which is 293 at room temperature, F is the Faraday constant 96,500, Z is the charge of the ion, and ln is the natural logarithm (that is, $\log_e$). $[ion]_o$ is the outer concentration of the ion, and $[ion]_i$ is the inner concentration (all units omitted). Keep in mind that the equilibrium potential is determined for each ion.

Now, when the equilibrium potential is calculated for the K^+ concentration in the table using the Nernst equation, it is -75 mV. This is a value very close to the resting membrane potential. The equilibrium potentials of other ions are far from the resting membrane potentials. This is the reason why at the beginning of the 20th century, the true nature of the resting membrane potential was thought to be the equilibrium potential of potassium ions.

The main cause of the difference in ion concentration inside and outside the cell is the Na^+ -K^+ pump, which is a type of active transport that uses the energy of ATP. Although details are omitted, the decrease in intracellular K^+ concentration is slight even when the state shown in Fig. A is changed to the state shown in Fig. C, whereas the equilibrium potential varies significantly from 0 mV to -75 mV. By actively generating an ion concentration gradient using energy, the basis for obtaining a large electrical signal with a small change in concentration is formed.

Table. Intracellular and extracellular concentrations of major ions in the squid giant nerve. mM is millimole.

Ions	Inside	Outside
Potassium (K^+)	400 mM	20 mM
Sodium (Na^+)	50 mM	440 mM
Chloride (Cl^-)	51 mM	560 mM
Calcium (Ca^{2+})	0.4 mM	10 mM
Magnesium (Mg^{2+})	10 mM	54 mM

Voltage Clamp by Hodgkin and Huxley

Advances in technology and science are inseparable. This is also true in the elucidation of the nature of neuronal activity.

At the beginning of the 20th century, it was expected that the resting potential of neuronal membrane was equal to the potassium equilibrium potential due to the selective permeability of potassium ions, whereas the action potential was the temporary vanishment of the resting potential due to transient loss of the selective permeability.

Afterwards, however, the development of oscilloscopes with the invention of the cathode-ray tube, and advances in amplifiers with the invention of the vacuum tube, etc., made it possible to measure the small voltages of action potentials. With the advancement of these technologies, it became clear that the action potential is a pulse-like increase in voltage of short duration, with an overshoot above zero volts.

These new findings led to a new hypothesis that the charge of action potential is carried by sodium ions. This idea was supported by findings such that when the extracellular sodium ions were replaced by their radioisotopes, the isotopes were transported into the cell.

To confirm this hypothesis, it was needed to accurately measure the permeability of the membranes for sodium and potassium ions, separately. The permeability is defined as the conductance, which is the reciprocal of the resistance R, where $g = 1/R$. At that time, it was almost known that the conductance of potassium was higher in the resting state and that of sodium was higher in the state of excitation. However, if the membrane potential E changes in a short time, it is not possible to measure each conductance accurately. In other words, since the membrane potential E, resistance R, and current I are expressed by the equation $E = RI$, which can be transformed to $I = gE$, the change in conductance g cannot be evaluated correctly when the membrane potential varies.

The British physiologists Hodgkin and Huxley used a method called valtage-clump using negative feedback control, which was the latest technology at the time (Figure 2.3). In this method, which is still used today, the membrane potential is held constant and the current flowing through the membrane is measured. Feedback control is used to detect the difference between the potential set by the experimenter and the membrane potential, and inject current to cancel it out. If the membrane potential is kept constant (clumped) and the membrane current is measured, the time change of the conductance can be estimated, because the current and conductance are proportional. Giant squid axons with a diameter of one millimeter were also used so that a pair metal electrodes were inserted to obtain spatially-homogeneous membrane potentials.

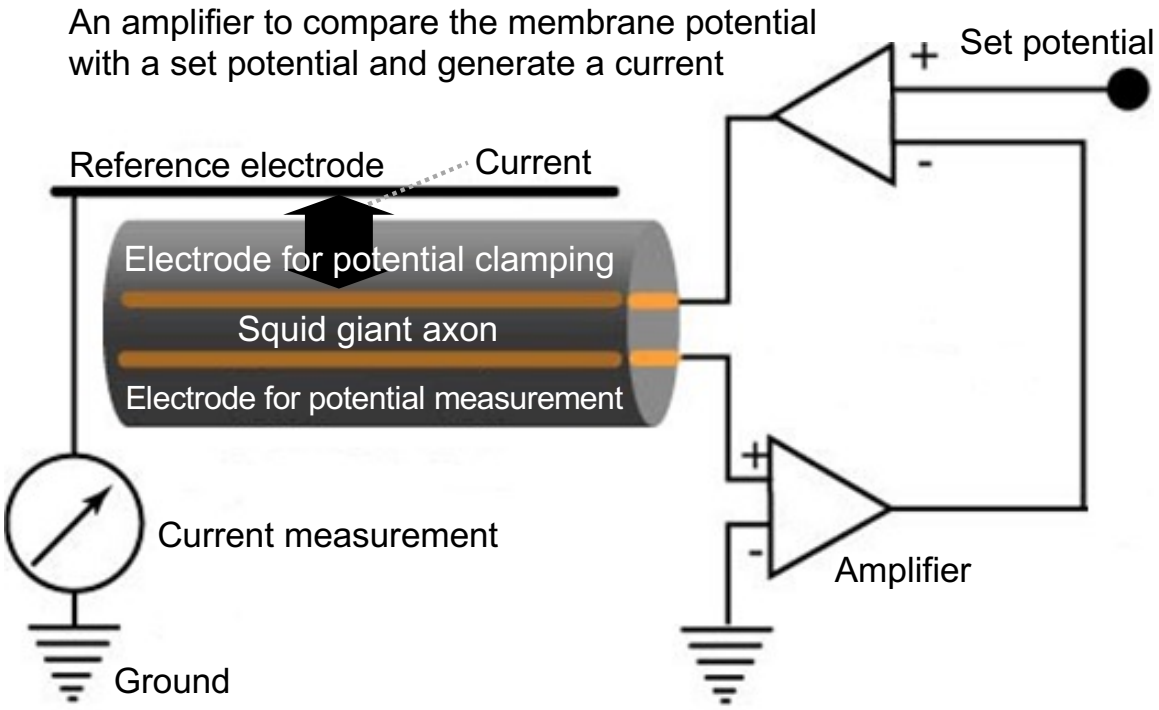


Figure 2.3. Voltage clump method. A current flows between the measurement and reference electrodes.

Voltage Clump Evaluates the Contributions of Different Ions to Action Potential Separately

What is the substrates responsible for the action potential of a neuron? To answer this question, it is necessary to isolate the potassium and sodium currents separately. Hodgkin and Huxley used the squid giant axon and the voltage clamp method to identify them.

First, the giant axon was immersed in normal seawater, and the membrane potential was clumped at the voltage that can induces a spiking activity under the normal condition (Fig. 2.4, top). Then, they observed a transient inward current followed by an outward current (Fig. 2.4 bottom, thick line). Next, the solution in which the giant axon was immersed was replaced by seawater without sodium ions. In this case, no sodium current is generated. When the membrane potential was clumped at the same voltage as above, only outward current was observed (Fig. 2.4 bottom, thin line). Then, the sodium current (Fig. 2.4 bottom, broken line) was estimated by subtracting the current measured under the without-sodium condition from the normal one. Since the current is proportional to the conductance when the membrane potential is clumped, they were able to estimate the time change of the conductance of sodium and potassium, respectively.

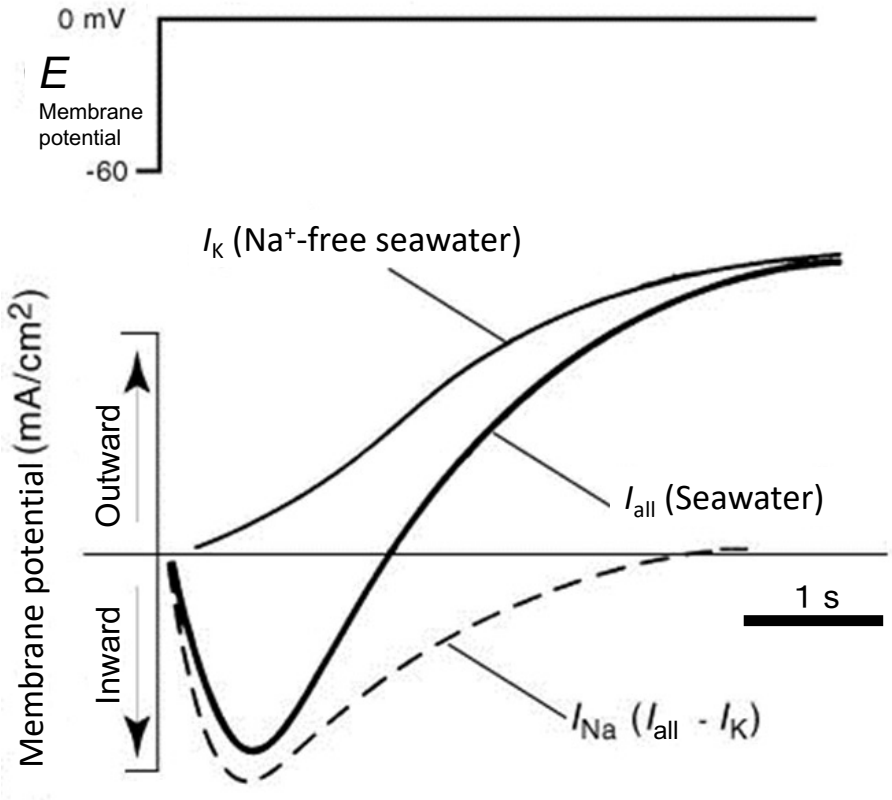


Figure 2.4. Separation of potassium current I_K and sodium current I_{Na} by voltage clump method.

BOX *The Squid Tank is a Product of Neurophysiology*

Some Japanese restaurants or sushi bars have squid tanks, and we can enjoy fresh squid sashimi. But, few people know that the it is a product of neurophysiology.

The choice of experimental materials is an important issue for scientists. The giant axon of the squid has contributed greatly to the development of neurophysiology. If it were not 1mm thick, Hodgkin and Huxley would not have been able to thread two electrodes through the axon. For such experiments, fresh squids are necessary. However, unlike fish, it used to be very difficult to keep squids in a tank.

One of the achievements of Dr. Gen Matsumoto, who regrettably passed away in 2003 at the age of 62, was the establishment of a method for breeding squid for neurophysiological experiments. Dr. Matsumoto, who switched from physics to neuroscience, was a sturdy man who believed that the most basic and fundamental research should be done first, and he poured his heart and soul into the establishment of the squid breeding method to the extent that he was said to be "crazy."

The problem was the accumulation of ammonia in the tank. Squid, unlike fish, are extremely sensitive to ammonia. After much effort, Dr. Matsumoto and his colleagues eventually solved the problem by building a purification tank using bacteria that decompose ammonia.

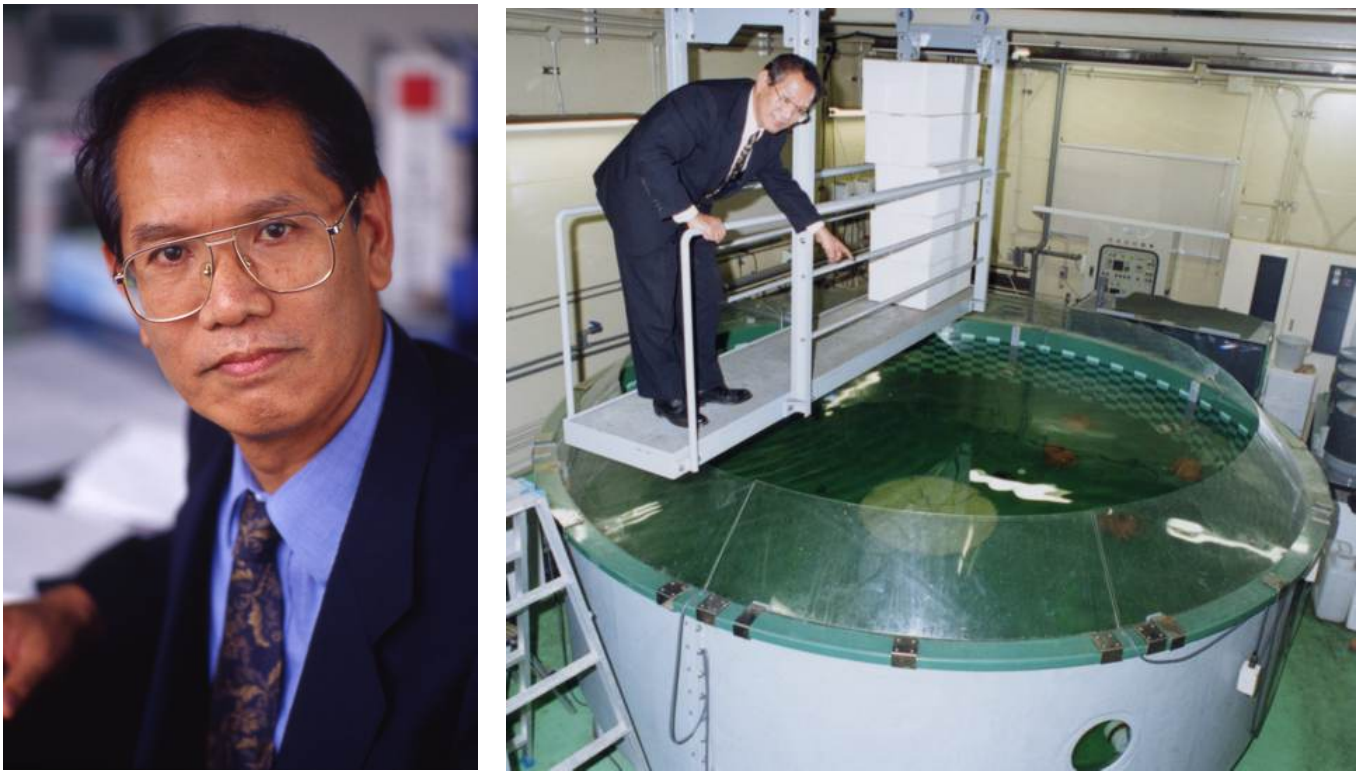


Figure. Dr. Gen Matsumoto and his squid tank (from Ref. 6).

Hodgkin-Huxley Equations

Following the success of isolation of sodium and potassium conductance in the voltage-clump experiments on squid giant axons (Fig. 2.5), Hodgkin and Huxley developed a theoretical model of neurons based on the ion-channel hypothesis that minute channels specific to each ion exist on the neuronal membrane. This model is known as the Hodgkin-Huxley equations (H-H equations). Details are given in BOX.

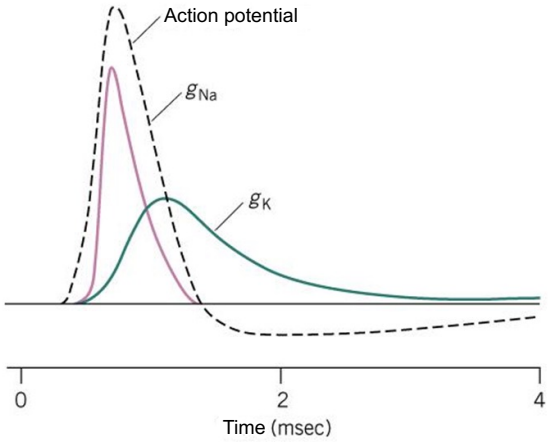


Figure 2.5. Time change of action potential, sodium conductance g_{Na}, and potassium conductance g_K.

The H-H equations have following characteristics: (1) sodium ions pass through sodium channels and potassium ions pass through sodium channels; (2) each type of ion channel has different mechanisms of opening and closing, and the states of the factors that regulate opening and closing (gate factors) are dependent on the membrane potential (voltage-dependent channel), etc. The computer simulations of the H-H equations reproduce the electrical activity of a neuron very well.

The existence of ion channels are so fundamental to life science that it is even mentioned in high school textbooks. However, Hodgkin and Huxley's hypothesis of ion channels was quite prescient at the time of 1952 when there was no evidence at all. They were awarded the Nobel prize in Physiology in1963 for their works and subsequent developments of related researches.

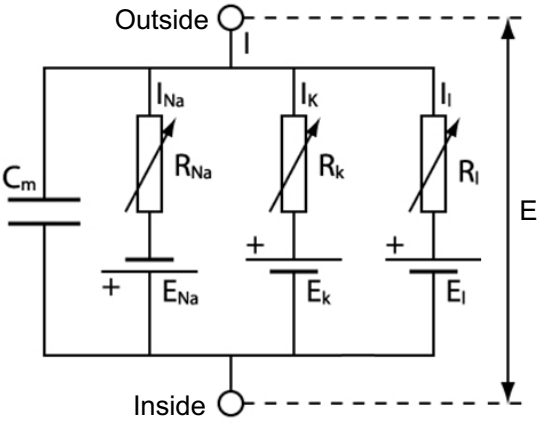


Figure 2.6. Equivalent circuit of the Hodgkin-Huxley equations. I is the total current. I_{Na}, I_K, and I_l are sodium, potassium, and leakage currents, respectively. E is the membrane potential. E_{Na}, E_K, and E_l are sodium, potassium, and leakage equilibrium potentials, respectively. R_{Na}, R_K, R_l are sodium, potassium, and leakage resistance, respectively (inverse of conductance g); Cm is the capacitance of the membrane.

The H-H equations are also highly scalable. If a new type of ion channel is discovered, you can add the equations for that channel based on the experimental results (Fig. 2.7A). You can also build a theoretical model that takes into account the shape of a neuron including dendrites (Fig. 2.7B). Modeling of a neuron based on the H-H equations enable to understand physiological findings in terms of electrical circuits, and its usefulness will never be lost in the future.

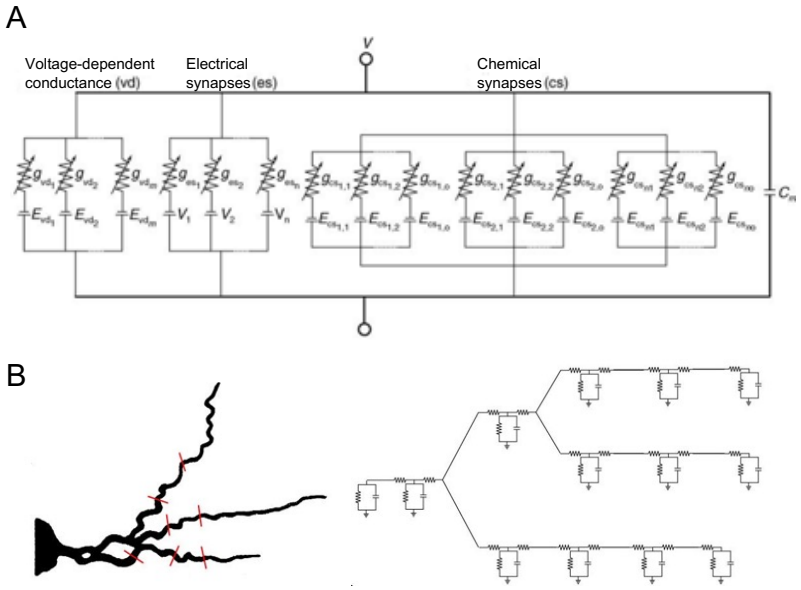


Figure 2.7. The H-H equations are scalable. It can be used to model additional channels (A) and structures (B) as well.

BOX *The Hodgkin-Huxley Equation as a Fourth-Order Nonlinear Differential Equation*

In the Hodgkin-Huxley equation (H-H equation), under the ion channel hypothesis, each channel has an independent gating mechanism, and the state of each factor related to the opening and closing of the gate (gate factor) is represented by three differential equations in a form dependent on the membrane potential, as shown below:

$$\frac{dm}{dt} = a_m(1-m) - b_m m,$$

$$\frac{dh}{dt} = a_h(1-h) - b_h h,$$

$$\frac{dn}{dt} = a_n(1-n) - b_n n,$$

where m, h, and n represent the probability that the gate factors involved in the activation of sodium channels, inactivation of sodium channels, and activation of potassium channels, respectively, are in the state where the channels are open. a_m, b_m, a_h, b_h, a_n, and b_n are each functions of the membrane potential E as follows:

$a_m = 0.1(25\text{-}E)/[\exp(2.5\text{-}E/10)\text{-}1]$,
$b_m = 4\exp(\text{-}E/18)$,
$a_h = 0.07\exp(\text{-}E/20)$,
$b_h = 1/[\exp(3\text{-}E/10)+1]$,
$a_n = 0.01(10\text{-}E)/[\exp(1\text{-}E/10)\text{-}1]$,
$b_n = 0.125\exp(\text{-}E/80)$.

Using these and the sodium and potassium conductances when all channels in the unit area are open, $\overline{g_{Na}}$, $\overline{g_K}$ (constants), each conductances can be expressed as:

$$g_{Na} = \overline{g_{Na}} m^3 h,$$
$$g_K = \overline{g_K} n^4.$$

These equations were determined to be m-cubed and n-quadratic, respectively, based on experimental results.

The current flowing through the channel, which is independent of the membrane potential, is collectively treated as leak current I_l and expressed as follows:

$$I_l = g_l(E - E_l),$$

where g_l is the conductance of the leak current and E_l is its equilibrium potential.

Since the cell membrane, which is a lipid bilayer, is an insulator, it forms a kind of capacitor. Therefore, when the membrane potential is changing, a capacitive current I_c also flows, as shown in the following equation (the voltage clump method was invented to cancel this current),

$$I_c = C_m \frac{dE}{dt}$$

where C_m is the membrane capacity per unit area.

Finally, each current is summed to obtain the total current I flowing in and out of the cell. If you use the inverse of conductance instead of resistance R, the equation becomes an addition of currents and can be expressed simply as follows,

$$I = C_m \frac{dE}{dt} + g_{Na}(E - E_{Na}) + g_K(E - E_K) + g_l(E - E_l),$$

where E_{Na}, E_K are the equilibrium potentials of sodium and potassium, respectively.

The Hodgkin-Huxley equation consists of four nonlinear differential equations: three for the gate factors and one for the sum of the currents. In this method, if a new ion channel is found, the equations for its opening and closing and current can be added in the same way.

But isn't it complicated? We now know what the equation models, but it is not so easy to imagine the behavior of the membrane potential without such prior knowledge, just by looking at the equations.

II. Design of a Simple Neuronal Model

Characteristics of Neuronal Activity

A model explains and reproduces a phenomenon by removing unnecessary things and extracting the essential characteristics of the phenomenon.

The above mentioned H-H equation is a physiological model that aims to map chemical phenomena, such as the composition of the concentration of each ion inside and outside the cell and the binding of neurotransmitters to ion channels, to electrical phenomena, such membrane potentials. This equation still provides a basic framework for understanding the behavior of various types of neurons.

However, the H-H equation is a physiological model, therefore not suitable for understanding the mathematical and mechanical aspects of the phenomenon. In particular, it is not useful to a mathematical or mechanical understanding of the following qualitative properties:

1) Neural excitation requires a minimum magnitude of input, or threshold.
2) When the input exceeds the threshold value, it oscillates with an impulse-like waveform (also called spike or firing).
3) As the magnitude of the input increases, the frequency of the oscillation increases monotonically.

It is never easy to imagine the above 1) to 3) properties just by looking at the H-H equation.

Seeking a Simple Neuronal Model

Various neuronal models have been proposed that are simpler than the H-H equation, a fourth-order nonlinear differential equation, but reproduce the above three properties. Among them, the KYS oscillator, named after Prof. Shinichi Kimura of Tokyo University of Science, Prof. Masafumi Yano of Tohoku University and Hiroshi Shimizu of the University of Tokyo, is a simple and beautiful formula reproducing them qualitatively[5), 6)].

The KYS oscillator is expressed as a second order differential equation as follow:

$d^2x/dt^2 = -f(x)dx/dt - g(x)x + D$,
$f(x) = a_1x^2 + b_1x + c_1$,
$g(x) = a_2x^2 + b_2x + c_2$.

In the following, I explain it using the metaphor of a spring pendulum, or a ball rolling around in a bowl. x can be likened to the horizontal position of a pendulum or a ball in a bowl, corresponding to membrane potential. D is the input. $f(x)$ and $g(x)$ are the expressions for friction and stiffness in terms of a spring, respectively. $g(x)x - D$ means the gradient of the bowl at position x, and $G(x)$ is its integral over x (with the integration constant of 0), representing the shape of the bowl. Using this "ball in a bowl" metaphor, I explain the behavior of the KYS oscillator without using equations in the next two pages.

Modeling of Neuronal Activity – Negative Friction and the Shape the "Bowl" are Key

In this section, I will give an intuitive overview of how the KYS oscillator qualitatively reproduces the electrical activity of a neuron.

When a ball is placed in a bowl, it rolls downward by gravity, and eventually stops at the bottom due to friction (Fig. 2.8A). As the horizontal position of the bowl represents the membrane potential, the dumped oscillating voltage can be illustrated as in Fig. 2.8A.

Next, a negative friction region is set around the bottom (Fig. 2.8B). The term negative friction may be unfamiliar, but it means negative resistance. Positive frictional force acts on the ball in the opposite direction to the motion and stops it, while negative friction force drive it in the same direction. Consequently, the ball keeps on oscillating without falling to the bottom.

Let the bowl have two bottoms and set a negative friction region around one of them, the ball falls down to the other bottom (Fig.2.8C). When the bowl shape is changed by the input, as in Fig. 2.8D, one bottom disappears. However, because of the negative friction region around the remaining bottom, the ball cannot fall there. With this ball behavior, the KYS oscillator reproduced the neuron's property of starting oscillation above a threshold value.

Let's look at Fig. 2.8D a little more carefully. After the ball passes through the negative friction region, it reverses and is attracted to the bottom. However, the positive and negative sides of the area have different forces of attraction.

On the positive side, the bowl's slope is so steep that the ball is strongly pulled and quickly returns toward the bottom. On the other hand, on a gentle slope on the negative side, the ball is not pulled back strongly, resulting in a slow return to the bottom. This is the mechanism by which the KYS oscillator oscillates with an impulse-like waveform.

As the input increases, the bowl shape is further distorted. In particular, the slope around the former bottom in Fig. 2.8C becomes steeper (Fig. 2.8E). This causes the ball to immediately return to the bottom of the remaining single bottom. This is how the KYS oscillator exhibits various frequencies depending on the input.

You may skip following discussions, but we will design the KYS oscillator to behave in a neuron-like manner. That is, its parameters, a_1, b_1, c_1, a_2, b_2, and c_2 are determined by considering their respective meanings.

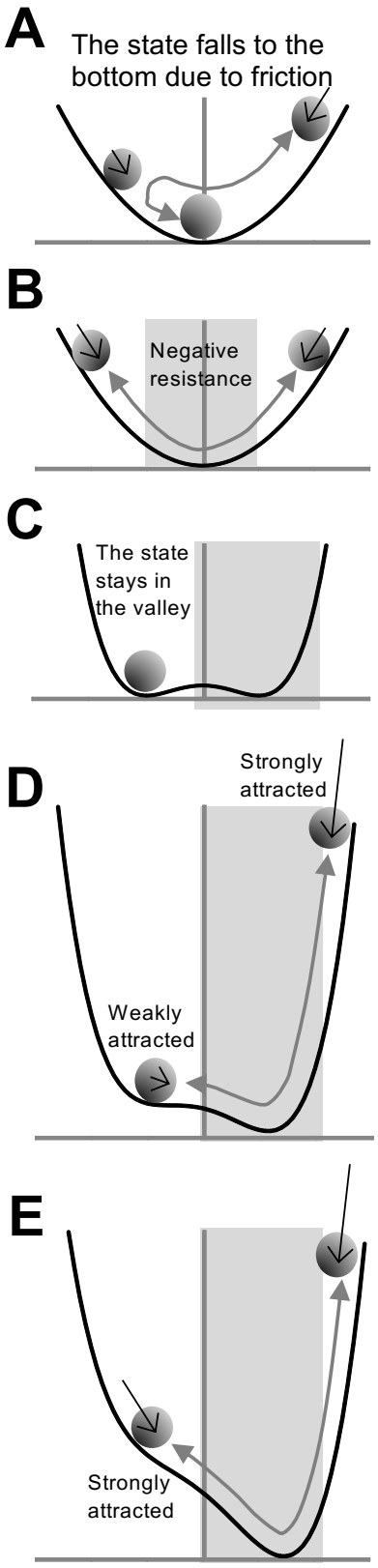


Fig. 2.8. The KYS oscillator qualitatively describes the behavior of a neuron. Shaded areas, negative friction areas.

The BvP Oscillator – Asymmetric Negative Friction Region

Consider the following three oscillators (see Appendix A2 for details): a linear oscillator without friction (corresponding to the KYS oscillator with zero constants other than the positive constant c_2), a linear oscillator with friction (with zero constants other than the positive constants c_1 and c_2), and the Van der Pol oscillator (VdP oscillator. For example, $a_1 = 1$, $c_1 = -25$, $c_2 = 10$, and the others are zero. That is, $g(x)$ is a positive constant, while $f(x) = x^2 - 25$.

Consider a potential. Let $G(x)$ be the integral of g(x)x – D with respect to x with the integration constant as zero. In the three examples raised here, $g(x)$ is a constant c_2, so $G(x) = c_2x^2/2$, which is quadratic about x. $G(x)$ is bawl-shaped with $x = 0$ at the bottom, as shown in Fig. 2.9A. The metaphor of a ball in the bowl will help you understand the following discussion. The ball moves toward the bottom according to the slope of the bowl.

In the case of a frictionless linear oscillator (Fig. 2.9A), the ball starts to fall at a certain height, it passes through the bottom at $x = 0$, reverses at the same height on the opposite side, and oscillates endlessly. In contrast, with a friction, the ball eventually stops at the bottom (Fig. 2.9B).

The VdP oscillator (Fig. 2.9C) has a negative friction region of $x = \pm 5$, so the ball cannot stop at the bottom, goes out of the negative friction region and returns in the positive friction region, resulting in the continuing oscillation. The negative friction region is set symmetrically with respect to the bottom of the bowl, resulting in a vertically symmetric waveform.

Now, let's consider the case where, for example, $a_1 = 1$, $b_1 = -10$, $c_1 = -1$, $c_2 = 10$ and the others are zero, i.e., $f(x) = x^2 - 10x - 1$ (Fig. 2.9D). This corresponds (not strictly) to a negative friction region of the VdP oscillator translating 5 or less in the positive direction. In other word, the negative friction region is asymmetric with respect to the bottom. This is called the simplified Bonhoeffer-van der Pol (BvP) oscillator.

You can recognize an asymmetric waveform in the vertical direction, that is, a slightly spike-like oscillation, which is qualitatively explained as follows. The time spent in the negative friction region is negligible, which means the asymmetric waveform is determined by the times spent in the both positive friction regions above and below the negative friction region. The time spent depends on the strength of the force pulling the ball to the bottom, i.e., the slope gradient. The slope in the positive side is steeper, and the ball is pulled strongly by the steeper slope immediately after entering the positive friction region of the positive side of the negative friction region.

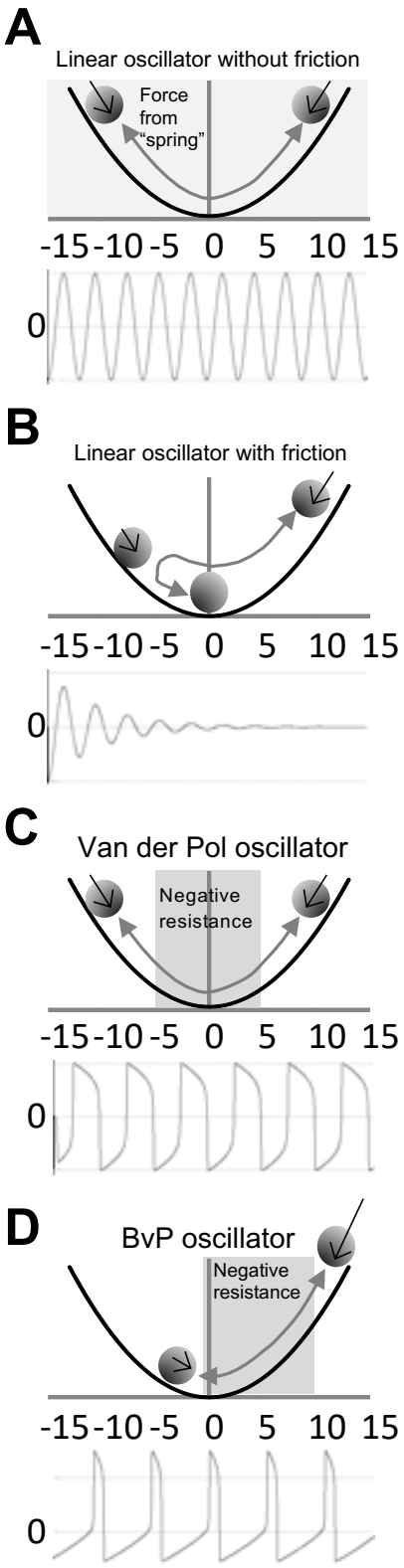


Figure 2.9. Schematic illustrations of poten-tial (bowl) and friction to understand various oscillations. The slope of the bowl corres-ponds to the force the ball receives from the spring. The ball is pulled strongly as it leaves the bottom.

The Duffing Oscillator – Jump of Convergence Point

In the previous section, we obtained slightly sharper and neuron-like waveforms by adding an asymmetric and nonlinear term in the friction of linear oscillator, in preparation for understanding the KYS oscillator that qualitatively reproduces the behavior of neuron's membrane potential. In this section, we will consider the effect of adding a nonlinear term to the stiffness of linear oscillator.

For example, let $a_2 = 1$, $b_2 = 0$, $c_2 = -25$, i.e., $g(x) = x^2 - 25$, and the others are zero, which implies that there is no friction ($f(x) = 0$) and D or the input to the KYS oscillator is zero (Fig. 2.10A). In this case, the potential $G(x)$, which is the integral of $g(x)x - D$ with the integral constant zero, is $x^4/4 - 12.5x^2$, a quartic equation with two bottoms. The oscillation of this oscillator depends on the initial condition of x. That is, if the potential at the initial value is higher than the potential at $x = 0$, it shows large oscillation symmetric to $x = 0$. In contrast, the initial potential is lower than the hill at $x = 0$, it oscillates within the valley near the initial position. In any case, they keep oscillating, since there is no friction.

Next, let us increase the friction without changing the shape of the potential $G(x)$. For example, suppose f(x) = 1 (Figure 2.10B). In this case, the ball falls to the bottom near the starting point. In general, there are two destinations in this system, depending on the initial value.

Then, we gradually increase the input to KYS oscillator D gradually from zero. When D = 20 (Fig. 2.10C), the potential D(x) is $x^4/4 - 12.5x^2 - 20x$. You can see that the left valley has become shallower. When the ball starts falling at x = -5, it stops at the left bottom.

As the D value is further increased, the left valley becomes shallower and shallower and eventually disappears. For the parameter values used here, the left valley vanishes around D = 38.5. If the initial value is set to -5, of course, the ball falls to the single remaining bottom (Fig. 2.10D). If the friction is sufficiently large and D is 38.4, the ball reaches the left bottom of the very shallow valley. A slight increase in D results in a jump of the convergence point. This sudden change is one of the bifurcation phenomena described above.

This oscillator is called the Duffing oscillator (the one with trigonometric function as input D is sometimes used as an example of chaos in the text books of nonlinear dynamics). In the next section, I will show how the Duffing oscillator is involved in the qualitative reproduction of the properties of a neuron.

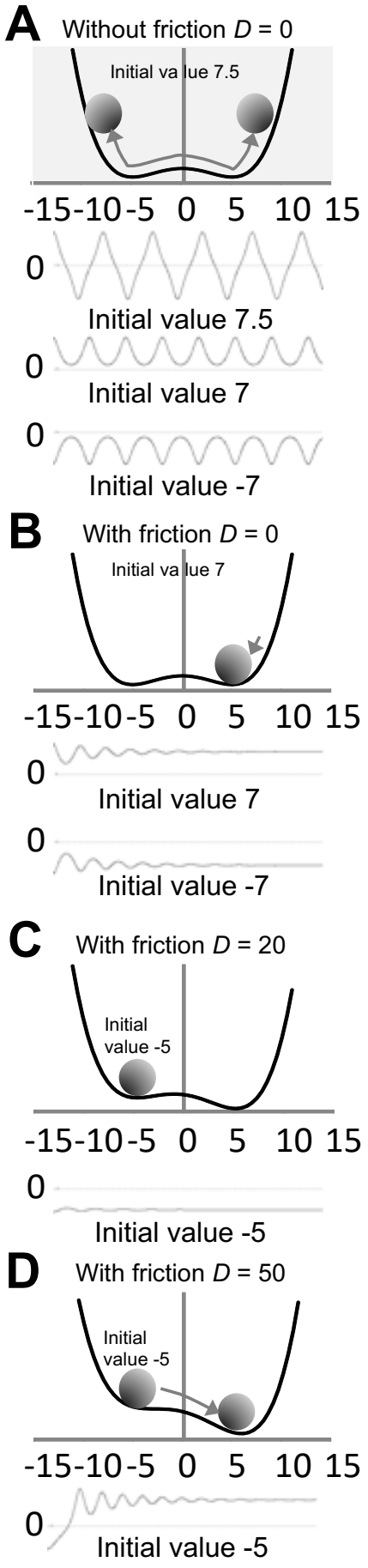


Figure 2.10. The Duffing oscillator. With (A) or without (B) friction. Different waveforms in A and B are of different initial values. Different D values result in different behavior of oscillation (B, C, D).

KYS Oscillator – Combining BvP and Duffing Oscillators

Here I describe the equation of the KYS oscillator that reproduces neuron's behavior.

$$d^2x/dt^2 = -f(x)dx/dt - g(x)x + D,$$
$$f(x) = a_1x^2 + b_1x + c_1$$
$$g(x) = a_2x^2 + b_2x + c_2$$

$f(x)$ and $g(x)$ are the friction and stiffness terms, respectively. Previously, in the section of the BvP oscillator, we overviewed what happens when $f(x)$ includes nonlinearity and asymmetricity, especially, $f(x) = x^2 - 10x - 1$. In this case, the waveform becomes a little sharper, looks slightly like a spiky waveform of a neuron. This oscillator is called a simplified version of the BvP oscillator. In the previous section, the nonlinearity of the stiffness term was examined. In particular, when $g(x) = x^2 - 25$ and the friction term has a large constant, we saw that the converging points alters depending on the initial value of x. In addition, we also observed that as D is increased, the number of the bottom reduces from two to one, resulting in a single point of convergence. This is called the Duffing oscillator.

The KYS oscillator is understood as the combination of these two oscillators. That is, here we consider the equation with $f(x) = x^2 - 10x - 1$ and $g(x) = x^2 - 25$.

Similar to the last section, we begin with the initial value $x = -5$, and increase D. When $D = 0$ (Fig. 2.11A), the potential $G(x)$ (the integral of $g(x)x - D$) is a quadratic function $x^4/4 - 12.5x^2$ with two valleys and symmetric about the origin. The convergence point is the bottom closest to the starting point. Dumping oscillations do not occur either. Even when $D = 20$, the shallow left valley remains, so the ball falls to the left bottom without dumping oscillations (Fig. 2.11B).

Next, let's look at the case where $G(x)$ has one valley as D is increased. For this parameter set, the potential has one valley when D is greater than 38.5. Fig. 2.11C is for $D = 50$. The ball is inevitably attracted to the single bottom. However, unlike the Duffing oscillator, the KYS oscillator has a negative friction region around the one remaining valley (shade areas in Fig. 2.11), which is inherited from the BvP oscillator. Therefore, the ball is subjected to the negative friction forces, and keeps oscillating without stopping at the bottom. In this sense, the KYS oscillator qualitatively reproduces the property of neurons that fire beyond a certain threshold.

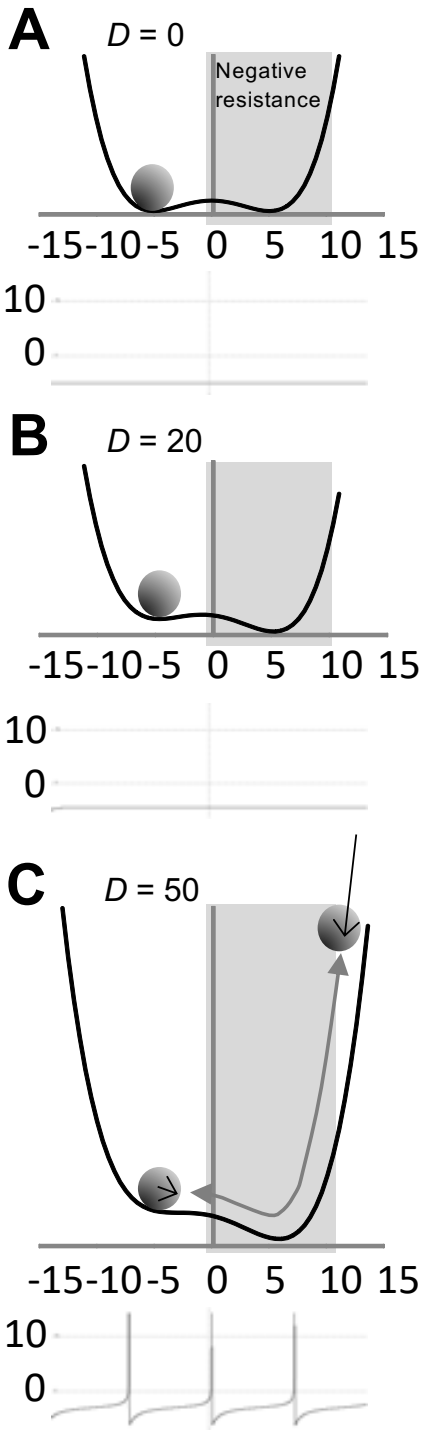


Figure 2.11. Properties of the KYS oscillator. The initial value of x is -5 in all cases. Thick solid lines represent the shape of the potential. A, B. When $D = 0$ and 20, the ball falls to the left bottom. C. For $D = 50$, the potential has one valley, but also has a negative friction region (shaded areas) around the right bottom. This structure make the ball to oscillate. The force that attracts the ball to the bottom is asymmetric on the both side of the bottom.

Correspondence between the KYS Oscillator and the Behavior of Neurons

The KYS oscillator, a second-order nonlinear differential equation that qualitatively reproduces the behavior of neurons, has nonlinearities in both the friction and stiffness terms. Here, we take a closer look at how it reproduces the neurons' properties.

As we saw in the previous section, when the input D exceeds a certain threshold, the number of valleys in the potential decreases from two to one. However, the ball, which is a metaphor for oscillator's behavior, does not settle at bottom of the single valley, and begins to oscillate because of the negative friction region around it, This is the reproduction of neuron's firing after the threshold is exceeded.

Its waveform is sharper than that of the BvP oscillator, and reproduces well the spike shape of actual neurons. These sharp waveforms are partially due to the asymmetry of the negative friction region with respect to the bottom, but are mainly due to the asymmetry of the potential shape. Fig. 2.12 illustrates the case of $D = 50$, as an example that D exceeds the threshold. The slope on the right side of the potential is so steep that the ball cannot stay in the positive friction region after exiting the negative region and is forced to return immediately. This is the mechanism of the sharp waveform.

The contribution of the left slope is more important: after D exceed the threshold, the left valley has already disappeared. However, if D takes a value slightly greater than the threshold, the left slope is so gentle that the system, or the ball, can move slowly around the former valley. This is a situation similar to climbing a mountain and taking a short rest on the gentle slope. This former valley is called an attractor ruin.

The gradient of the left slope can be controlled by changing the value of D. As noted earlier, for the values of D slightly greater than the threshold, the gradient around the attractor ruin is nearly horizontal, and the KYS oscillator oscillates at a very low frequency; as D increases, its gradient becomes steeper, and it oscillates faster (Fig. 2.12B). For the parameter set used here, the KYS oscillator varies its oscillation frequency over a wide range of the D, from 38.5 to about 700. This wide frequency range is a key feature of the KYS oscillator compared to the Van der Pol and the BvP oscillators.

As described above, the KYS oscillator qualitatively reproduces the firing properties of a neuron: it fires when its input exceeds a threshold value; the oscillation waveform is spike-like; and it can distinguish between the wide range of input values by its frequency.

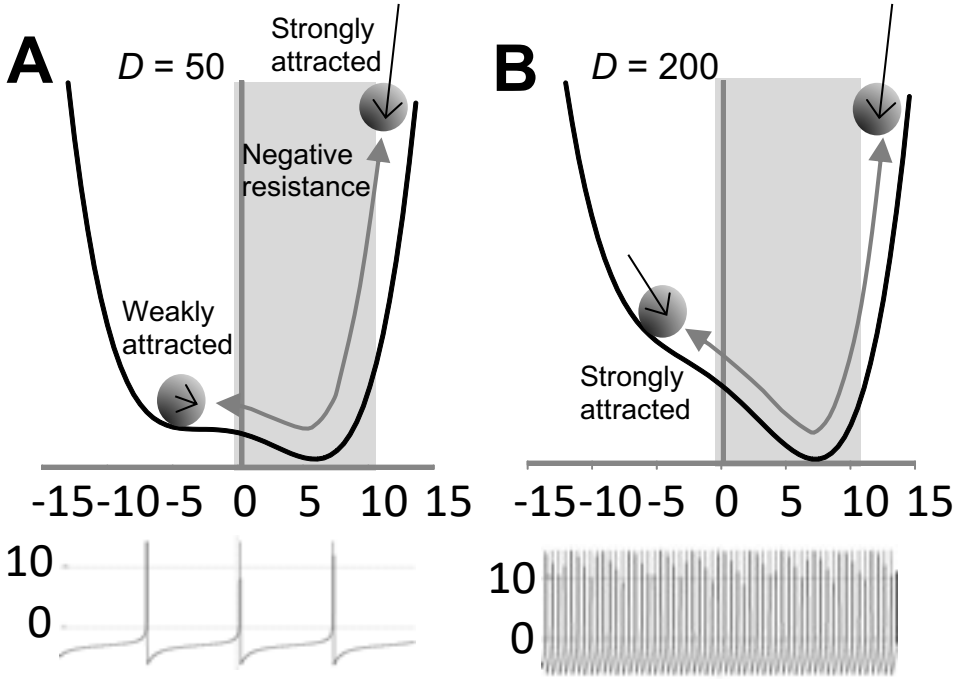


Figure 2.12. The behavior of the KYS oscillator after the input exceeds a threshold. A. $D = 50$, B. $D = 200$.

Why Do Neurons Emit Impulses? – An Evolutionary Perspective

The KYS oscillator qualitatively reproduces the characteristics of neuronal firing by combining two oscillators, the BvP and Duffing oscillators, which seemingly have nothing to do with neurons. This seems very strange to me.

As discussed in Chapter 1, Section II, oscillations are critical in living systems as nonlinear systems. The information processing in simple organisms such as slime molds depends entirely on the interactions between oscillators. However, the waveform of oscillation in slime molds is not spike-like. I might have been a slime mold in a previous life, but I do not remember it, so I do not how slime molds feel. But, they seem to live slower and less sensitive than I, a human being. Since the KYS oscillator oscillates with a spike-like waveform, it can flexibly take various inter-spike intervals and can sensitively encode a wide range of inputs by different frequencies. This neuronal property underlies our ability to notice and represent slight changes of environments. The wide range of inter-spike interval of the KYS oscillator is achieved by increasing the nonlinearity and asymmetricity of the friction and stiffness terms. The same thing must have happened during the evolution of life. Neurons must have developed to express information sensitively and richly. Of course, such an idea does not fit with scientific verification.

References

1) Delcomyn F. *Foundation of neurobiology*. WH Freeman, New York (1998)
2) Toyama K. ed. *Introduction to life science in the Nobel prize: Brain and neural functions*. Kodansha, Tokyo (2010) in Japanese
3) Sugi H. *Neural and synaptic science*. Kodansha, Tokyo (2015) in Japanese
4) Bear MF, Connors BW, Paradiso MA. *Neuroscience (3rd)*. Lipincot Wiliams & Wilkins, Philadelphia (2006)
5) Kandel E et al. eds. *Principles of neural science (5th)*. McGraw-Hill, New York (2012)
6) http://www.brainvision.co.jp/genspage/
7) Matsumoto G. Maruzen, *Phenomena and entities of neural excitation*. Tokyo (1981) in Japanese
8) Kimura S et al. A self-organizing model of walking patterns of insect. *Biol. Cybern.*, 69:267-283 (1993)
9) Sakamoto K. Dynamic properties of neural equations. In *Guideline of experiments B in electrical, communication, electronics and information course*, School of Engineering, Tohoku Univ. (1997) in Japanese

Chapter 3 Perceiving "Coherence" - Figure-Ground Separation and Synchronicity

Chapter 2 overviewed that neurons can be understood as a nonlinear oscillator. Chapter 1 Section II outlined that slim molds can be models as a network of nonlinear oscillators, and that its integrity as a whole is achieved through local interactions between the oscillators (see Appendix A3 for details).

Specifically, the primary visual cortex (V1), the main entry point for visual input in the cortex, will be discussed, along with the basic properties of neurons in V1, called orientation-selective cells, and how nonlinear oscillators can be involved in information processing in V1. Through this discussion, we will address the issue of figure-ground separation and the problem of synchronicity. The figure-ground separation problem refers to the problem of separating an object (figure) from its background (ground) and extracting it as a coherent whole in perception and recognition. The problem of synchronicity is the problem of meaningful temporal coincidence.

Simple organisms, such as slime molds, tend to take stereotyped actions in response to limited sensory stimuli. On the other hand, animals with well-developed brains need to segregate meaningful objects from their environment by integrating basic clues from a large variety of perceptual input. In the latter, the problem of synchronicity becomes more pronounced.

I. The Primary Visual Area V1 and Figure-Ground Separation

Orientation Selective Cells – Neurons that Respond to a Specific Orientation of a Line Segment

Light entering the eye forms an image on the retina. Light at each position in the image is detected by photoreceptor cells in the retina. The area of the visual field from which photoreceptor cells receive input is called the receptive field. The term "receptive field" is an important concept in neuroscience. It is widely used to describe the "area of responsibility" to which neurons respond.

Signals detected by photoreceptor cells reach the primary visual cortex V1 via contrast processes in the retina and lateral geniculate nucleus of the thalamus (Fig. 3.1). Along these processes, all signals are not mixed at once. Instead, the receptive fields are enlarged in each process through some computations and integrations.

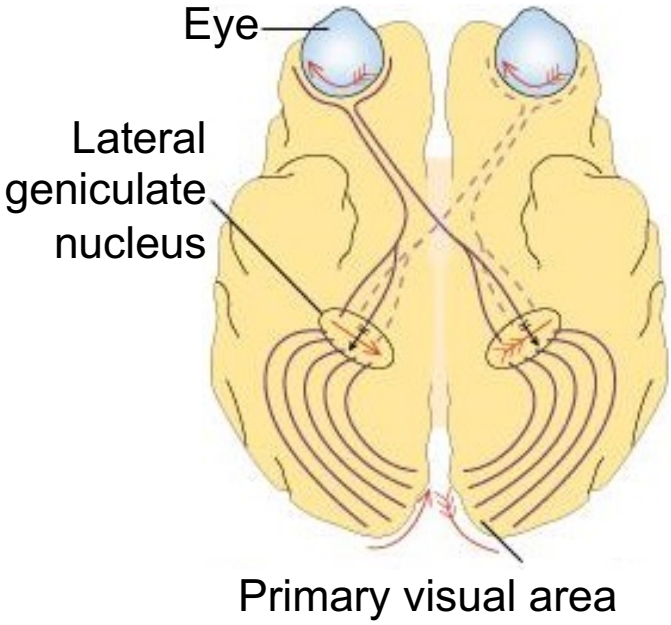


Figure 3.1. From retina to V1.The images in the right and left visual hemifields of each eye and project to the left and right hemisphere, respectively, while preserving the approximate positional rela-tionship in the visual field.

V1 neurons have receptive fields. That is, these neurons respond to stimuli presented in a limited area in the visual field, which was revealed by the examining the firing properties of V1 neurons by penetrating a microelectrode. However, they do not respond to all types of stimuli presented to the receptive field. Most of them prefer a line segment with a certain tilt, referred to as orientation. Namely, these neurons are called orientation-selective neurons (Fig. 3.2).

In 1981, David Hubel and Torsten Wiesel were awarded the Nobel Prize in Medicine for their discovery of orientation-selective neurons. Their story of discovery after long efforts is a good example of serendipity. Hubel says “After about five hours of struggle, we suddenly had the impression that the glass with the dot was occasionally producing a response, but the response seemed to have little to do with the dot. Eventually we caught on: it was the sharp but faint shadow cast by the edge of the glass as we slid it into the slot that was doing the trick.”[1)] A divine revelation of science may come in this way.

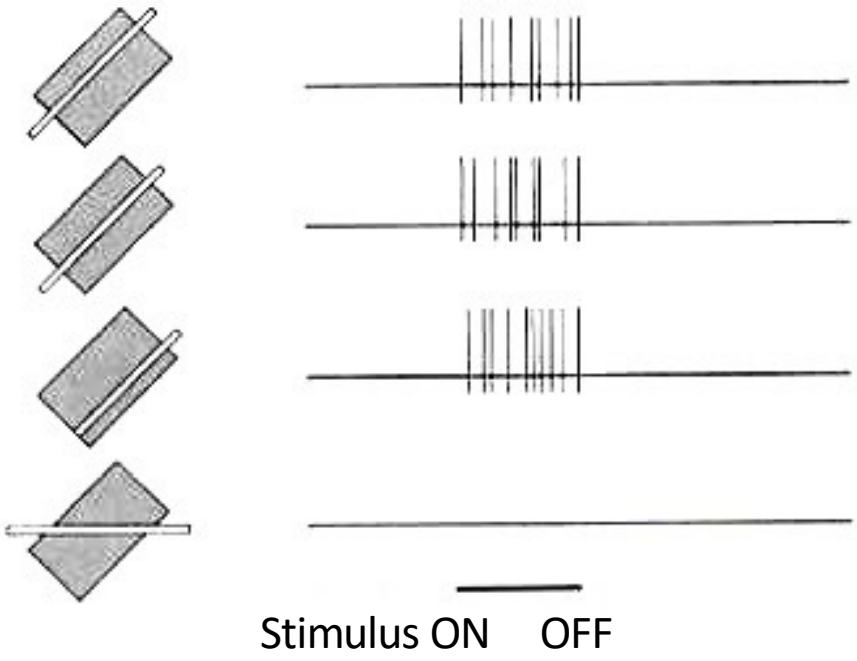


Figure 3.2. Responses of a V1 orientation selective neuron (complex cell). Left: stimuli (white bars) and the receptive field (gray). Right: Firing responses to the stimuli：It does not respond to the bottom-most stimulus[1)].

Finding Relationships – Basic Processing of Figure-Ground Separation

You may not be familiar with the phrase, figure-ground separation. It refers to the extraction of meaningful information from the background in perception and recognition. Some readers may think, "That means, in essence, signal detection." Many people seem to consider that perception and cognition is nothing more than the detection of a signal from an object that has some physical properties to be perceived or recognized.

Figure 3.3. Christ's image.

Then, what do you perceive in Figure 3.3? In this figure, something is drawn in black and white. But, physically, that's all. This is an image of Christ. If you find a relationship between the beard, mustache, hair and eyes, you would recognize it as a Christ's image. Physically, however, the statue and the background are indistinguishable.

Thus, perception and cognition are not passive processes for detecting physical signals in the outside world, but active processes for finding consistent relationships between clues.

The figure called Kanizsa's triangle clearly indicates the importance of finding consistent relationships between clues in perception and recognition (Fig. 3.4)[2]. This triangle is visible only when the three Pac-Mans have a coherent relationship, and not otherwise. Surprisingly, the inner side of the triangle appears brighter than the outer side, and even a contour, called illusory contour, is perceived between the Pac-Mans.

Some readers may still think, "Oh, this is a pattern-matching process in the brain. I mean, it's matching the visual image to a template of shapes in the brain, such as Jesus or triangle, acquired through experience."

What do you see in Figure 3.5? This shape is certainly unfamiliar. But you must be able to perceive it or any kind of illusion contour without difficulty. However, it is inefficient and unlikely that you possess every templates for every shape you have encountered in the brain.

These figures indicate that percepts are not unilaterally determined by external physical causes, but emerge through active and autonomous brain processes that find consistent relationships among various clues and create a coherent whole.

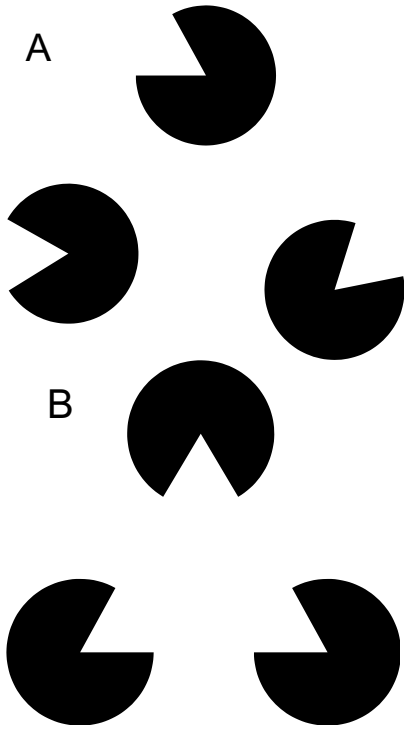


Figure 3.4. Three Pac-Mans (A) and Kanizsa's triangle (B).

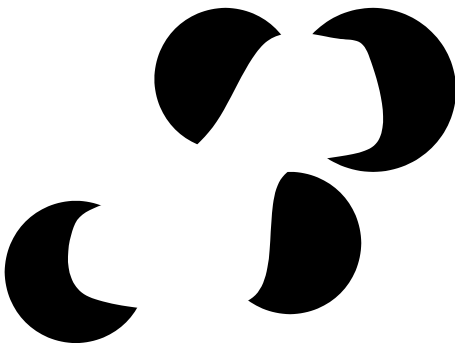


Figure 3.5. Illusory contours can be perceived for arbitrary shapes.

Holovision – A Computational Model for Figure-Ground Separation[3), 4)]

One of the most fundamental processes in perception and recognition is to segregate a meaningful entity from the background. We can easily perform figure-ground separation even for novel objects. This is the reason why even inexperienced part-time workers at archaeological or fossil digs can distinguish ancient pottery or fossils from ordinary stones.

As discussed in the previous section, figure-ground separation requires finding good and consistent relationships among the detected cues or features, and generating clusters or wholes in an active and autonomous manner.

A computational vision model called holovision was developed by Dr. Yoko Yamaguchi of Prof. Hiroshi Shimizu's laboratory, where I studies when I was a student at the University of Tokyo (Fig. 3.6). This model aimed to perform figure-ground separation by mutual entrainments between nonlinear oscillators. In this model, orientation selective cells in the primary visual cortex V1 of the cerebral cortex, which detect line segments with a specific orientation displayed within a certain region of the visual field, are regarded as nonlinear oscillators. Figures or perceptual coherence are self-organized by binding the line segments coded by each oscillator. That is, the consistent relationship between line segments emerges as synchronization between oscillators. Mutual entrainment was considered to be advantageous for evaluating the global consistency of entire figure without being bound by local minima (see Appendix A for details).

Figure 3.6A shows an overview of holovision: the R-plane corresponds to the retina; information from each domain (receptive field) on the R-plane is input to the S-units; each unit of the S-units corresponds to an orientation-selective cell and responds well to the input of its preferred orientation line segment. Each unit in S-units strongly is connected to its neighbors in the tangential direction of the preferred orientation.

Figures 3.6B and C depict examples of inputs and outputs, respectively. The area surrounded by the dashed lines in Figure 3.6B contains a triangle. You can see synchronous oscillations between the units that encode the line segments of the triangle (No. 1-6), not between the other units (No. 7, 8).

So far, many computational models using oscillators have been proposed. However, it is surprising and epoch-making that Holovision was proposed in the mid-1980s before the Singer's experiment, which will be described in the next section.

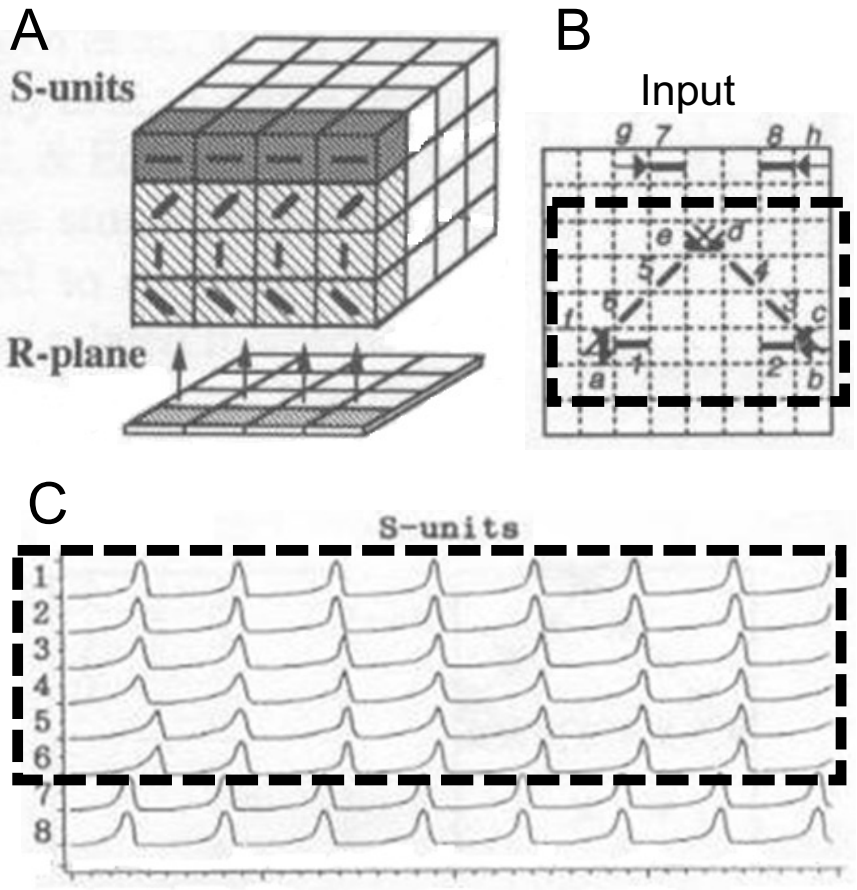


Figure 3.6. Holovision. A: Overview of the model. B: input example. C: the output for input B. Dashed lines in B and C are by the author (From Ref. 3).

Synchronous Firing between V1 Neurons

To achieve figure-ground separation, i.e., the perception/recognition of an integrated object segregated from the background, the observer has to actively establish some consistent relationships between the detected cues and features. In the previous section, a vision model called holovision, which was aimed to achieve figure-ground separation, was introduced. This model used mutual entrainment to generate consistent relationships among line segments such as triangles, by regarding orientation selective cells, the line segment detectors in the primary visual cortex V1, as nonlinear oscillators.

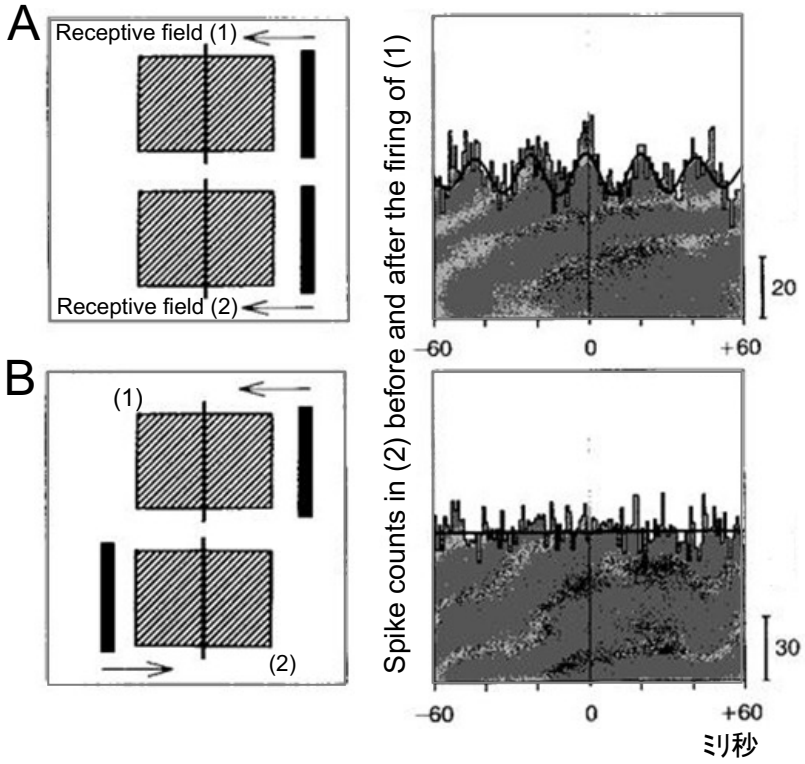


Figure.3.7. Synchronous or unsyn-chronous firings between orientation-selective neurons. A,B. The orientation of the stimulus is optimal for the two receptive fields. However, these activities synchronize when the two bars move in the same direction (A), but not when they move in the opposite direction(from Ref. 6).

Experimental results demonstrating the validity of this mid-1980s model were published in 1989 by Wolf Singer and colleagues in Germany[5]. They inserted two microelectrodes into V1 and analyzed the activities of neurons whose receptive field (the spatial extent of cellular response in the visual field) were adjacent to each other and the preferred orientations are tangentially aligned (Fig. 3.7). In general, the orientation-selective firing activity is greater for moving than for stationary line segments, but is less dependent on the direction of movement. How, then, does neuronal activity distinguish between two bars moving in the same direction (Fig. 3.7A left) and moving in opposite directions (Fig. 3.7B left)? In particular, in the former case, are there any neuronal phenomena that correspond to the impression of a perceptual cluster?

They found that the neuronal activities recorded from the two electrodes are not only well activated but also oscillate synchronously when the two bars move in the same direction (waves with a peak at time zero in Fig. 3.7A right), while such synchrony does not appear when the two bars move in the opposite direction (the flat profile in Fig. 3.7B right). From these results, they proposed the binding hypothesis of features through synchronous oscillations (Fig. 3.8). That is, they considered that visual coherence is achieved by synchronous firing between orientation-selective neurons that respond to the line segments that compose it.

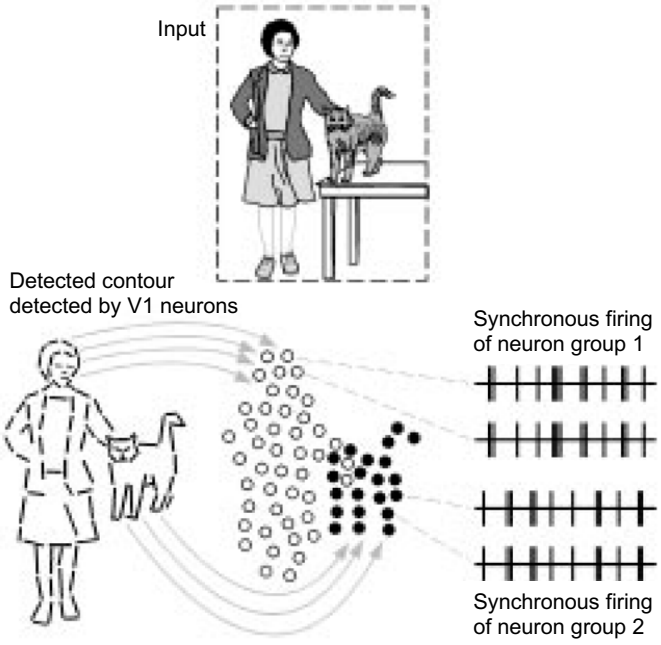


Figure 3.8. Binding hypothesis through synchronous firing (from Ref. 7).

II. Indefinite Environment and Synchronicity

Coherence and Synchronicity

In the previous section, we have overviewed figure-ground separation and synchronous phenomena. To extract a “figure” as a perceptual coherence as from a complicated back-“ground”, some consistent relationship must be established between elemental cues detected and represented. As a neural correlate to such perceptual/cognitive process, I have discussed synchronous oscillations of neuronal activities and introduced a corresponding model and experiment.

This means that a whole or cluster in perception and recognition emerges as a synchronic order of its constituent features. Neurons as nonlinear oscillators can flexibly interact with each other. They interact with the surroundings, saying, “Now I’m representing XX, but what about you?” There are tremendous number of such interactions, which circulate if there is no inconsistency among them. Such circular or coherent interactions between neurons can emerge some temporally-corelated or synchronous neural activities in a broad sense in the help of neurons’ intrinsic tendency to synchronize. In other word, the neural activities composing the circular interaction cannot be uncorrelated nor fall into disorder. Here, I refer such emergence of coherence as synchronic order.

You See It because It Is There, and It Is There Because You See It

The circulation of action, which contains no inconsistencies, is not limited to just between neurons. It is also between ourselves and the environment that surrounds us. The figure-ground separation problem is a good example. An image or percept extracted as a "figure" from complicated "ground" is not always be obtained by mere physical signal detection. As outlined in the first section, it is obtained only from an active observation or discovery of consistent relationships among clues. However, this does not mean that it is just an illusion of each individual. If this were the case, it would be impossible to share experiences with others. There is something we can share our experiences in the environment. In this sense, an image or percept emerges in the circular structure of "you see it because it is, and it is there because you see it." The fact that you can see something is a synchronic order that emerges between ourselves and the environment.

Synchronic order does not emerge only in perception and recognition. In systems where elemental things A and B are mutually dependent, desirable states emerge synchronously. In Chapter 1, I mentioned that many of the systems in which human society faces serious problems are ones in which observers or actors strongly interact with the environment or with others. In these systems, some consistent relationship emerges as a temporal or synchronic order in the broad sense.

Serendipity – Finding Synchronic Order

I have argued that in complex systems with strong interactions, such as between neurons or between observer and environment, some coherent relationship emerges synchronously in a broad sense of synchronicity. Here I will further emphasize that the issue of synchronicity is more deeply related to the way we, living systems, live in this world.

The world we live in is definitely more complex than ourselves. No matter how complex we may be, this world is inherently more complex than we are because we are part of it. In fact, it is so overwhelmingly complex that the unexpected things can happen at any moments no matter how hard we try to predict it. We always face small or big happenings.

In the following, I will use the term "indefinite environment," so that we can be clearly aware that the real world inevitably includes an essentially unpredictable aspect in which, when you cast dice, a number other than one from six can come up, namely, the probability space of the dice cannot be defined. I believe that the author's mentor, Prof. Hiroshi Shimizu, has transformed the ancient question, "What is life?" into "how can biological systems adapt to an indefinite environment?" with his outstanding insight. He also repeatedly emphasized the importance of synchronicity in living in such a environment.

Prof. Shimizu came up with the above idea in the context of modern science. In the following, I will show that his idea is not ridiculous by introducing the ideas of Jung and Aristotle.

In his article 'Synchronicity: The principle of non-causal relations'[9], famous depth psychologist C.G. Jung argued that we find meanings through synchronicity or coincidence in vast realms in the real world, where causal explanations do not hold. Causality here means a prediction from laws derived from experience. He pointed out the importance of noticing the co-occurrence of things that cannot happen by chance from forecasts based on experience for discovering new aspects of the real world. This ability to find meanings in synchrony is also called serendipity. As mentioned earlier, the discovery of orientation-selective neurons in the primary visual cortex was a good example of serendipity.

On the other hand, Aristotle, in his "De Anima"[10], asked a simple but essential question: Why do we have different modalities such as sight and hearing? He then sought for the answer in synchronicity. He said, "when different senses detect something coincidentally, you perceive a single percept, not different senses." In other words, when the different sensory organs or related brain regions detected a signals simultaneously, we feel an entity or cause that gave rise to different sensations. This also means that we have multiple modalities to know the substance or cause behind them through synchronous detection.

Meaningful relations emerge synchronously in the indefinite environment, while, by actively capturing the synchronic order, we, the observers or agents, achieve harmonious relations with the environment in a synchronic manner on the spot. This is, I believe, how we manage to adapt to the ever-changing indefinite environment.

Hebbian Rule – Structuring the Synchronic Order

I have been repeatedly stated that something meaningful including consistent relationships between the constituents emerge synchronously in an indefinite environment where things interact strongly. Synchronic order is valuable, but disappear in an instant. However, the brain has a mechanism to capture it and keep in its structure, which is called learning.

We still have many things to be revealed in the learning mechanism in the nervous system, but what has been studied most is the synapse, at which a neuron (presynaptic neuron) sends signals to the next neuron (postsynaptic neuron) shown in Fig. 3.9A. Numbers of studies have been conducted on the assumption that gain changes in signal transmission from the presynaptic cell to the postsynaptic cell are the most fundamental mechanism of learning.

One type of the mechanisms of synaptic gain change captures synchronicity. This type of synapse is called Hebbian synapse after Donald Hebb who proposed this concept. In this type, gain increases when pre- and postsynaptic neurons are co-activated within a certain time-window, which is called Hebbian rule. That is, "neurons fire together, wire together." The reverse is also true. Namely, asynchronous firing can weaken synaptic gain.

This implies that when different inputs enter a neuron at the same time, they are strengthened. A postsynaptic neuron receiving numerous synaptic inputs cannot be made to fire with a single input. For a postsynaptic neuron to fire, multiple presynaptic neurons have to fire synchronously (inputs from X and Y to x in Fig. 3.9B). Conversely, presynaptic neurons that fail to make the postsynaptic neuron to fire eventually lose their inputs (input from Y to y in Fig. 3.9C).

However, neurons are not passive. When the spontaneous activity of a neuron is high, synaptic gain increases for weak inputs. Many of you may have had the experience of not being able to learn when you passively took a class, while you were able to understand them well when you studied actively. Similar cases can also occur at neuronal and synaptic level. In the black-and-while picture shown in Fig. 3.3., once you recognize Jesus, you can never go back to not recognizing Him. This can also be explained in the same way. That is, the Hebbian rule is also the cellular basis for maintaining active recognition results.

The brain has developed a mechanism to preserve good encounters over the long process of evolution.

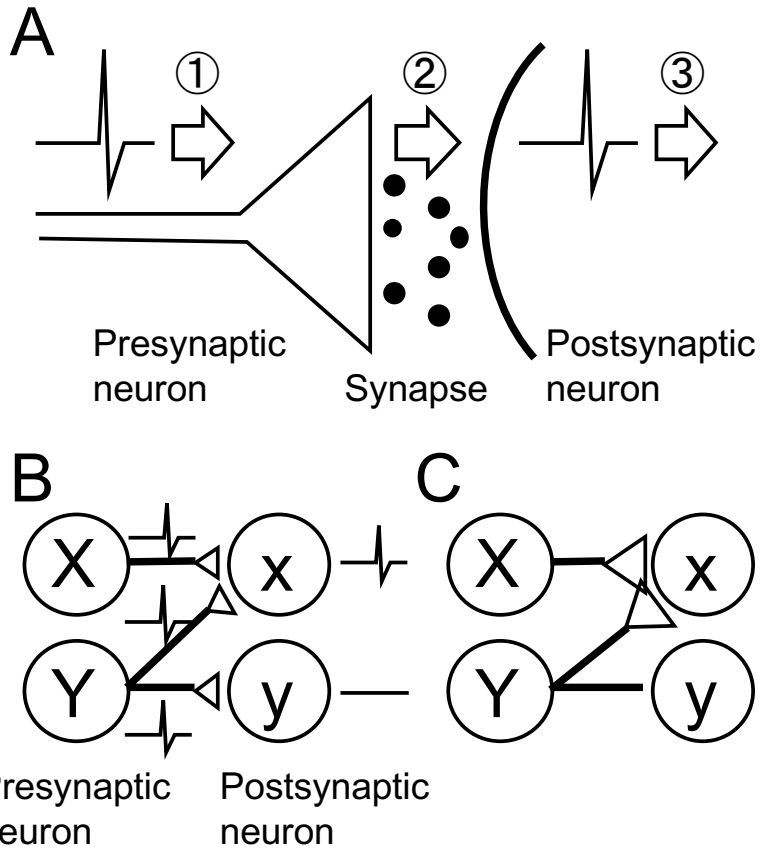


Figure 3.9. Synapse and Hebbian rule. A.Presynaptic cell activity ① causes neuro-transmitter release ②. When the membrane potential exceeds the threshold, the postsynaptic neuron fires ③. B, C. X and Y can fire x but not y, resulting in synaptic gain difference(indicated by triangle size).

How to Make Consistent Relationships

In this chapter, trough the discussion of figure-ground separation, I have discussed from multiple perspectives that our percept, recognition *etc*., emerge as a consistent relationship or synchronic order between observer of the world and indefinite environments including fundamental unpredictability.

However, we do not have clear principles on how to create a synchronic order and what kind of synchronic order is desirable to implement a system that emerges such order, because we have too many possibilities and degrees of freedom. In fact, the holovision, which attempted to solve figure-ground separation using mutual entrainment between nonlinear oscillators, could not handle real images despite tremendous efforts of Dr. Yoko Yamaguchi. One reason may have been the lack of oscillators with high entrainment performance such as the KYS oscillator at that time. However, I believe that the problem was not that particular issue, but something more fundamental.

The problem, in my opinion, lay in the firm policy for having rules in deciding what should be entrained and what should not. We can call these rules constraints. Many of the problems that the brain faces and solves are ill-posed problems, i.e., problems for which states or solutions cannot be uniquely obtained from the given conditions and information alone. For example, in the figure-ground separation problem, it is not clear which clues should be bound together. The holovision was faced with the problem of what constraints were needed to solve the ill-posed problem of figure-ground separation.

What makes the problem even more difficult is that the constraints often have to be generated in situ. For example, if a new fossil is partially buried in the soil, we can easily recognize "something" that is not just a stone. Recognizing shape requires constraints to bind limited clues together. However, to deal with novel objects, we need to generate constraints, or at least recall appropriate ones, though evaluating their adequacy is not easy. In any case, we are pursuing the principles of how to "create" in life and in the brain. Thus, the problem of constraint generation cannot be avoided.

The problem of constraint generation has been a major challenge for Prof. Hiroshi Shimizu since just before his retirement from the University of Tokyo, and he has published several works on the subject[11), 12)]. However, I do not think that he was able to tackle this problem through specific scientific themes, since he retired from the forefront of science. This problem is too difficult for me to solve with my poor talent. Therefore, I cannot provide the answer to you, but will address this issue head-on in Part II.

References

1) Hubel DH. *Eye, brain, and vision*. Scientific American Library, New York (1988)
2) Kanizsa G. *Organization in vision : Essays on gestalt perception*. Praeger, New York (1979)
3) Shimizu H et al. Pattern recognition based on holonic information dynamics: towards synergetic computers. In Haken H ed. *Complex systems - operational approaches in neurobiology, physics and computers*. 225-239, Springer, Berlin (1985)
4) Yamaguchi Y, Shimizu H. Pattern recognition with figure-ground separation by generation of coherent oscillations. *Neural Networks*, 7:49-63 (1994)
5) Gray CM et al. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334-337 (1989)
6) Engel AK et al. Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci.*, 15:218-226 (1992)
7) Engel AK et al. Role of the temporal domain for response selection and perceptual binding. *Cereb. Cortex*, 7:571-582 (1997)
8) Shimizu H. *Seimei wo Torae-Naosu (2nd)*. Chuko Shinsho, Tokyo (1990) in Japanese
9) Jung CG. Synchronicity: The principle of non-causal relations. In Jung CG, Pauli W. *The interpretation of nature and the psyche*. Bollingen Foundation, New York (1955)
10) Aristotle *De Anima (on the soul)*.
11) Shimizu H. *Seimei-Chi to shite-no Ba no Ronri*. Chuko Shinsho, Tokyo (1996) in Japanese
12) Shimizu H. *Ba no Shisou*. University of Tokyo Press, Tokyo (2003) in Japanese

Chapter 4 Creative Planning of Behavior: Neuronal Dynamics in the Prefrontal Cortex

Our daily lives are filled with unexpected events to a greater or lesser extent. No matter how much you try to anticipate, what you never imagine can happen at any time. For example, you may have an experience that you went to a certain place and were in trouble to see the place has changed drastically. In this book, we refer to the real world, which inherently contains unpredictability, as an “indefinite environment.”

Animals do not give up easily when faced unpredicted events. They often managed to do something at any level. In the previous chapter, we discussed the problem of figure-ground separation in vision. We can segregate a figure from the ground from relatively early stages of development. In addition, when an insect has one leg amputated, it is able to walk with the remaining legs, although this is new to that individual. Such flexible adaptation has not been achieved by current robots. Thus, creativity in the broad sense may be realized at any level of the living system.

But, this chapter will focus on creativity in a more narrow sense, on what we consider creative. Specifically, I will show you mainly our results on the issue of problem solving and behavioral planning requiring inspiration and related neuronal activities in a region of the cerebral cortex called the prefrontal cortex (PFC).

Difference between Inspiration and Complex Behavior

When it comes to creativity, you may consider things that are not related to the ordinary, such as Picaso’s masterpieces that no one else can paint or Bach's improvisational and complex compositions. But the essential aspects of creativity are no stranger to us and need not be amazing nor complex.

One day, a famous robot was in the entertainment, walking around the party room and shaking hands with the guests. Naturally, the guests were delighted. But, when the robot was requested another handshake, it was unable to do it again. This robot was famous for its complex actions such as spectacular dances. However, even the simple action of shaking hands, if it had not been supposed in advance, was seriously difficult for the robot, The robot should have improvised to do another handshake, because it did not have time to learn it.

Certainly, it would be very helpful to have a robot that can perform a determined and complex task as accurately and quickly as an industrial robot. However, if we need a robot to help humans in their daily lives, we need some kind of "one more handshake" inspiring mechanism. This is because, unlike on a factory production line, there are many unexpected things that happen in everyday life. Like, "Who took the ear buds out of here?" or "What? Don't you need lunch today?” etc. At least in our house, we are in a very indefinite environment!

Immediate Goals are Required to Achieve the Final Goal

We have been working on the issues of problem solving and behavioral planning to elucidate the brain mechanisms of creativity and inspiration. Elucidating the mechanisms related to these issues seems to be more practical and easier to tackle as scientific research than working on painting or music problems, since it is clear what works and what doesn't.

Problem solving and action plans are difficult to define. However, generating specific measures and actions is critical to solve a problem and achieve a goal. The example in Figure 4.1 illustrates this aspect well. This famous Japanese police officer came up with the specific action of hooking the nose to the crane to achieve his final goal of taking an Anpanman doll.

Figure 4.1. Example of problem solving and behavioral planning. From Osamu Akimoto, "KochiKame: Tokyo Beat Cops" *Shonen Jump*, No.8 (1991) with permission.

Path-Planning Task

To explore the neuronal mechanisms related to problem solving and behavioral planning in experiments, it is necessary to record neuronal activities while animals perform a behavioral task. To this end, Prof. Mushiake of Tohoku University School of Medicine devised a path-planning task for Japanese monkeys[1]. The task involves moving the cursor toward the final goal on a grid-like maze presented on the screen (Fig. 4.2 and 4.3). First, a final goal was presented in a corner of the maze, followed by a delay period. Any cursor paths were allowed, though a part of the path was blocked in some trials. This task is an ill-posed problem in the sense that path cannot be uniquely determined. After the above preparatory period, a Go signal was delivered, cueing the animal to move the cursor to the next crossing, which was referred to as the immediate goal. The monkey could reach the final goal in three steps in the shortest, but there was no limit to the step number. The animal was rewarded for reaching the final goal.

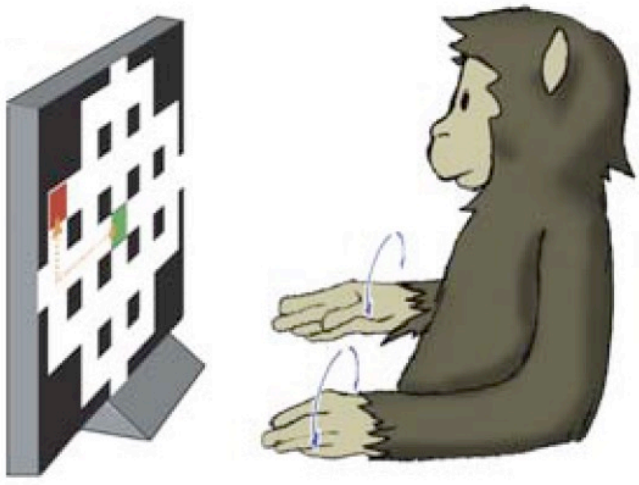


Figure 4.2. Overview of the path-planning task (from Ref. 2).

The monkey grasps the manipulanda with both hands. The animal was allowed to make four hand movements: right supination, right pronation, left supination and left pronation. These movements move the cursor up, down, left and right. The correspondence between hand movements and cursor operation was switched every fixed number of trials.

It took more than a year to train a monkey to learn the task. The monkeys have experienced every combination of final-goal, block and hand-cursor assignment. Hence, some may doubt that the monkey only performs the learned behavior, not think out nor create a behavioral sequence in each trial. Actually, the animals did not show a novel sequence in every trial. However, to understand the task structure that the subject is required to move the cursor to reach the final goal seems more efficient than to learn every variety of trials. In fact, they sometimes took wrong cursor movements, but corrected immediately after they recognized it during a single trial. These facts suggest that they decided the movement sequence on each trial, which was also supported by the neuronal activities I show in the following sections.

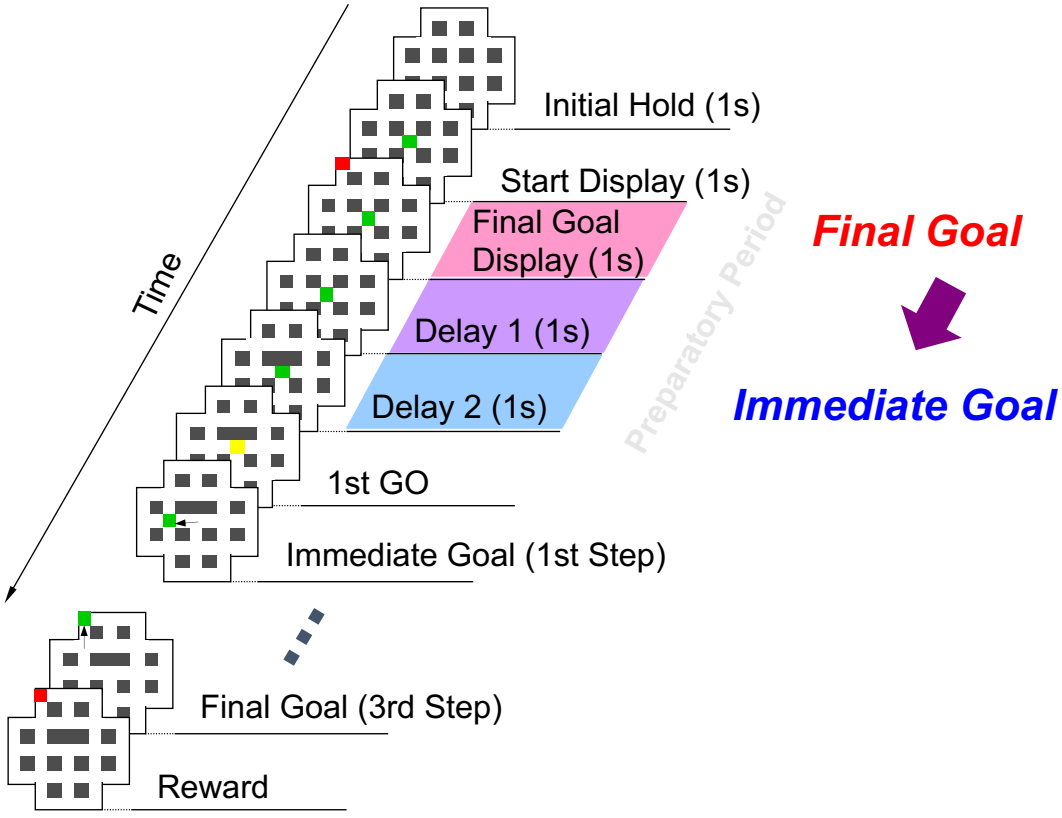


Figure 4.3. The time-course of a trial of the path-planning task (from Ref. 3).

Planning and Problem Solving: The Functions of the Prefrontal Cortex

The prefrontal cortex (PFC) of the cerebral cortex is known to play crucial roles in problem solving and behavioral planning. Anatomically, PFC is located away from the primary sensory areas such as V1 as the front ends of external signals, and the primary motor cortex (M1), the area closest to muscles in the cortex. That is, PFC is in between sensory/cognitive and behavioral/motor processes. As Figure 4.4 shows, "smarter" animals seem to have larger PFC. Indeed, the PFC patients are characterized by impairments in organizing behavior, such as behavioral planning, based on perceptual information.

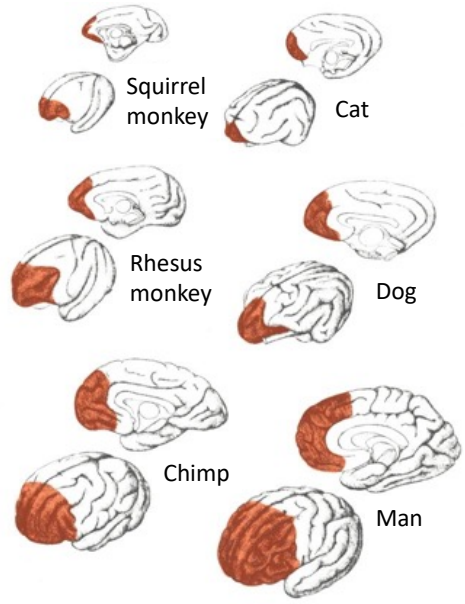


Figure 4.4. Location and size of PFC in different animals (shaded areas) (from Ref. 4).

Early single-unit studies of the monkey prefrontal cortex primarily used delayed response tasks requiring working memory. Working memory actively stores information for short periods of time and is indispensable as the basis for thinking and conceptual manipulation. It corresponds to computer memory.

The oculomotor delayed-response task is as follows. When the monkey fixates at the center spot, another spot appears in the surrounding for a short period of time as the target. The animal has to keep looking at the center for several seconds after the peripheral target disappears. Thereafter, the center spot turns off, which serves as the go signal. The monkey is rewarded if he/she shift his/her gaze to the position where the target used to locate.

When neuronal activities are recorded from the lateral PFC during the task, you can find neurons that show persistent activity even after the target has disappeared. That is, this type of neurons fire as if they memorized the target location(Fig. 4.5).

There is no doubt that working memory is a major basis of intellectual processes. However, it must exist for thinking and problem solving, which have not been well studied at neuronal level.

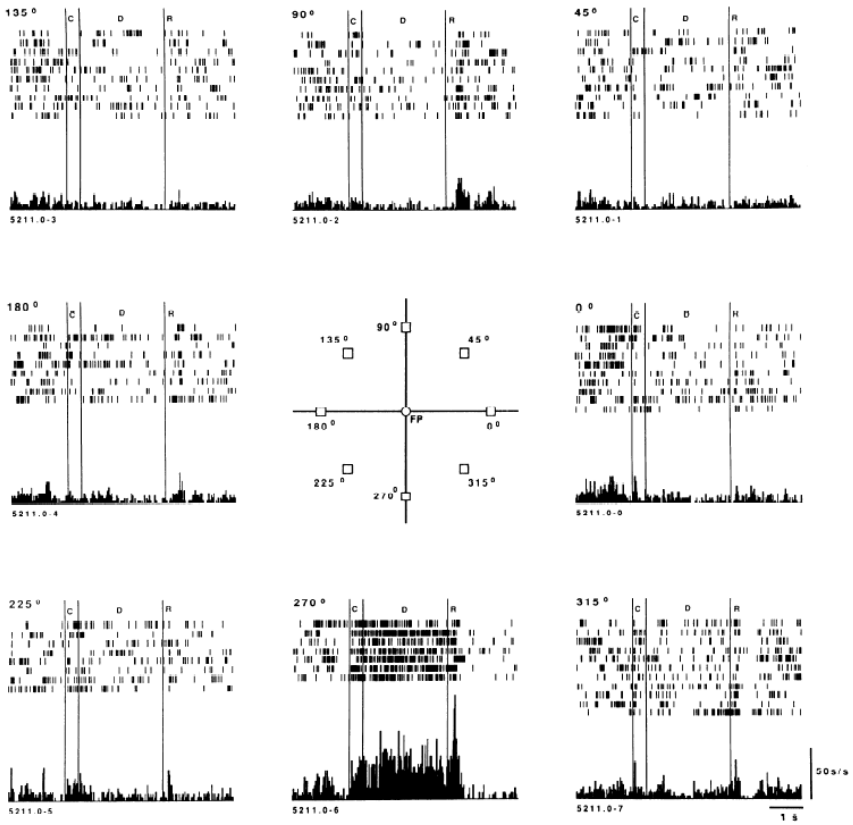


Figure 4.5. An example of lateral PFC neurons' activities for eight different target positions during an oculomotor delayed response task. This neuron shows a persistent activity for the bottom target. The firing activities (black dots) of each trials within about ten trials were plotted (top) and summed to provide histograms (botton) for each target (from Ref. 5).

BOX *The Higher Motor Areas in the Frontal Lobe*

There are several areas between the lateral prefrontal cortex, discussed in this chapter, and the primary motor cortex, which projects to the spinal cord and is the most directly involved in muscle contraction among areas of the cerebral cortex (Fig. 1). The distinction of these areas is the accumulated results of anatomical and physiological findings. Prof. Jun Tanji, who supervised me for many years, and his laboratory at Tohoku University made great contributions to the identification of these higher motor areas in the frontal lobe from the 1980s to the mid-2000s.

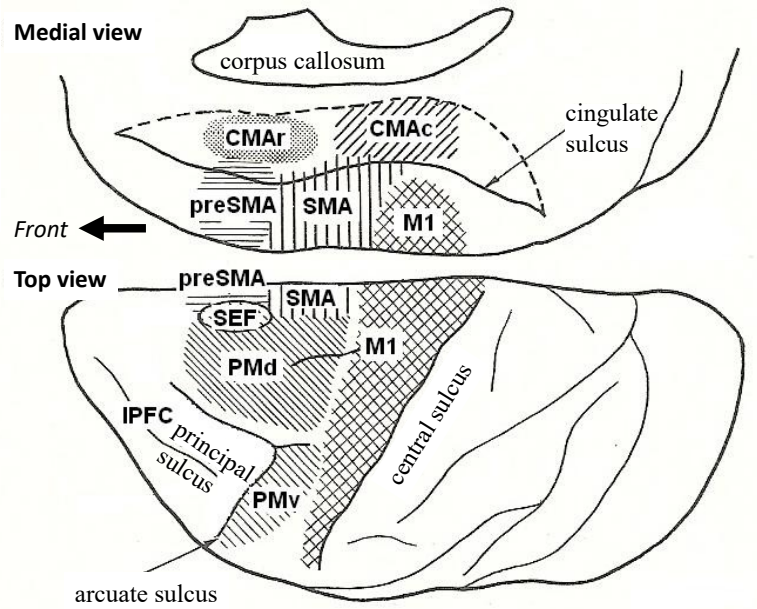


Figure 1. Schematic drawing of multiple motor areas in the cerebral cortex of monkeys. Medial (top) and top (bottom) views of the left hemisphere (from ref. 6). CMAr: rostral cingulate motor area; CMAc: caudal cingulate motor area; preSMA: pre-supplementary motor area; SMA: supplementary motor area; M1: primary motor cortex; SEF: supplementary eye field; PMd: dorsal premotor area, PMv: ventral premotor area, lPFC: lateral prefrontal cortex.

One work by Prof. Mushiake of Tohoku University, my long-time collaborator, is very famous. The premotor and supplementary motor areas were not well distinguished before the 1990s. To clarify the distinction, they trained monkeys to execute a sequential button-pushing task (Fig. 2). It included two modes: one is visually guided and the other is memory guided. In the former, the monkeys were required to press the sequentially illuminated buttons in the order 3→1→4, for example, which was fixed in a trial block. The animals were so smart that they learned the order during the visually guided trials. So, when the task mode was switched to memory guided one by turning off the button lights, they could easily press the buttons in the correct order.

Neuronal activities in the primary motor area directly reflect hand movements to be executed. Therefore, the neurons fired well when the monkey push the button, regardless of whether the animal performed visually guided or memory guided modes of the task (Figure 2 left). On the other hand, neurons in the premotor area were activated during visually guided trials, while they were quiet during memory guided trials (Figure 2 middle). In contrast, neurons in the supplementary motor area did not show strong discharges during the visually guided mode, while they exhibited strong activities during the memory guided mode (Figure 2 right). Subsequently, the premotor area was distinguished into dorsal and ventral parts. The famous mirror neurons are often observed in the ventral premotor area.

In addition, prof. Tanji's group has made significant contributions to revealing the functional differentiation of the frontal lobe through research by their high experimental skills and intelligence of Japanese macaques, including the identification of neurons encoding sequential movements in the supplementary motor area by Dr. Keisetsu Shima and discovery of the presupplementary motor area by Dr. Yoshiya Matsuzaka.

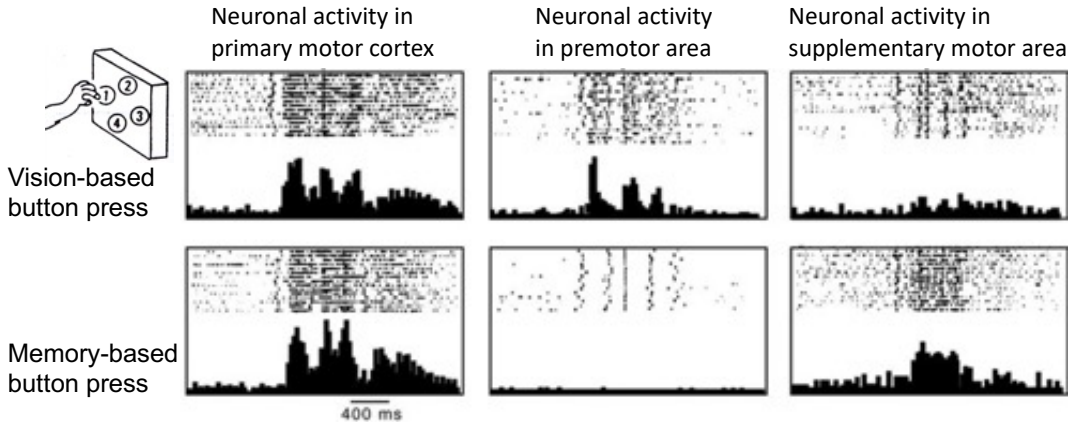


Figure 2. Sequential button press task and single neuron activities in the primary motor area (left), premotor area (middle) and supplementary motor area (right) (from Ref. 7).

Transition of Information Coded by Firing Rate

The prefrontal cortex plays a crucial role in problem solving and behavioral planning. Non-invasive measurements such as functional nuclear magnetic resonance imaging (fMRI) are not sufficient to investigate brain mechanisms in detail. It is indispensable to analyze the activity of individual neurons recorded from animal brains.

Neuronal activities related to working memory, the basis of thinking, are characteristic of the prefrontal cortex. What about neuronal activities related to behavioral planning?

The path-planning task requires the subject to move the cursor in a stepwise manner toward a final goal. The location of the final goal have to be memorized to succeed in a trial. Correspondingly, we observed a number of working memory-related neurons that varied thier firing rate depending on the final-goal location and showed sustained activities for their preferred location throughout a trial. In this respect, our results were in good agreement with previous studies on working memory.

The second major type of neurons were those whose activities appeared to reflect coming up a concrete action to achieve the final goal, an important aspect of problem solving and behavioral planning as discussed above. The neuron shown in Figure 4.6 A fired well for the Right Down final goal during the final goal display period. However, this neuron was active in the Delay 2 period when the monkey moved the cursor to the right after the Go signal was presented, as if the monkey had already decided on the specific action.

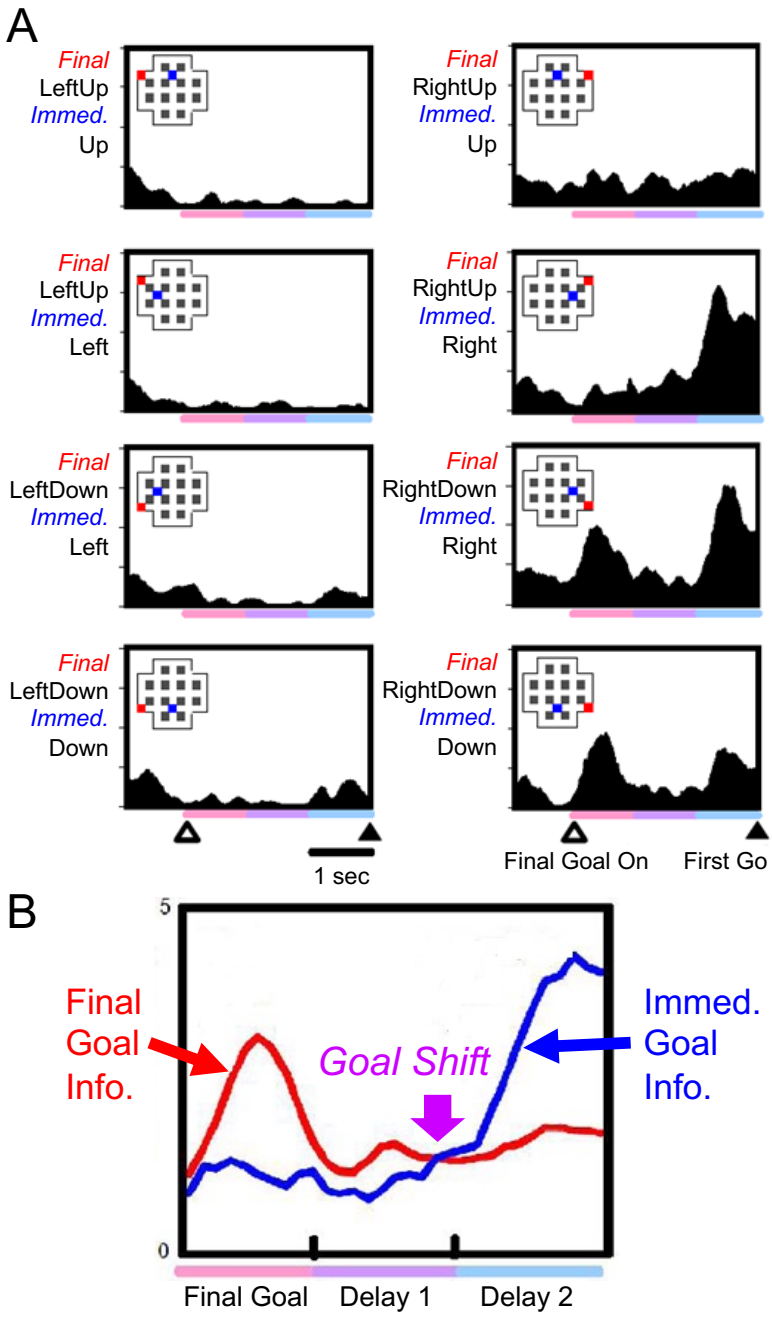


Figure 4.6. An example prefrontal neuron showing final goal – immediate goal shift. A. Firing rates for each Final and Immediate goals. B. The time course of goal information coded by the firing rate (From Ref. 3).

What the firing activity of a neuron changes in response to is technically referred to as what the firing activity is encoding. Early in the preparation period, this neuron is encoding the final goal position, while near the end of the preparation period, it is encoding the immediate goal, i.e., the direction of the first cursor movement. The temporal evolution of the degrees of goal information encoding (specifically, the temporal evolution of the normalized multiple regression coefficients) are shown in Fig. 4.6B. These plots visualize the shift of the encoded goal information, i.e., how this neuron changed the encoded goal information in response to planning the immediate goal from a given final goal. Such dynamic encoding by individual neurons is itself a new picture of modern neuroscience. Such neurons will be referred to below as final goal-immediate goal-shift neurons. Also remember how goal shift is represented in Figure 4.6B.

Finding Consistent Relationships behind Behavioral Planning

Coming up with specific ideas and actions to solve a problem or achieve a goal is a critical aspect in behavioral planning. We have found neurons reflecting this aspect called the final goal-immediate goal-shift neurons. These neurons changed their activities in response to the final goal location early in the preparatory period of the path-planning task, while, just before a Go signal, their firing were largely modulated by the direction in which the animal intended to move the cursor in the first step, referred to as the immediate goal.

The monkeys had to think out an immediate goal on each trial, although they had been well trained to perform the task with ease. If the immediate goal was to be created on the spot, I expected to find self-organizing phenomena, i.e., autonomous generation of order characteristic of complex systems in the prefrontal neural circuits and neuronal activities. In particular, I hoped for synchronization of neuronal firing.

In the first place, concrete measures and actions to solve a problem or achieve an final goal must satisfy the conditions that the problem or goal explicitly or implicitly includes. In particular, a good solution or action should satisfy those conditions properly and simply.

Let's look back at the story of the police officer who is the protagonist of the famous Japanese cartoon discussed in a previous section. He came up with a concrete action of hooking the crane to the doll's nose to achieve the final goal of taking it. Before obtaining the idea, he considered several things shown in Fig. 4.1: machine's movements and speeds differ in different manufacturers; dolls' shapes have to be considered; experienced skills are required in the sense of distance. Also in Fig. 4.7, he pointed out the importance of grabbing the doll vertically, taking into its center of gravity into account, and not worrying. The action of hooking the crane to the nose satisfies these requirements simply and consistently.

In the previous chapter, I noted that consistent relationships, in a broad sense, emerge as a synchronic order. If concrete actions or immediate goals are created as consistent relationships between explicitly and implicitly considered factors, including the final goal, there may be an increase in synchronization between prefrontal neurons contributing directly to spontaneous coming-up-with-idea processes. In the next section, we present the results of the authors' study, which shows that this is in fact the case.

Figure 4.7. Concrete actions satisfy the requirements. From Osamu Akimoto ""KochiKame: Tokyo Beat Cops"." Shonen Jump No.8 (1991).

Synchronous Firing Synchronizes with Behavioral Planning

In the final goal-immediate goal-shift neurons that we found in the monkey lateral prefrontal cortex during the path-planning task, the information encoded by the firing rate shifts from the final goal to the immediate goal (the direction of the first cursor movement). Is there any evidence that this shift is a self-organizing or emergent phenomenon of neural circuits as complex systems? In particular, is there any correlation with the synchronization phenomenon as a synchronic order that I have repeatedly pointed out?

Synchronous firing means that different neurons fire at the same time. Although here are various criteria for defining temporal coincidence, w we considered different neurons recorded simultaneously to be synchronous if they fired within a 25 ms-time window. Among the simultaneously recorded neurons, we selected neuronal pairs that fired during the action planning period and contained at least one f final goal-immediate goal-shift neuron. Simultaneous increase of the activity (firing rate) of the two neurons comprising pair inevitably increases the probability that the two neurons will fire synchronously by chance. However, what we wanted to evaluate now is not such coincidental synchronization, but synchronization as a self-organization or emergence phenomenon in complex systems. I will not go into the detail, but we have carefully eliminated the influence of firing rate from the index of synchrony and evaluated it that cannot be explained by a simultaneous increase in firing rate.

These neuron pairs showed temporal changes in final goal information, immediate goal information, and firing synchrony as shown in Figure 4.8. From the figure, it can be seen that there is an increase in synchrony around the time of the final goal-immediate goal shift. Since this increase occurs before and after block presentation in the maze, one might be concerned that the increase in synchrony might be related to it, but this was not the case. The peak time of synchronization of these neuron pairs correlated with the final goal-immediate goal shift time.

These results are only indirect evidence for the phenomenon of brain self-organization. In the absence of block information, pathways cannot be uniquely determined. In this sense, this behavioral task is an ill-posed problem. The moment of solving such a problem, i.e., the moment when the information encoded by the activity of the neurons shifts from the final goal to the immediate goal, corresponds to the moment when the idea is conceived. This coincidence of increased neural synchrony with a shift in the neural representation of goals strongly suggests that the mechanism by which plans are "created" in the brain is the emergence of a synchronous order.

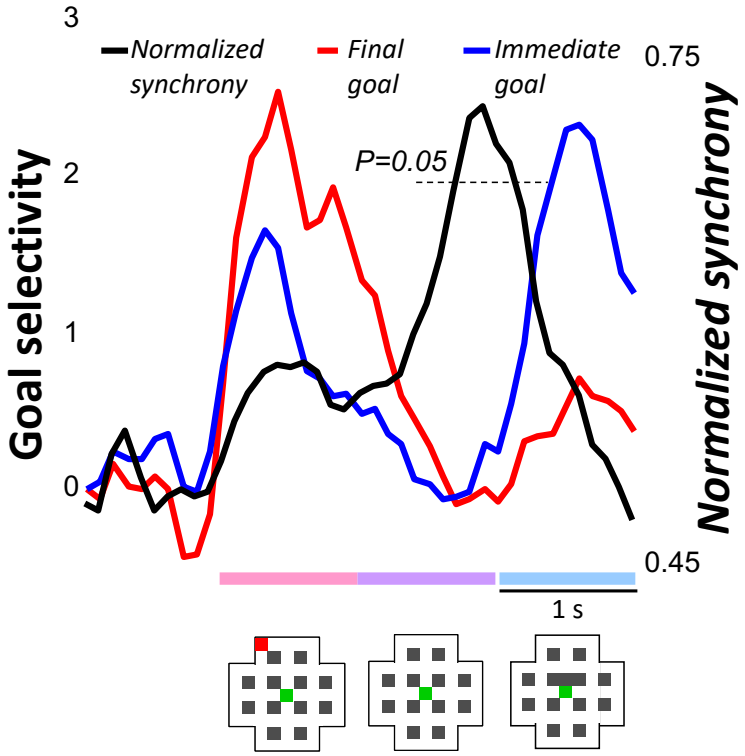


Figure 4.8. Average temporal changes in the strength of synchronous firing (black), final goal (red), and immediate goal (blue) encoded by neuron pairs containing final goal-immediate goal shift neurons during the preparatory period (from Ref. 3).

Increased Firing Variability as a Precursor to Inspiration

So far, I have described that, in many neurons of the monkey lateral prefrontal cortex during the path planning task, the information encoded in the firing rate shifts from the final goal to the immediate goal or specific action. I have also shown that this shift, which is considered to be the neural basis for coming up with an action plan, occurs simultaneously with an increase in synchronous firing between neurons, a phenomenon characteristic of complex systems.

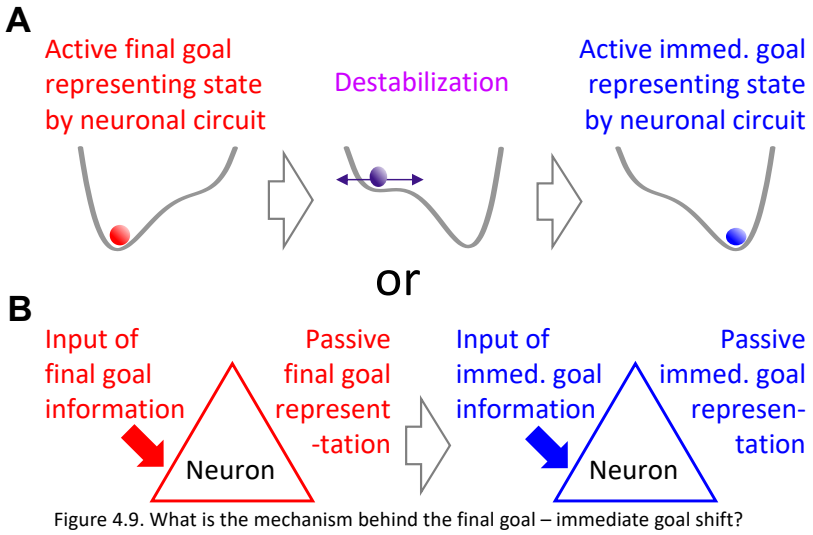


Figure 4.9. What is the mechanism behind the final goal – immediate goal shift?

Here, a further question arises. Do the neurons or neuronal networks transition by themselves from the final-goal representing state to the immediate-goal representing state (Figure 4.9A)? Or does the goal-representation shift simply reflect input change (Figure 4.9B)? That is, how can you rule out the latter possibility that those neurons fire initially in dependence on inputs carrying final goal information encoded elsewhere, and are then driven by another input about the immediate goal determined in yet another brain region? Answering these questions is crucial to clarify whether the prefrontal cortex creates immediate goals autonomously.

In Chapter 1, I argued that state transitions in complex systems can be understood as bifurcations, and that critical fluctuations, 1.e., increases in fluctuations in the measures obtained from the system, are precursors to bifurcations. Then, do the fluctuations in the neuronal activity increase before these neurons switch the information they are encoding?

Neuronal firing can be regular or irregular, even at the same firing rate (number of firings per unit time) (Figure 4.10A). We used a measure of firing irregularity that is independent of firing rate as an index of fluctuations in neuronal activity and compared the degree of irregularity before and after the final goal-to-immediate goal shift. We then found (as expected?) that firing fluctuations were increased before the shift (Figure 4.10B). Using a simple theoretical model, we also demonstrated that firing irregularity increases before the bifurcation of the neural network.

Critical fluctuations are only indirect but powerful evidence of bifurcation. Our results are the first ever observation of critical fluctuations suggesting bifurcation in higher cortical areas such as the prefrontal cortex.

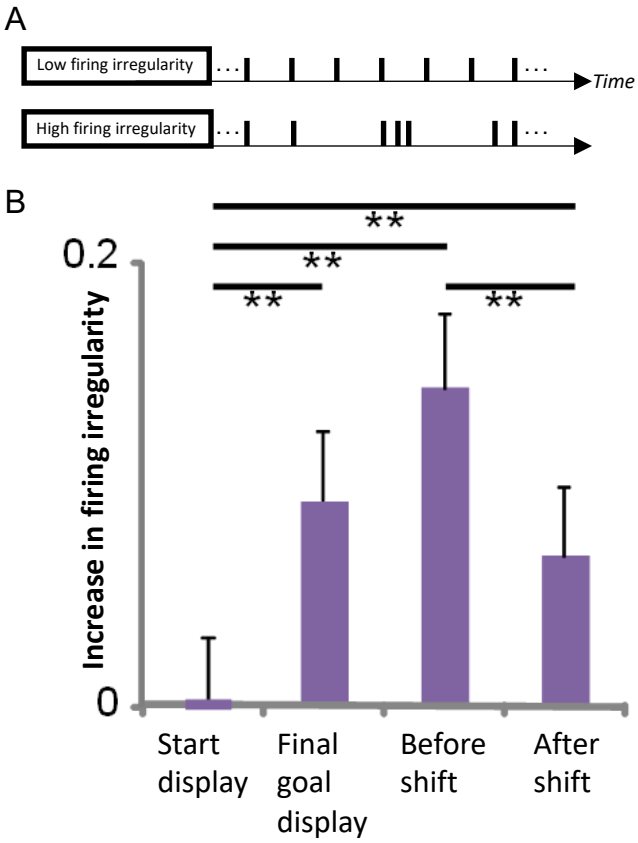


Figure 4.10. Firing irregularity increases just before the final goal-immediate goal shift. a. Schematic diagram showing the difference in firing irregularity at the same firing rate. b. Change in firing irregularity with respect to before the trial start. $p < 0.01$ for ** (from Ref. 11).

Behavioral Planning as an Emergent Phenomenon of the Prefrontal Cortex

So far, I overviewed our experimental results. By recording neural activity from the monkey lateral prefrontal cortex during a path-planning task, it was found that a number of neurons transition from the final goal to the immediate goal (first movement direction) in terms of the information encoded by the firing rate. These cells tended to show transient synchronous activity with other cells during the transition. In addition, an increase in firing fluctuation was observed immediately before the transition. Synchronization and increased fluctuation just before state transitions are phenomena unique to complex systems. The following schematic illustrates a scenario in which a neural circuit composed of final-immediate goal transition cells in the lateral prefrontal cortex generates a plan of action (Figure 4.11).

First, neuronal activities encode the final goal when it is presented. That is, the prefrontal network takes a state in which the up-and-down of the firing rate is dependent on the final goal location. This state is transient but stable.

This is followed by network destabilization. Good actions and measures are obtained when one is free from some biases, prejudices and implicit assumptions. These things do not provide good ideas from a global perspective, but they have some reasons for minds to be captured, i.e., some local stabilities, as if a ball rolling down a slope is caught not at the bottom but by a small dimple along the way. To have globally good ideas requires to avoid such small dimples or local minima. The increase in firing fluctuations we found presumably reflects the network's readiness to have a globally good state.

Next, there is a transient increase in synchronous firing. This is the moment when an immediate goal or concrete measure is created in the brain as a synchronic order that consistently and simply satisfies the factors and conditions considered.

Prof. Yuichi Katori of Future University Hakodate and prof. Kazuyuki Aihara of the University of Tokyo have created a theoretical model that explains a simplified version of this phenomenon. In this model, the network in a particular state spontaneously destabilizes and bifurcate to take another state.

The prefrontal network adapts to indefinite environments by flexibly taking various states with limited neuronal resources.

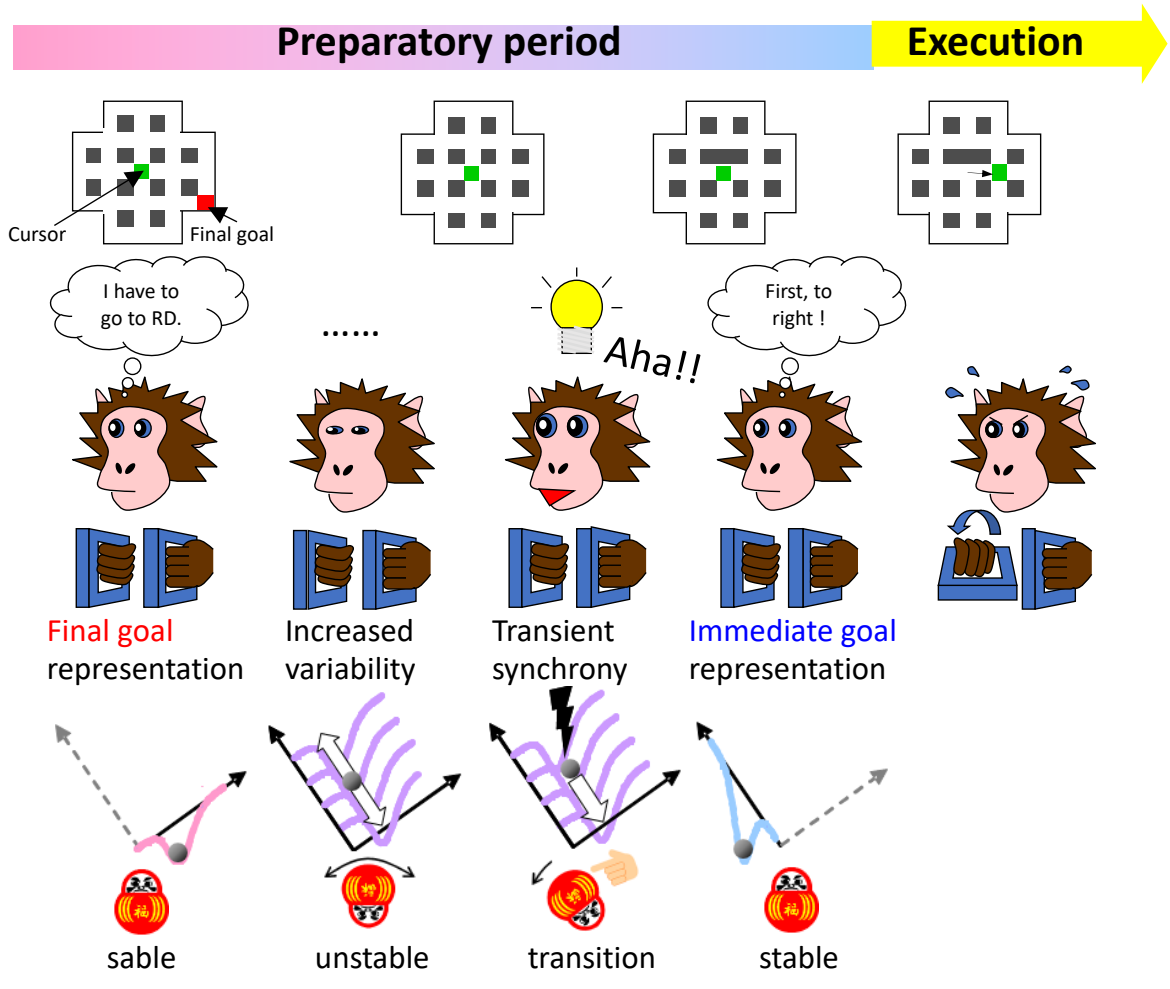


Figure 4.11. Complex system dynamics scenario in which prefrontal neural circuits generate an immediate goal from a final goal in a path planning task.

BOX *Idea Generation Based on the Nature of the Prefrontal Cortex*

Scientists live under constant pressure to come up with good ideas. Here I would like to share with you a method that I practice naturally. I learned this method from my elementary school teacher, Mr. Shigeya Ando, and I call it Nikolay Nosov's "School Boys" method.

"School Boys" is a Russian novel for children written by Nikolay Nosov (Figure A). Mr. Ando instructed us to solve math problems using the procedure described in the book. The important point is to wright down all the questions and things you notice, as illustrated in Figure B.

When you execute this process, you will naturally notice that the area of trapezoid CDFG is equal to the area of quadrangle ABFE, leading you to the correct answer of 48 cm^2.

This method is more effective when done on a simple computer text file for the following two reasons.

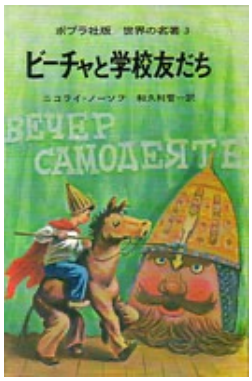
First, you can write down any sentences coming into your brain, and change their order freely. This process make it easier for you to find the logical structure, such as "Oh, these two questions are essentially the same," or "I have to solve this problem first before that one." You can escape from going in circles.

Second, when you look at the written text, you don't have to remember what comes to mind, so you seem to come up with and notice a lot more things than if you don't write them down.

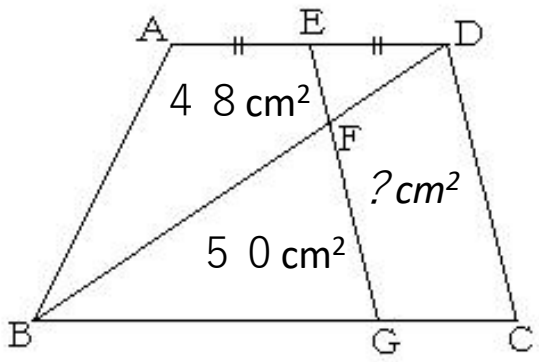
This approach is very reasonable given the function of the prefrontal cortex. Writing out reduces the load on the working memory of the prefrontal cortex. In turn, the prefrontal capacity can be allocated to other things. The saved prefrontal capacity can be used for reasoning and other tasks. Even rearranging the order of sentences by hand, rather than reasoning with the mind, reduces the load on the prefrontal cortex. In other words, using another visible "prefrontal cortex" on a text file can greatly facilitate creativity. I strongly recommend this method to my readers.

This creative thinking method reminds me of my elementary school days, along with my deep appreciation for Mr. Ando.

A



B



In the figure on the right, AD and BC, CD and GE are parallel and AE = ED, respectively. If the area of quadrangle ABFE is 48 cm^2 and the area of triangle BGF is 50 cm^2, find the area of quadrangle CDFG.

Here's what I noticed...

The area of triangle ABD is twice the area of triangle BDE, isn't it.

Ah…, the area of triangle BDE is equal to the area of triangle DEG.

And, twice the area of triangle DEG is the area of parallelogram CDEG.

Aha! Then the area of parallelogram CDEG is equal to the area of triangle ABD!

Also, parallelogram CDEG and triangle ABD have a common part triangle DEF.

Oh well!!! I don't need to know (for now) the area of triangle DEF.

Figure. The "School Boys " method. a. "School Boys" by Nikolay Nosov. b. Example of the use of this method. Partially modified from the 1987 La Salle Junior High School entrance examination.

BOX *"Anticipating" Neurons*[2)]

In the main text, I showed the final goal-immediate goal-shift neurons in which the information encoded by their firing activities shifts from the final goal direction to the direction of the first cursor movement during the preparatory period or behavioral-planning period of the task. On the other hand, we also found many neurons that simply encode the first cursor movement without coding the final goals (Figure Top). Surprisingly, there are also neurons that encode the second (Figure Middle) or third (Figure Bottom) cursor movement, respectively. Prof. Hajime Mushiake of Tohoku University, who found them, named these three types of neurons "anticipating" neurons. The simultaneous firing of these neurons during the behavioral-planning period corresponds to an important aspect of action planning in which we plan multiple steps of the action we intend to perform before execution. We do not prefer to start doing something haphazardly. We often start things with some planning. This is the first scientific result reporting neuronal activities that reflects this important aspect of action planning.

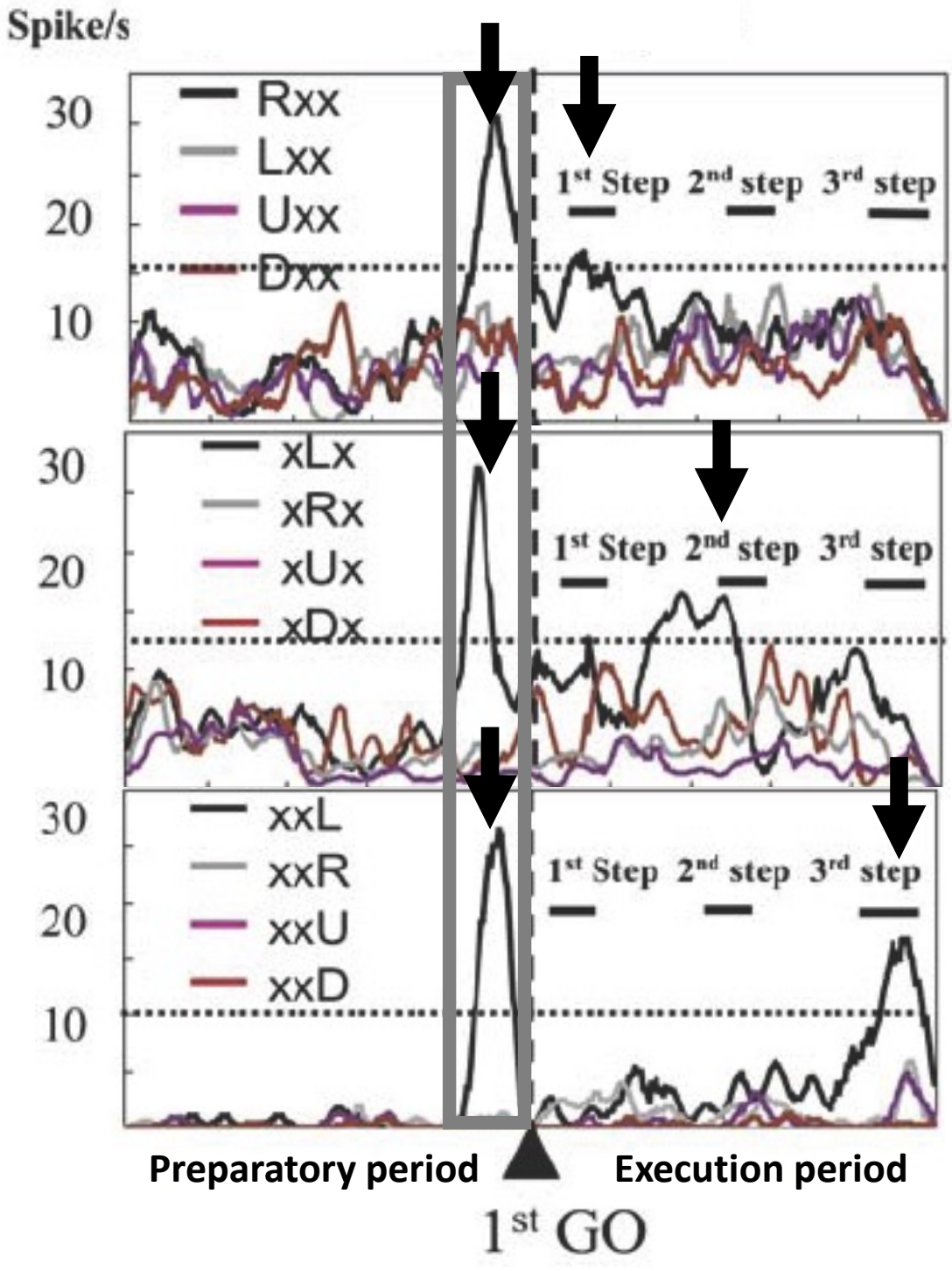


Figure. "Anticipating" neurons in the lateral prefrontal cortex.

BOX *The Tomato Robot Competition*

Industrial robots used in factories perform many sequential movements accurately, but these are pre-programmed for the production line. On the other hand, the current robots cannot do an immediate action such as “shake hand again,” so it is by no means easy to create a robot that conceives and executes sequential actions in an environment where not everything is regulated, as in a factory, i.e., in the indefinite environment described in this book. However, something must be done to develop robots that can handle such an environment. Robot competitions provide a good opportunity for developers to recognize what can and cannot be done with current technology. But in the real world, the bar is too high. Therefore, it is very realistic and fruitful to hold competitions in an intermediate environment between a real environment and a factory production line.

The “Tomato Robot Competition,” organized by an old friend of mine, professor Kazuo Ishii of Kyushu Institute of Technology, who specializes in robotics, is an excellent competition that fulfills the above conditions. The progress of agricultural technology has made remarkable progress, and state-of-the-art tomato farms look like a factories. However, even though their growth is strictly controlled, each plant has different branching and fruiting pattens. In other words, a most modern tomato farm provides a good intermediate environment between the factory production line and the real environment.

I participated as an observer in the third conference held in late 2016. The required task was to approach a plant, move the front-end (corresponding to the hand) close to the fruit, and harvest it. For now, the order of operation is fixed, so there is no need to generate it flexibly. However, I was strongly reminded of the difficulty of figure-ground separation, i.e., perceiving the fruit as a whole, even though it is occluded by leaves.

Prof. Ishii is working hard to review and improve the regulations so that the competition will be fruitful for all participants. I sincerely hope that the competition will continue and grow.

BOX *Studies That Can Be Black and White and Those That Cannot*

There are many cases in which a novel experimental method has advanced a research field. In biology, in particular, unlike physics, there are few cases where a theory has preceded an experiment (for example, the existence of gravitational waves predicted by the theory of relativity was verified using a giant observation device), and breakthroughs in experimental methods have often lead to major advances in the research field. On the other hand, however, I always ask myself whether it is truly possible to make breakthroughs that will deepen our understanding of how the brain works.

At the molecular and neuronal level, it is still possible. There are many relatively clear-cut issues, such as whether long-term potentiation of synapses is associated with increased calcium concentration in postsynaptic cells, or whether the number of dendritic spines increases with learning. If the problem is easy to make black-and-white, breakthroughs are likely to occur through new experimental techniques.

On the other hand, at scales beyond the neuronal network, what should be obvious is never clear. Even if the spatiotemporal pattern of network activity changes dramatically under certain experimental conditions, the cause or functional significance of the change is neither known nor understood. This ambiguity stems from the complexity and degrees of freedom of the experimental subject. Even if we were to record the activity of every neuron in the brain, what could be extracted from the data would remain an open question.

When interpreting data obtained from objects with a high degree of freedom, room for "beliefs" comes into play. Many of my experimental results presented in this book are indirect evidence, and it is undeniable that my interpretation of the data reflects my "beliefs." This is partly due to the author's competence, but also because of a fundamental problem in brain research. So I think I need to refine my "beliefs" about how the brain should be.

References

1) Mushiake H, Saito N, Sakamoto K, Sato Y, Tanji J. Visually based path planning by Japanese monkeys. *Cogn. Brain Res.*, 11:165-169 (2001)
2) Mushiake H, Saito N, Sakamoto K, Itomaya Y, Tanji J. Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans. *Neuron*, 50:631–641 (2006)
3) Sakamoto K, Mushiake H, Saito N, Aihara K, Yano M, Tanji J. Discharge synchrony during the transition of behavioral-goal representations encoded by discharge rates of prefrontal neurons. *Cereb. Cortex*, 18:2036-2045 (2008)
4) Squire L et al. *Fundamental neuroscience (2nd)*. Academic Press, New York (2002)
5) Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding visual space the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.*, 61:331-349 (1989)
6) Tanji J. *Nou to Undou – Action wo Jikkou Saseru Nou*. Kyoritsu, Tokyo (1999) in Japanese
7) Mushiake H, Inase M, Tanji J. Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during visually guided and internally determined sequential movements. *J. Neurophysiol.*, 66:705–718 (1991)
8) Tanji J, Shima K. Role for supplementary motor area cells in planning several movements ahead. *Nature*, 371:413–416 (1994)
9) Shima K, Tanji J. Role for cingulate motor area cells in voluntary movement selection based on reward. *Science*, 282:1335-1338 (1998)
10) Mastuzaka Y, Aizawa H, Tanji J. A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: neuronal activity during a learned motor task. *J. Neurophysiol.*, 68:653–662 (1992)
11) Sakamoto K, Katori Y, Saito N, Yoshida S, Aihara K, Mushiake H. Increased firing irregularity as an emergent property of neural-state transition in monkey prefrontal cortex, *PlosONE*, 8: e80906 (2013)
12) Katori Y, Sakamoto K, Saito N, Tanji J, Mushiake H, Aihara K. Representational switching by dynamical reorganization of attractor structure in a network model of the prefrontal cortex, *PloS Comput. Biol.*, 7: e1002266 (2011)

Part 2. Seeking the Principles of Creativity

Chapter 5　Solving Problems with Assumptions - Theory of Brain Computation and Constraints

In Part I, we looked at the correspondence between brain function and complex systems phenomena. I hope that you now understand that behind situations that require creativity (in a broad sense), there exist complex emergent phenomena in the brain and nervous system. However, the correspondence between brain function and complex phenomena does not mean that brain creativity has been clarified. In Part II, we will organize our thought on what we need to consider in order to get close to the principles and mechanisms of creativity.

Some readers may say, “The phenomenon of complex systems emergence may indeed be the basis of the brain's creativity. It would be impossible to elucidate a very complex system like the brain only by experiments. So why don't we just build a theoretical model of complex systems based on experimental results?" Some may think, "Yes, that would be a good idea. In fact, there is a lot of theoretical model research going on, ranging from relatively abstract models that combine a large number of nonlinear oscillators to quite realistic neural circuit models such as the Hodgkin-Huxley model, which is constructed by considering the anatomical wiring of the neuron model. Some of them show quite interesting behavior. However, frankly speaking, there are many brain/neural models that merely exhibit interesting spatiotemporal patterns or, worse, interesting "properties" as complex systems.

This chapter delves into the problems with such approaches and the vague questions and concerns about them.

I. Breaking the Problem Down into Several Levels for Understanding the Brain

Three levels of David Marr

Many researchers, including myself, are examining in detail the conditions under which certain brain regions and neurons are activated or deactivated in order to understand how the brain works. But is there anything wrong with focusing only on examining the details?

One might think that this is because that alone does not tell us how the neurons are wired together to form a neural circuit.

If so, we should wire up elements like neurons and run model simulations, whether real or abstract, based on the experimental results.

Neurons are complex and diverse, and no perfect model exists. Even if we refer to anatomy, we would be doing our best to mimic wiring trends. However, in complex systems, not just neural circuits, small differences often make a big difference in behavior.

David Marr, a theoretical neuroscientist who died in 1980 at the age of 35, asserts that it is impossible to understand the entire information processing and computation in the brain from a single perspective or equation. He further emphasizes that we must not confuse the issues, that is, we must distinguish between the question of what should be processed and the question of how it should be processed. We then present the following three levels of issues/perspectives that are necessary (in slightly different words) to understand a machine (in this case, including the brain) that performs some kind of information processing or computation.

Levels of Theory of Computation (first level): What is the purpose of the computation? What should be computed?

The level of algorithms and representation (second level): How is the objective realized? How is the information represented and processed?

Level of hardware realization (third level): How are the algorithms and representations physically realized?

Distinguishing between these three levels of problems leads us in the direction of understanding the entire information processing and computation in the brain, which can be reproduced in robots and other devices. When the brain's information processing cannot be successfully reproduced, Marr's three levels bring clues to the solution to the problem, whether the purpose of the computation is wrong, the algorithm or information is poorly represented, or hardware limitations prevent it from working.

The First level. The Purpose of the Computation

What is the purpose of specific information processing by the brain, or the first level of understanding the system that carries out the information processing? For example, to see a color is to see a color, to see a three-dimensional object is to see a three-dimensional object, what more is there to say?

If we look back carefully at the brain's information processing, it is not always obvious what exactly is the purpose of even the information processing in our daily lives. An example of this can be seen in the surface color perception experiments conducted by Edwin H. Land.

Many people may think that seeing the color of an object's surface means detecting the wavelength of light projected from that object onto the retina. Through a sophisticated experiment, Land, founder of the Polaroid Company (famous for its instant cameras), showed that this is not the case, that is, the color we see on the surface of an object is based on the surface reflectance of the three primary colors of light, red, blue, and green, or the degree to which each wavelength is reflected from the surface.

There is an abstract painting called the Mondrian figure, in which various colors are arranged in patches (Figure 5.1). Land asked subjects to focus on a particular patch and judge its color. The figures were illuminated with light of red, blue, and green wavelengths. The intensity of each light was adjusted so that the intensity of the light reflected from the patch the subject was asked to focus on is kept constant across the three wavelengths of light.

Surprisingly, the subjects judged the red portion of the Mondrian figure to be "red" and the blue portion to be "blue," even though the intensity of each red, blue, and green reaching the eye from the patch of interest was physically constant. This property of color vision is called color constancy.

So what is it that our color vision "sees"? Land's experiment also answered this question. For example, suppose that the intensity of the irradiated red was 100 and the intensity of the red reaching the eye was 50. In this case, we can calculate the red reflectance of the patch to be 0.5. When this estimate is made for each experimental condition, it becomes clear that what we see as color is the reflectance of red, blue, and green on the surface of the object. In other words, the purpose of surface color calculation is to calculate the reflectance of light on the object surface.

The wavelength of light in our daily environment varies greatly with time of day, weather, and other factors. Despite such variations, calculating the surface reflectance, which is an inherent property of objects, is useful for life.

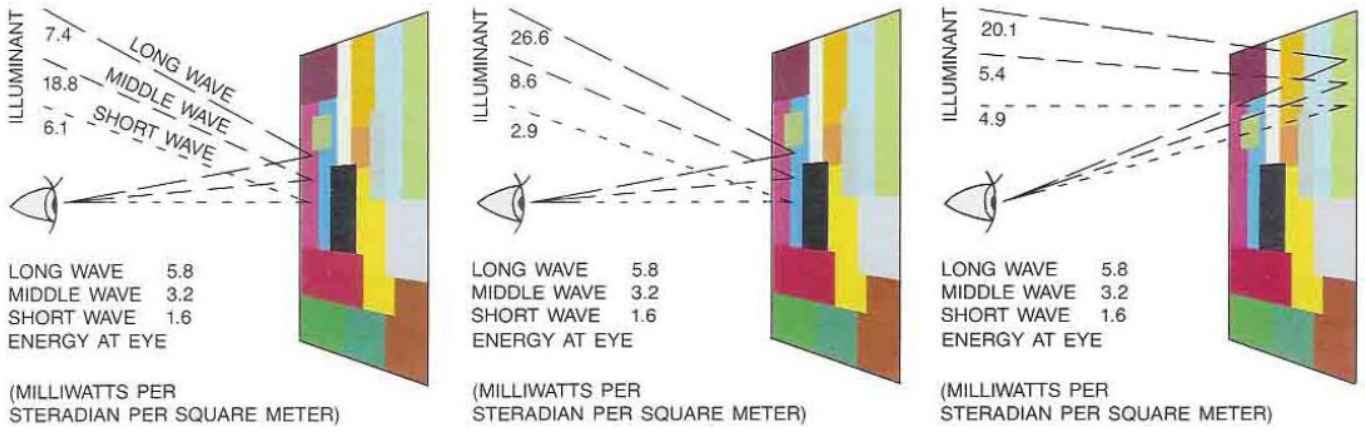


Figure 5.1. Land's experiment on color constancy using Mondrian figures (from Ref. 2).

The Second Level. Algorithms and Representations

The second level of understanding of how the brain works, the question of how to realize the purpose of information processing, that is, how to represent and process information, will also be addressed using the surface color problem as an example.

The purpose of the surface color calculation is to calculate the reflectance (ratio of reflected light intensity to irradiated light intensity) of red, blue, and green on the object surface, as indicated by Land. As for the representation, the three primary color representations of red, blue, and green are fine. The problem is processing. Only wavelengths that enter the eye are available. How can we calculate the reflectance of a particular surface? Land proposed the idea that if we compare the intensity between two neighboring points, we would know the surface reflectance ratio of the two points, even if we do not know the surface reflectance.

The algorithm is illustrated as follows (Figure 5.2). (1) Select one of the wavelengths (red, blue, or green) in that order. The process of the selected wavelengths is as follows. (2) Scan two points on the image step by step. (3) Calculate the intensity ratio of the two points. If the intensity ratio is near 1 (for example, within ±0.05), it is assumed to be 1; if it is outside the range, the intensity ratio of (3) is adopted. (5) Find the serial products of the intensity ratios along the scan (i.e., multiply them all through). (6) Divide the value of all the cascade products by the maximum value of the cascade products (normalization). (7) Perform the above for other wavelengths.

The result of the above calculation is a set of three simultaneous products for a point, which is considered to be the surface color of that point.

Several implicit assumptions are made for this algorithm. First, if the surface area at all three wavelengths exhibits a cascade product of 1, then that area is considered white. Another assumption is that illumination intensity does not change rapidly in space. Algorithm (4) is used for this purpose and helps eliminate the effect of gentle gradients in illumination intensity and obtain a stable surface color.

This algorithm has some aspects that do not match human surface color perception; for example, there are shadows or spot-lit areas in the middle of the image. However, even if you know the purpose of the calculation (in this case, the surface reflectance ratios of red, blue, and green), you understand how to calculate it is a different problem, and one that is required to be puzzled. It is also extremely important to note that some plausible assumptions must be made. This point will become the most important issue later in this book.

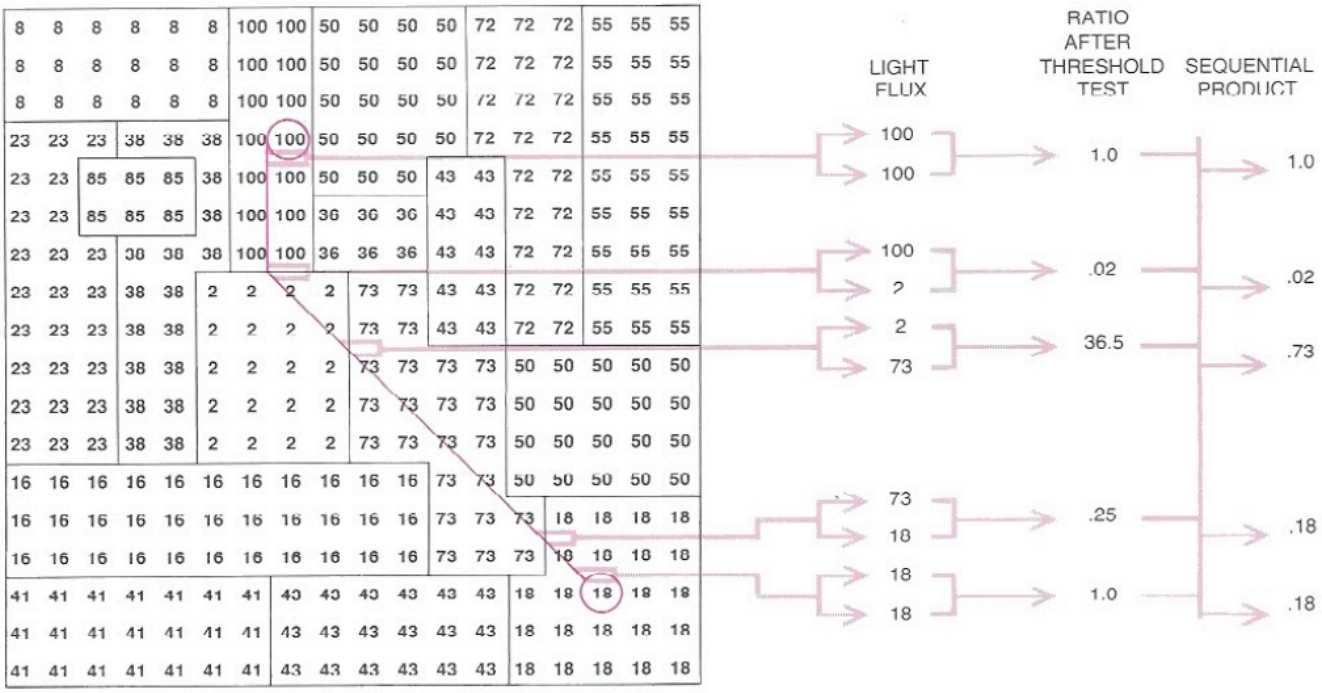


Figure 5.2. Overview of Land's algorithm for calculating the surface color. The surface reflectance ratio is obtained by comparing the luminance between two adjacent points. It is then integrated along the scan path and normalized by the maximum value (from Ref. 2).

The Third Level. Hardware Implementation

Even if there is an algorithm that can achieve certain information processing, its implementation in hardware is not uniquely determined. It is reasonable to consider that the way it is implemented in semiconductors and the way it is implemented in the brain would be different owing to differences in the materials used.

In the example of surface color calculation, the algorithm that compares the light intensity between two adjacent points is valid for inferring the surface reflectance ratios of red, blue, and green. If this algorithm were to be implemented as a computer program, step-by-step vertical or horizontal scanning would be appropriate, as Land contemplated. In contrast, the way the brain implements this algorithm appears to be different. We are not scanning the scene sequentially to recognize surface colors. The brain must be doing simultaneous, or parallel, processing in two dimensions.

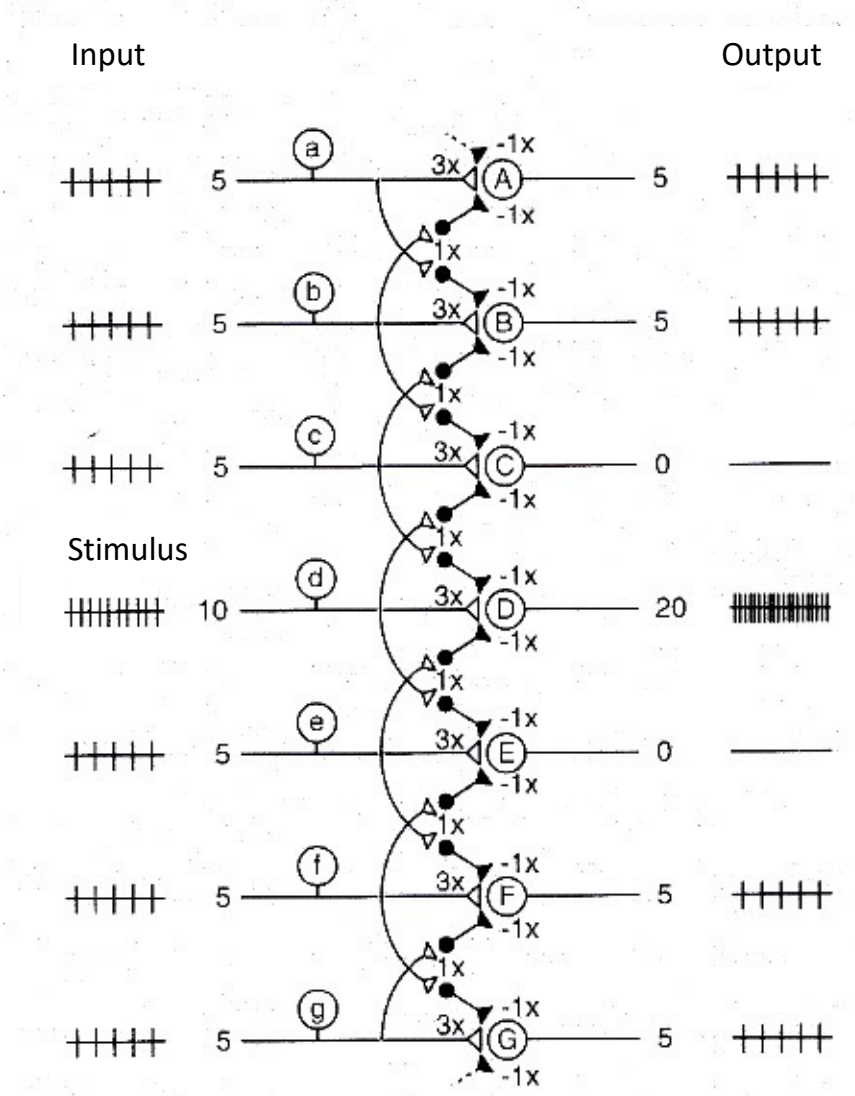


Figure 5.3. Lateral inhibition emphasizes spatial variations in input intensity.

B. K. P. Horn's model realized the key processing of surface color computation in a more or less brain-like manner[3]. He used a structure called "lateral inhibition," which is widely employed by the brain to achieve contrast enhancement. Lateral inhibition is a phenomenon in which when one neuron is stimulated, the stimulated neuron produces an excitatory response while simultaneously inhibiting the neighboring neuron to prevent it from producing an excitatory response (Figure 5.3). As a result, small changes in light intensity are suppressed, and large changes are emphasized. In other words, gradual illuminance gradients in the image can be eliminated (Figure 5.4).

Lateral inhibition was discovered by Nobel Prize winner H. K. Hartline in his research using the eyes of horseshoe crabs (!). Horseshoe crabs have simple eyes and are easy to study. Material for the right experiment. This is a typical example of this.

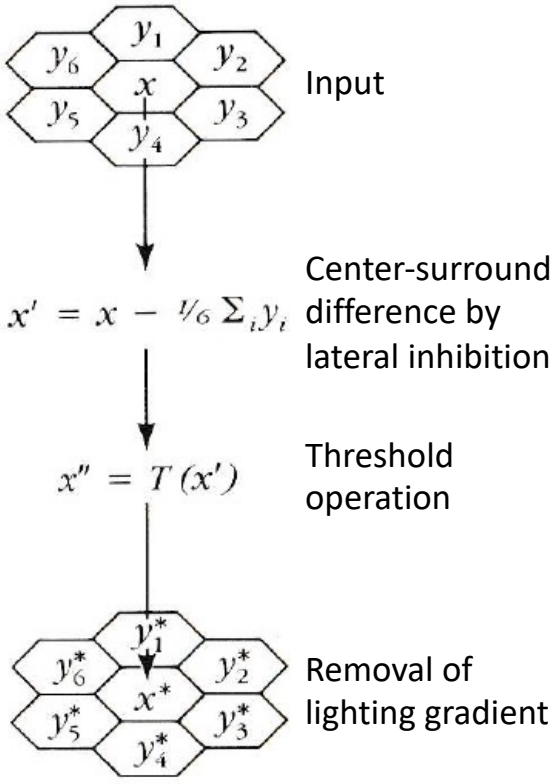


Figure 5.4. Horn's model using lateral inhibition.

Understanding the Brain at Different Levels is the Essence of the Brain

Understanding how the brain functions cannot be accomplished at a single level. I believe that David Marr's assertion continues to ring alarm bells for brain researchers today[1)]. This assertion is tantamount to stating that we will never understand how the brain works if experimentalists are just obsessed with figuring out neurotransmitters and ion channels, or if theorists are happy that they wired up model neurons and found interesting patterns. Significant research is the only significant step forward. No single discovery or invention can reveal everything about the brain.

I believe Ma's argument is persuasive because it reveals how the brain works and how it differs from today's computers and other machines. Today, machines do not have the ability to tell you that something is wrong. Even if it malfunctions, the machine itself does nothing. For example, in an arithmetic problem such as "If you buy something for 50 yen and pay 100 yen, how much is the change?" even if the machine gives an incorrect result of 150 yen, humans will notice that something is wrong because they are concurrently processing another perspective and level of processing such as "There is no way that the amount of change will be more than the money I paid in the first place." By establishing a good and consistent relationship between the different levels of processing, we attempt to find the essence behind it. This is where we see the most important aspect of how the brain works, which current machines are unable to do.

BOX *Neurons in Cortical Area V4 Respond to Surface Color*

The surface color problem is a good example of studies that have three levels of understanding of the brain: the purpose of processing, the algorithm, and the hardware realization. S. Zeki experiment made the surface color problem more complete by adding physiological results.

Zeki and his colleagues recorded neuronal activities from each of two areas of the monkey's cerebral cortex called primary visual cortex (V1 area) and V4 area; V1 is the first area of visual information entering the cortex and V4 is a downstream area receiving direct and indirect input from V1 area. The apparatus and stimuli were the same as those used in Land's psychological experiment. In other words, a Mondrian figure consisting of patches of various colors was illuminated with red, blue, and green light, the components of each light reflected from a patch were kept constant, and the light was used to stimulate cells in the V1 and V4 cortices. The results showed that cells in the V1 cortex changed their response only to the wavelength of light from the patch, while the activities of V4 neurons changed in response to the surface color, regardless of the wavelength. They revealed that color constancy is calculated between the V1 and V4 areas.

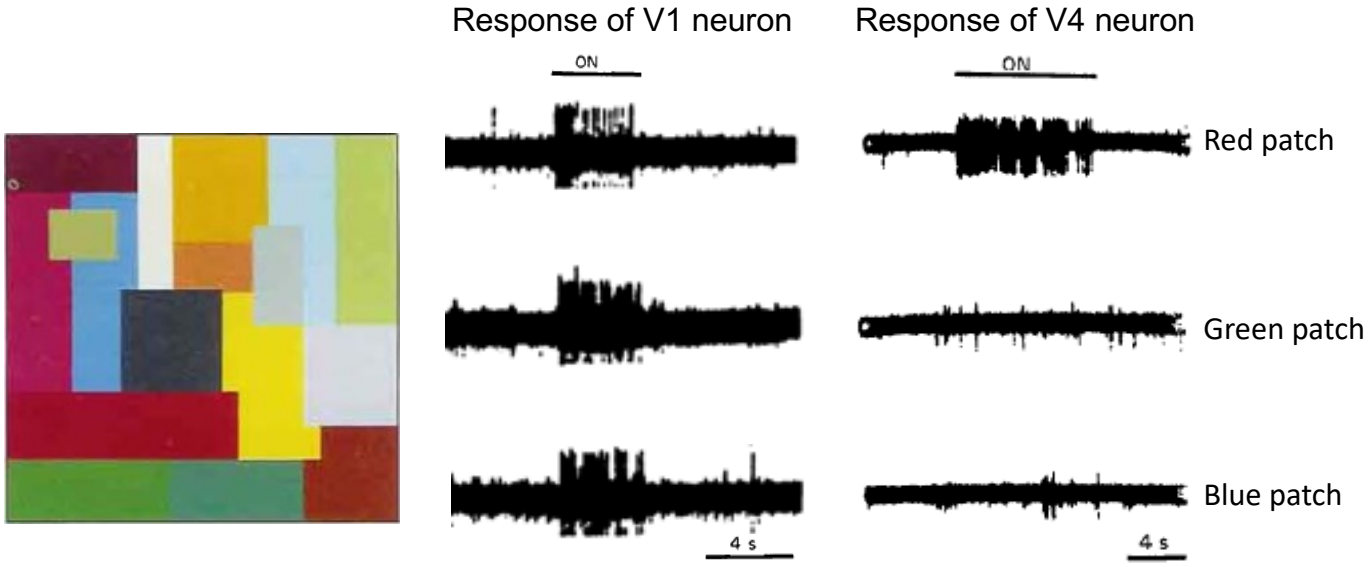


Figure. Zeki's experiment. (Left) Mondrian figure. (Middle) A cell in the V1 area. Stimulation with the wavelength most preferred by the cell. (Right) Cell in the V4. The response changed depending on the color of the patch, even though the stimulated wavelength was constant (from Ref. 4).

II. Constraints Necessary to Solve the Problem

Compute 3D Images from 2D

In many situations in daily life, processing for recognition and motion lacks the information needed to obtain a single computational result. Problems for which a single answer cannot be obtained only from the given information and clues are called ill-posed problems. Although, at first glance, it may seem difficult, we can easily come up with an answer to this problem. The most familiar example is binocular stereopsis, or the problem of obtaining three-dimensional perception from a two-dimensional image, that is, the retinal image.

Many people may think, “No, I have two eyes. Isn't that how you see three dimensions? What's so difficult about that?" or, "I can see perspective quite well with one eye.” Certainly, the retinal image of one eye obtained from the environment contains many clues for three-dimensional perception, such as the so-called detail perspective, in which nearby objects appear larger and distant objects appear smaller.

One of the figures in which the three-dimensional image can be seen only when all such monocular cues are eliminated and the left and right images are integrated in some way is called the famous random dot stereogram by Julesz (Fig. 5.5)[5]. Look at the figure on the right with the right eye and the figure on the left with the left eye. Do you see the figure pop out?

It can be said that the three-dimensional image we perceive is actually created by the brain. For an image to appear in 3D, there must be a strong correlation between what our right and left eyes see. However, there is no clear indication of where these two images should correspond and any correspondence is theoretically possible. Because there is no single correct answer, the problem of solving random dot stereograms is considered an ill-posed problem.

However, the fact that anyone can see a popped-out figure indicates that our brains are equipped with rules on how to correlate binocular images. This is a plausible and implicit assumption. This plausible implicit assumption is called constraint. Constraints are generally required to solve ill-posed problems.

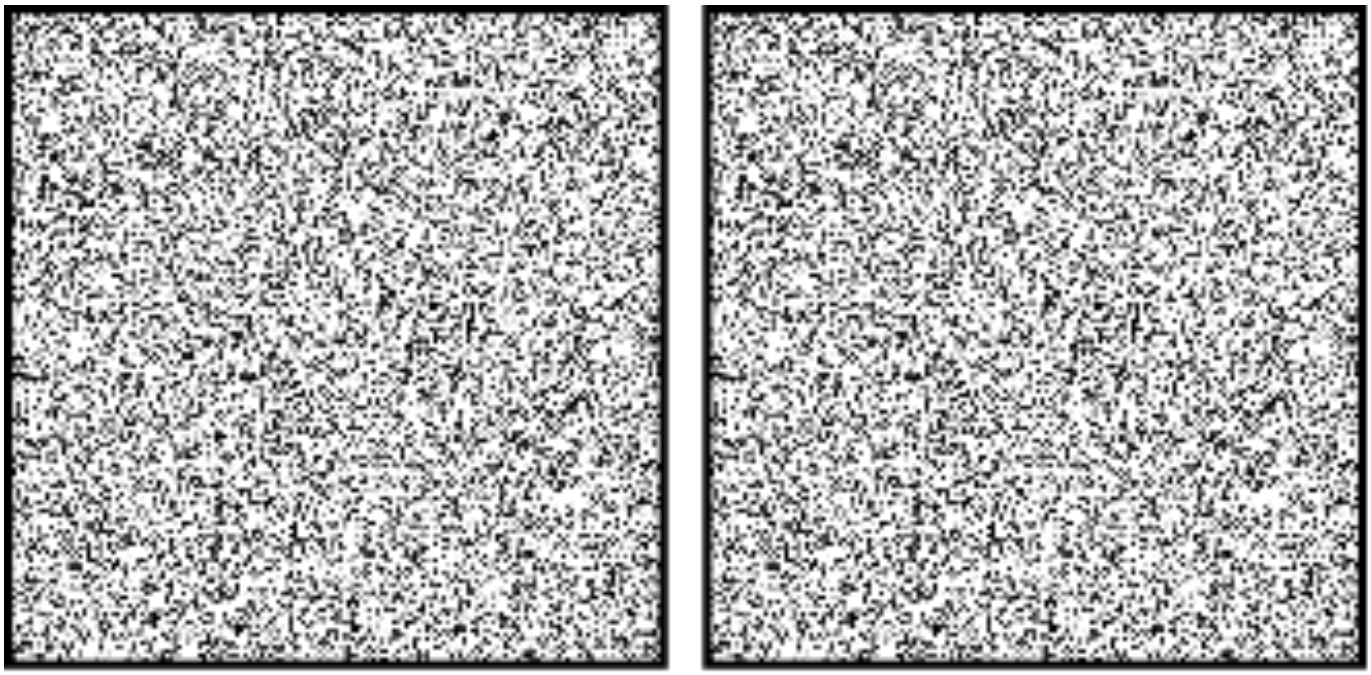


Figure 5.5. Random dot stereogram. There are no three-dimensional cues in any of the figures. The figure appears only when each figure is viewed with different eyes.

BOX *Familiar Ill-Posed Problems*

There are many ill-posed problems around us, that is, problems for which the information and conditions given by the environment are not enough to determine a single answer to the calculation. Such problems are not limited to visual processing.

Even a simple arm extension to a certain target is an ill-posed problem (Figure). The environment does not uniquely determine the arm trajectory to the target. In many cases, it is arbitrary. Even if the trajectory is determined, there still remains much arbitrariness in joint angles. Furthermore, even if the joint angles are determined, there still remains much arbitrariness in the muscle force. This problem is known as the Bernstein problem.

The path planning task that I did was also an ill-posed problem in the sense that the cursor path was arbitrary.

Bats flying around at night recognize the structure of three-dimensional space and the location and shape of their prey by listening to the bounce of ultrasonic waves they emit. In the sense that they are calculating this from the sounds coming into their only two ears, bats also solve this ill-posed problem.

How about looking for ill-posed problems around you?

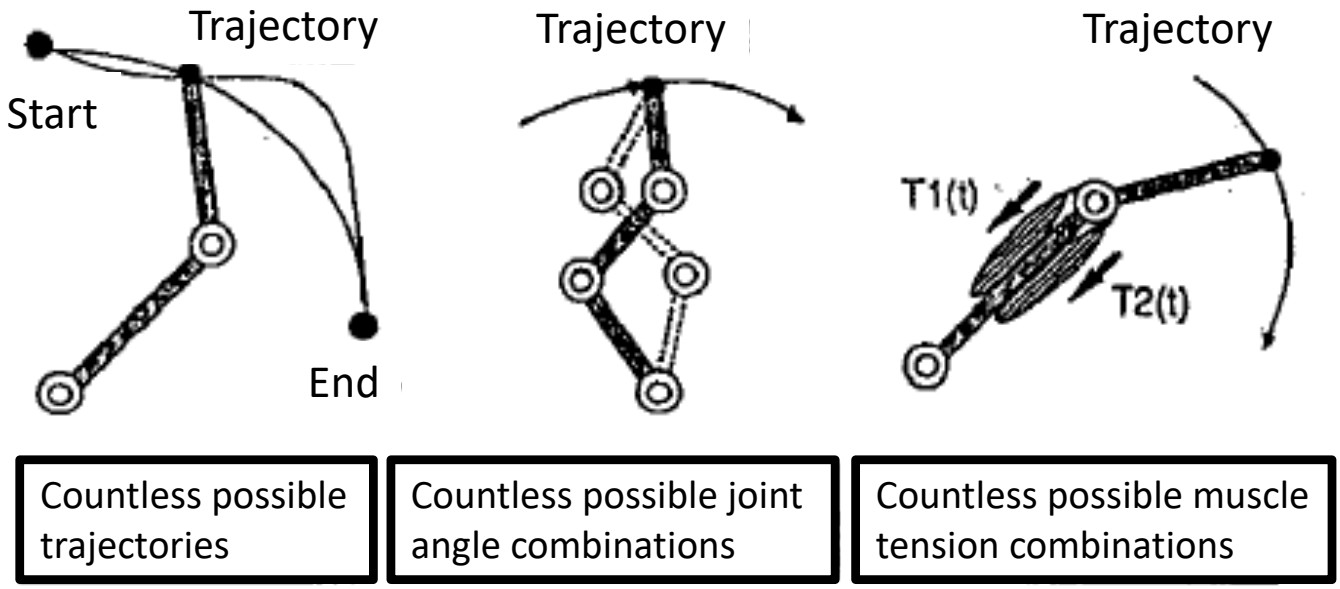


Figure. Even a simple "arm extension" movement includes ill-posed problems (from Ref. 6).

BOX *Understand Culture as an Ill-Posed Problem*

Many people may think that culture is a very humanistic concept, elusive, and far from being scientifically elucidated. I was one of them. However, if we understand that many of the problems that organisms, especially humans, face on a daily basis are ill-posed problems, that is, problems that cannot be answered or concluded with a single answer or conclusion based solely on information or clues given from the environments, we can consider why the brain needs culture, an intangible thing, from a brain science perspective as well.

As a slightly oversimplified example, consider the situation in which the values of X and Y must be determined by the equation X + Y = 1. This is another ill-posed problem in that we are not given another equation to determine one solution. If you come from an affluent rural area and have been taught since childhood that all people are equal, you might think X = Y = 0.5. On the other hand, if you were born in a cruel old mining area and have been told by your scary older brother that "people live or die," you might think that X = 1 and Y = 0, or X = 0 and Y = 1.

Many people may say that one of the pleasures of the World Cup in soccer is getting the taste of the national character that a team exudes. In the soccer field, where the situation changes from moment to moment, there is a mountain of ill-posed problems, from decision making to motor control. Naturally, there are areas that cannot be filled by the tactics of the coach or the players' individual experiences. There will also be room for the national character and culture of the country to creep in quietly.

If culture is viewed as one of the tacit assumptions, unconscious rules, and constraints necessary for human beings to solve problems, it may open the way for a scientific understanding of culture.

Smoothness Constraint

Solving ill-posed problems in which the answer cannot be determined by the given information and clues alone requires plausible assumptions and constraints. Obtaining a three-dimensional image by looking at a figure called a random-dot stereogram is a clear example. To obtain a stereogram, the right eye image must correspond appropriately to the left eye image, but how they should correspond is not given. Viewers can only make plausible assumptions and constraints.

D. Marr et al. proposed an algorithm to obtain a stereoscopic image from a random-dot stereogram using the smooth constraint condition (and the one-to-one correspondence constraint condition)[8]. To simplify the discussion, let us consider a one-dimensional random-dot stereogram (Figure 5.6, top). There are many possible ways to correspond a point in the right eye image to a point in the right eye image. For example, if they are mapped as in Figure 5.6 middle, the resulting stereo surface is bumpy. Certainly, there are many bumpy things around us, but in most cases, within a certain area, the surface is almost constant depth, i.e., smooth. In addition to the one-to-one correspondence constraint that the left and right points correspond one-to-one, Marr et al. solved the ill-posed problem in the random-dot stereogram using a smooth constraint, i.e., the depth variation between two adjacent points in the obtained 3D image should be as small as possible, an assumption plausible from everyday experience (Fig. 5.6, bottom).

D. Marr et al. proposed an algorithm to obtain stereoscopic images from random-dot stereograms using smooth constraints (and one-to-one correspondence constraints)[8]. To simplify the discussion, we considered a one-dimensional random-dot stereogram (Figure 5.6, top). There are many possible ways to map a dot in the right-eye image to a dot in the right-eye image. For example, if we map them as shown in Figure 5.6 middle, the stereo surface will be bumpy. Certainly, there are many bumpy things around us, but in many cases, within a certain range, the surface is of almost constant depth, i.e., smooth. In addition to the one-to-one correspondence constraint in which the left and right points correspond one-to-one, Marr et al. solved the ill-posed problem in the random-dot stereogram using a smooth constraint, that is, the depth variation between two adjacent points in the obtained 3D image should be as small as possible, an assumption plausible from everyday experience (Fig. 5.6, bottom).

It is not logically possible to derive why these constraints are correct, but they are not bad assumptions in light of everyday experience. In addition, the three-dimensional image obtained was simple. From this point of view, these constraints seemed plausible. This simplicity is an important property of plausible assumptions and implicit premises, which are not taught by anyone.

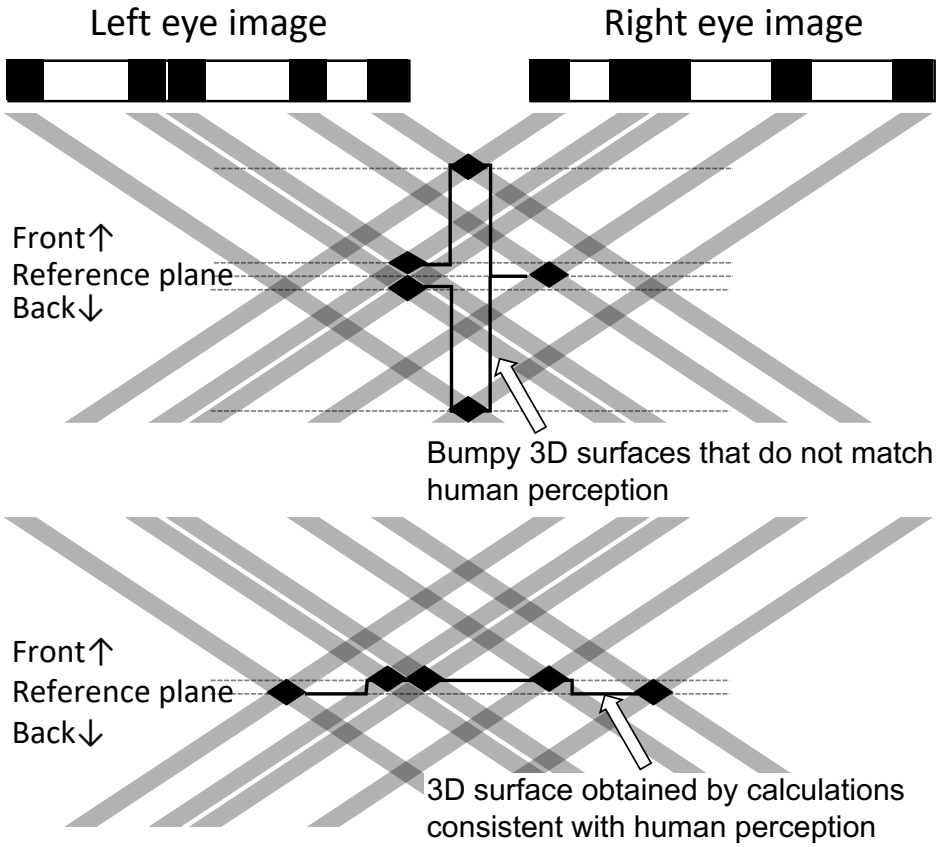


Figure 5.6. Overview of the algorithm for obtaining a 3D image from a random dot stereogram. Top: A simple example of a 1D random-dot stereogram. Middle: A bumpy 3D interpretation with one dot at the front and one at the back is also possible. The correspondence between the left and right sides is indicated by the black diamonds. Bottom: Correspondence based on Marr's smoothness constraint. The correspondence is made such that the change in the depth of the adjacent points is minimized. Consequently, a plane is formed in front of the reference plane, which is consistent with human perception.

Emergence and Ill-Posed Problems

This book attempts to view the creative aspect of the brain as a fusion of neuroscience and complex systems theory. How does the emergent phenomenon of complex systems relate to solving ill-posed problems, that is, problems for which the answer cannot be determined only by the given conditions? Dr. Naoyuki Sato and Dr. Masafumi Yano's model for solving the ill-posed problem in random dot stereograms (hereafter referred to as "RS") is a good example[9].

Figure 5.7. Random dot stereogram with superimposed surfaces.

What does the RS in Figure 5.7 look like? It is likely that a surface is composed of dots and another surface shows through beyond that surface. This is referred to as an RS with superimposed surfaces. This figure can only be answered by D. Marr's famous algorithm as "it looks bumpy". Solving an ill-posed problem requires implicit assumptions or constraints. Marr's algorithm cannot reproduce the appearance of the RS with superimposed surfaces because it uses the smoothness constraint and one-to-one correspondence constraint between the points of the right- and left-eye images in a single processing mechanism. Sato and Yano separated these two processes into separate mechanisms, with the latter utilizing nonlinear oscillator entrainment unique to complex systems (Figure 5.8).

Many oscillations in the real world, such as the beating of the heart and resonance of machinery, are nonlinear oscillations. The units that oscillate are called nonlinear oscillators (hereafter referred to as oscillators). Interacting oscillators are known to synchronize spontaneously under certain conditions, a phenomenon called entrainment. Using oscillators that mimic the behavior of neurons, Sato used oscillator entrainment to generate a point-to-point correspondence between images of the right and left eye.

One oscillator is placed at each point in the left and right images, and when the oscillators of the point of the right eye image and the left eye image are synchronized, an output is sent to the processing layer responsible for the depth plane corresponding to the disparity between those two points. This output strengthens the adjacent points on the same depth plane. It also strengthens the synchrony of the left-right pair of oscillators.

The use of nonlinear oscillators at the hardware realization level (Marr's third level) has led to success. In particular, nonlinear oscillators enable flexible correspondence without being caught by local minima.

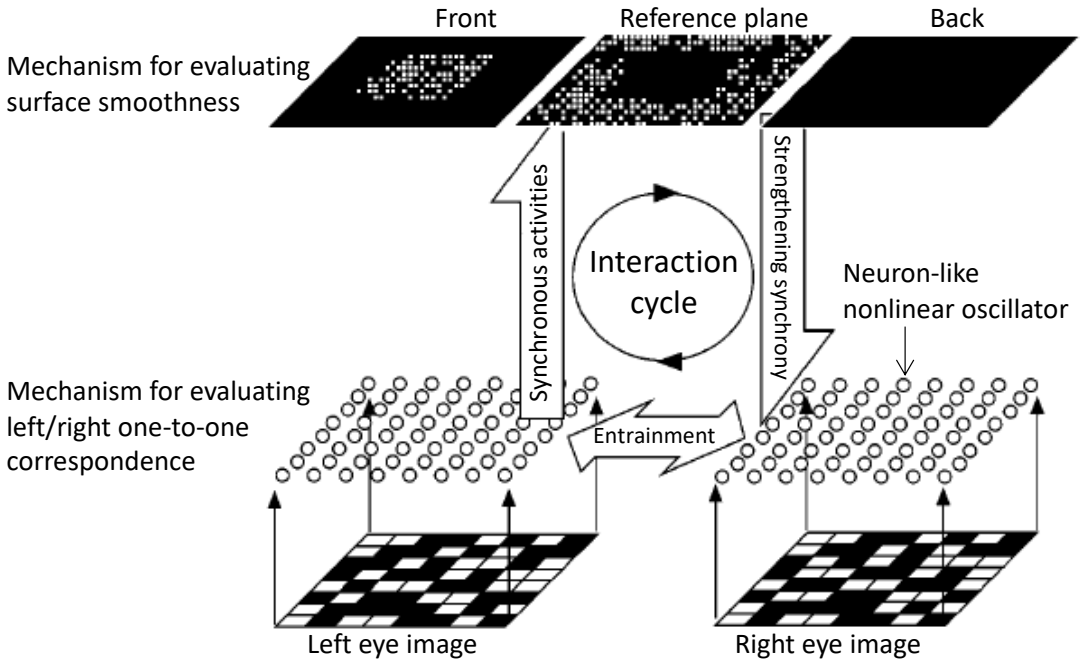


Figure 5.8. The Sato model with nonlinear oscillators (Ref. 9).

References

1) Marr D. *Vision*. WH Freeman, New York (1982)
2) Land EH. The retinex theory of color vision. *Sci. Am.*, 237:108-128 (1977)
3) Horn BKP. Determining lightness from an image. *Comput. Graphics Image Processing*, 3:277-299 (1974)
4) Zeki S. Color coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colors. *Neurosci.*, 9:741-765 (1983)
5) Julesz B. Binocular depth perception of computer generated patterns. *Bell Syst. Tech. J.*, 39:1125-1162 (1960)
6) Kawato M. *Nou no Keisan Riron*. Sangyo Tosho, Tokyo (1996) in Japanese
7) Matsuo M, Tani J, Yano M. A model of echolocation of multiple targets in 3D space from a single emission. *J. Acoust. Soc. Am.*, 110:607-624 (2001)
8) Marr D, Poggio T. Cooperative computation of stereo disparity. *Science*, 194:283-287 (1976)
9) Sato N, Yano M. A model of binocular stereopsis including a global consistency constraint. *Biol. Cybern.*, 82:357-371 (2000)

Chapter 6: Creating Implicit Assumptions - Abduction and Brain Wiring

The workings of the brain cannot be understood simply by examining it in detail or by constructing seemingly realistic large-scale theoretical models. For example, in the case of stereopsis, we must consider exactly what is being computed when we see a stereoscopic image, how to obtain the result of that computation, and how it is actually realized in the brain. What makes the problem even more difficult is that not much information is available to the brain to perform the desired calculation. In the case of binocular stereopsis, there is no information provided by the environment in which the right-eye image should correspond to the left-eye image, and a plausible assumption, called a constraint, is required. In this case, the rule of mapping the left and right images such that the resulting stereoscopic image is as smooth as possible seems to be a good one. However, if the brain is truly autonomous, it would be necessary to create constraints, as in a society where laws are made and the government is administered accordingly. The same should be done for basic processing, such as stereopsis. It is difficult to elucidate the mechanism that creates constraints. However, it is no exaggeration to say that this is one of the biggest problems that human beings have been struggling with since the dawn of history. Although there is no way to answer such a big problem, it is not a waste of time to summarize how I approached this problem.

I. Constraints Need to be Created

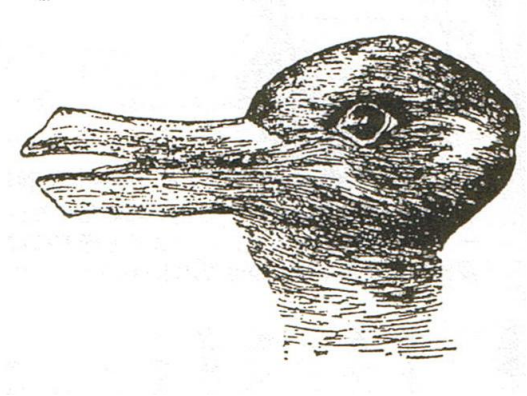


Figure 6.1. A rabbit or a duck?

Answers Vary Depending on the Situation: Part 1

What do you see in Figure 6.1? If you consider the left protrusions as ears, it looks like a rabbit looking to the right, and if you consider it a beak, it looks like a duck or other bird. Of course, which way of seeing is dominant will differ from person to person, but anyone can see either way if he or she wants to, i.e., the way to answer the question is arbitrary. Therefore, the answer to the question of how it appears depends on the situation.

If the picture in Figure 6.1 is surrounded by a meadow, we would probably assume that it is a rabbit, and if it is surrounded by a pond, we would probably assume that it is a duck. In other words, in this case, the way to answer the question depends on the situation/context of the picture; specifically, how to make the picture continuous/consistent with its surroundings.

Now look at Figure 6.2. The rightmost card has three stars. Which of the four cards on the left are the same as the three stars on the right? Again, the answer is "It depends on the situation and context." In other words, if it is a shape match, it is the second from the left, and if it is a number match, it is the second from the right.

It is called the Wisconsin Card Sorting Test, a well-known test that often detects frontal lobe disorders [1)]. The experimenter employs rules (e.g., correct or incorrect by color matching) without informing the subject. The subject is only asked, as shown in Figure 6.2, "Which card is the same as the other cards presented? and subjects must now estimate what the matching rule is through only the correct or incorrect answer they are informed of. After testing the same rule for a while, the experimenter changes the rule without informing the subject, but patients with frontal lobe impairment do not respond well to the rule switching.

This test is an ill-posed problem discussed so far in this book, in the sense that no one answer can be determined from the cards presented in a single test. The way the answer is given, the rules, or as we have discussed so far, the binding conditions of the decision will vary depending on the context. Subjects must discover and create their own rules or constraints that are consistent and simple to understand with past test results.

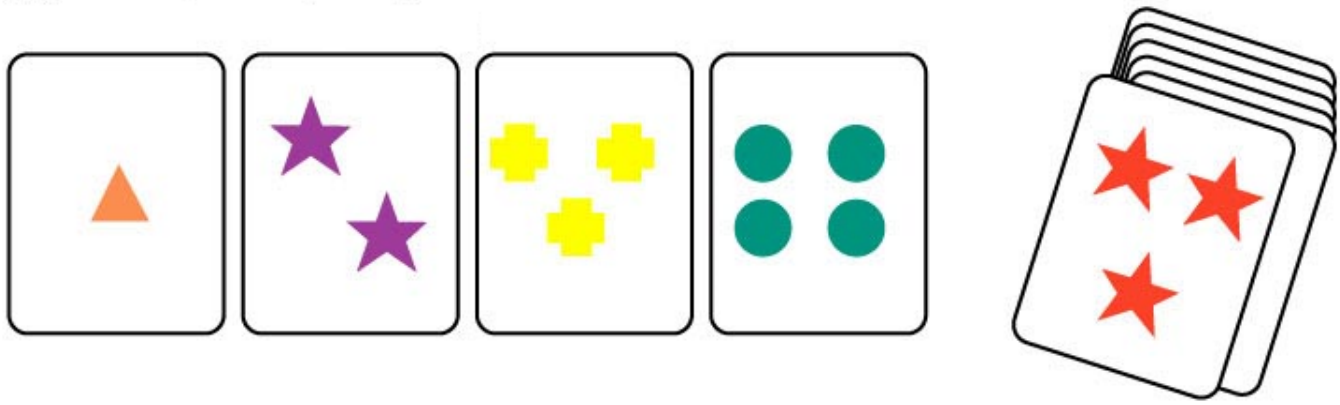


Figure 6.2. Wisconsin Card Sorting Test.

Answers Vary Depending on the Situation: Part 2

There is a motion perception problem called the rolling wheel problem. When one light spot is placed on the top row of Figure 6.3, that is, on an invisible virtual wheel, and the wheel rolls, the cycloidal motion of the light spot is perceived. Now, what if two light points are placed on the virtual wheel?

When light points are placed at both ends of the virtual wheel, as shown in the middle row of Figure 6.3, the two cycloidal motions are not perceived independently, but the motion of the two light points is organized, and the wheel, which should be invisible, is perceived as translating while rotating around the midpoint of the two light points. On the other hand, if one point is placed on the wheel and the other at the center of the wheel, as in the lower panel of Figure 6.3, the latter, not the midpoint, becomes the axis of rotation, and the former is perceived as rotating around it. In this phenomenon, the arbitrariness of the appearance is very low, i.e., everyone can only perceive it as such.

This rolling wheel problem is called an ill-posed problem, in the sense that the view is not uniquely determined from the visual motions presented. Since the brain solves this problem easily, there must be some constraints in the brain, in this case, rules about how to organize the two light points. Specifically, there should be rules regarding how the motion of the two light points is decomposed into the common motion as the motion of a point that represents the whole motion, such as the center of a virtual wheel (hereafter, the representative point), and the relative motion of the light points around the representative point.

Cutting and Proffitt stated that there are two rules for this motion decomposition: the common motion minimization rule, which determines the representative point so that the change in the common motion is minimized and the relative motion is perceived by subtracting the common motion from the physical motion of the light point; and the relative motion minimization rule, which determines the representative point such that the sum of the relative motion vectors of each light point around the representative point is minimized. Furthermore, the rule used depends on the arrangement of the two light points on the virtual wheel, with the common motion minimization rule used in the case of the bottom row of Figure 6.3, and the relative motion minimization rule used in other cases, such as the middle row of Figure 6.3. In other words, constraints change depending on the situation.

However, it is not possible to construct a computational model that can be executed by a machine, etc. as it is. The fact that common and relative motion decomposition depends on representative points means that motion decomposition is performed while determining the representative points or rules. In other words, it is necessary to clarify the meta-rule that determines which rule is appropriate: the common motion minimization rule or the relative motion minimization rule.

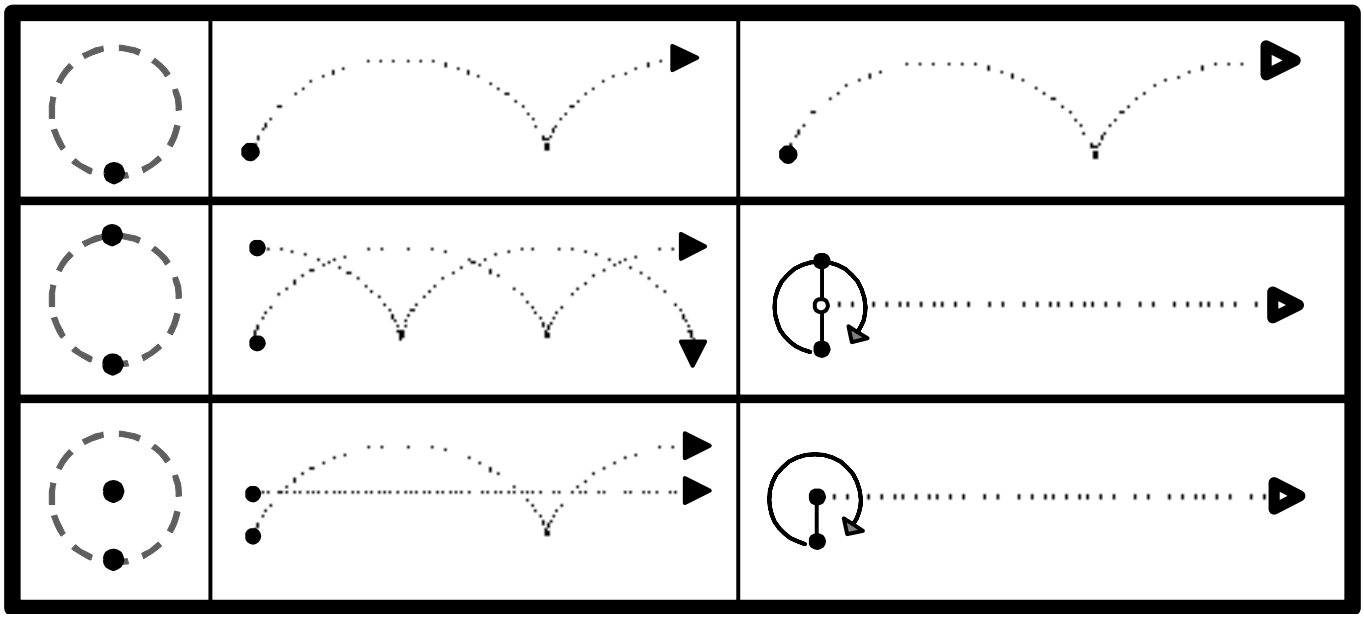


Figure 6.3. The rolling wheel problem. Left: Light points (black) placed on a virtual wheel (dashed line). Middle: the trajectory of each light point when the virtual wheel rotates. Right: perceived motion. Dotted lines indicate common motion, solid lines indicate relative motion.

Solving Problems While Creating Constraints

Among the problems that our brains must solve every day, there are many that cannot be answered with only the clues given by the outside world, so-called ill-posed problems. Constraints are required as implicit assumptions to solve such problems. Whether or not they are arbitrary, that is, whether or not they can be freely done by the power of consciousness (which is itself an important and interesting issue), constraints can change depending on the situation, which has been discussed in this book.

This strongly suggests the existence of a mechanism that determines constraints based on the situation. Of course, some constraints are innate and robust, such as the relational rules used in psychic photography when we recognize faces by making relations between trivial features. However, if we are truly pursuing a brain mechanism of creativity, does it not require the brain to create even constraints, as opposed to complex or self-organizing systems that generate spatiotemporal patterns under given constraints? Rather, is it the key to responding flexibly to the ever-changing external world? This is shared by companies that constantly update the rules and constraints of their manuals to keep up with the changing times.

Generating these constraints is a significant problem. The following is one direction in which the author considers the mechanism and implementation method to generate constraints.

Various Names for “Constraints”

The term constraint is an implicit assumption, or premise, necessary for solving ill-posed problems for which the answer is not uniquely determined. It has essential similarities with several other important terms, although they are used in different situations and have different nuances. We will understand these similarities later and list some of the relevant scopes of this document.

Constraints are implicit assumptions; therefore, it may be better to call them hypotheses. It would be better to use "tentative set" or “precondition” rather than hypothesis, which gives a stronger impression that it was generated on the spot to solve an ill-posed problem. However, throughout the following discussion, we would like to refer to it as hypothesis.

The author's mentor, Dr. Hiroshi Shimizu, uses the term field or place, which is a term from Kitaro Nishida's philosophy. Of course, he is currently engaged in a more in-depth discussion, but he definitely started using this term when he began to consider the generation of constraints.

It may not be clear to you that concepts and categories are constraints, but they are very similar in that they integrate the various things that make up this world, just as there are constraints that integrate dots in a random dot stereogram to generate a surface. The term prior probability may be used when the probability of an individual thing belonging to a concept or category is discussed.

The above was a rather rough discussion, but from now on in this book, except in special cases, what we call constraints will be unified as tentative sets and their generation as abduction.

II. Type of Thinking Tentative Sets

Type of Thinking (1) – Deduction

There are several types of our type of thinking. The first would be deduction. This may sound a bit rough, but in deduction, we think the following way. This is called the three-stage argument.

(1) Property p is possessed by all the elements belonging to population M.
(2) All the elements belonging to sample S belong to population M.
(3) Therefore, property p is possessed by all the elements belonging to sample S.

(1) is a major premise. We can call this an implicit assumption or a tentative set. For example, the law of universal gravitation states that all objects have the property of universal gravitation. (2) is a minor premise. For example, earth and apple are (belong to) objects. A specimen is a finite object that is concretely observable. If (1) and (2) are correct, we can naturally conclude (3). For example, we can conclude that (since the law of universal gravitation holds) the earth and the apple are attracted to each other, that is, the apple should fall to the ground.

Type of Thinking (2) – Induction

The second is induction. The induction is as follows.

(1) All elements belonging to set M originally possess (or should possess) property p.
(2) All the elements of sample S that should belong to set M possess property p.
(3) Therefore, it seems certain that property p is possessed by all the elements belonging to set M.

Note that even in induction, (1) the major premise, that is, an implicit assumption such as that "all objects have the property of universal gravitation," is necessary here as well. In addition, (2) it admits that a finite number of concrete observable objects have property p that they must have if the major premise is true. For example, we admit that not only apples but also other objects fall, which would naturally occur if the universal law of gravitation were true. Thus, we conclude that (3) the major premise is true. For example, "Although the number of observable objects is finite, they all fall, and planets orbit the sun, the universal law of gravitation is certain."

Again, it is important to emphasize that even in induction, the major premise is tentatively set up in advance, and it is not an issue of where it came from.

Type of Thinking (3) – Abduction

In contrast to deduction and induction, a third type of thinking was proposed by American philosopher C. S. Pearse is called hypothesis generation or abduction[3]. This type of thinking is also called the logic of discovery, in which hypotheses are generated as major premises or implicit assumptions as follows[4]:

(1) The (surprising) fact C is observed from all of the samples S belonging to the set M,
(2) (However) if property p is possessed by all elements of set M, then the observation of C from all elements of sample S is a fact that was bound to happen,
(3) (Therefore) property p is possessed (and there is reason to believe) by all elements of set M.

(1) is to admit (with surprise) that the event of an apple falling or a planet orbiting the sun is an event. (2) is that if all objects have the property of gravitation, then of course the apple will fall and the planets will orbit around the sun. Therefore, (3) the property of gravitation can be called a universal law that applies to all objects. However, (3) cannot be created from (1) and (2). In the first place, (2) is already based on a "major premise." If all objects were falling, the universal law of gravitation would be plausible, but there would be a leap in the universal law of gravitation. Assume that "all objects attract each other" because they fall or because the planets orbit the sun is somewhat oversimplified. However, it is a simple law that neatly explains various phenomena. Certainly, (2) and (3) are similar, and Peirce himself was quite confused by them, but they are clearly different. (2) is about the plausibility of a premise, as in the testing of a hypothesis in a scientific experiment, while (3) is about the creation of the premise itself, so that one can come up with the next scientific hypothesis from the experimental facts.

In terms of creating implicit assumptions as described above, abduction is the same as creating constraints for solving ill-posed problems, where the answer is not uniquely determined by the given conditions alone. Therefore, the problem of creating constraints is called "abduction" and is explained below.

Now, some readers may think that coming up with Newton's Universal Law of Gravitation means that abduction is a process of great discovery that has little to do with our daily lives in the history of humankind, but this is not the case. As we will discuss in more detail in the next chapter, we can all perform everyday tasks that are difficult for today's pattern recognizers, such as correctly inferring the whole picture of an unfamiliar object that is only partially visible, that is, generating a hypothesis about the object.

Probabilistic Pattern Recognition and Type of Thinking

Having reviewed deduction, induction, and abduction as types of thinking, since the real world involves uncertainty, let us review the correspondence between these and the discussion of probabilistic pattern recognition, i.e., when an event x is observed and it is necessary to make a probabilistic judgment as to whether it belongs to category M. For example, it is such a case that when one sees a fish of a certain size x, one must decide whether it is a rainbow trout or not.

The probability of obtaining a certain category M and a certain observation x, for example, the probability of catching a rainbow trout at a certain point in the river, can be written as $p(M, x)$, where the probability can be decomposed in two ways:

$$p(M, x)= p(x)p(M \mid x) = p(M)p(x \mid M)$$

$p(x)$ in the first equation is the probability of catching a fish of size x. For example, the probability of catching a fish of size 150 cm. On the other hand, $p(M \mid x)$ is the probability that given an observed value x, it belongs to category M, called the posterior probability. The probability that a 150 cm fish is a rainbow trout is extremely low. The second equation, on the other hand, is the product of the prior probability $p(M)$ and the probability (conditional probability or likelihood) $p(x \mid M)$ that x is observed when M.

For convenience, let us assume that all probabilities are 1 to consider the correspondence with the type of thinking.

Deduction would be equivalent to the task of finding $p(x \mid M)$ from $p(M)$ and $p(M \mid x)$. For example, the major premise (1) is the unfounded belief that there are definitely only rainbow trout at that point in the river (prior probability $p(M)$ is 1). Minor premise (2) is that if there is a fish x centimeters long, it is definitely a rainbow trout (posterior probability $p(M \mid x)$ is also 1). Thus, conclusion (3) is that since rainbow trout of size x cm continue to be caught (i.e., $p(x)$ is 1), if there is a rainbow trout, it is always x cm (i.e., likelihood $p(x \mid M)$ is 1).

The well-known Bayesian formula can be written as $p(M \mid x) = p(x \mid M)p(M)/p(x)$, which is a variant of the previous formula, and is equivalent to induction. If a person who is unfamiliar with rainbow trout casts at a point in the river (1) with the unfounded assumption that rainbow trout always exists there (prior probability p(M) is 1), (2) he/she will catch a fish of x cm, which is the standard size of rainbow trout, regardless of how many times he/she casts (i.e., both $p(x)$ and $p(x \mid M)$ are 1). Then, (3) we can be sure that there is only rainbow trout at that point in the river (the posterior probability $p(M \mid x)$ is also 1).

On the other hand, it is unclear how to determine prior probability p(M). According to the textbook[5)], the prior probability is very vague, for example, it "reflects a priori knowledge of how likely it is to catch rainbow trout" and "depends on the season and choice of points (partially modified)." There is no basis for this, but nothing can begin without a basis. In this sense, the prior probabilities are tentative.

The Properties of "Hypotheses"

I do not have any concrete ideas on how to generate tentative sets as implicit assumptions necessary for thinking and problem solving. However, there are certain properties that a hypothesis should have. Satoshi Watanabe lists them as follows[6]:

(1) A hypothesis itself is not directly observable,
(2) It is obtained from incomplete information,
(3) It enables various predictions,
(4) It is simple and beautiful.

The law of universal gravitation mentioned earlier is exactly this. The law of universal gravitation cannot be directly observed (1). Even if you look at the night sky through a telescope, the "law of universal gravitation" is not written anywhere. Moreover, Newton did not arrive at the universal law of gravitation after observing everything in the world, which means that the universal law of gravitation is derived from incomplete information (2). However, classical mechanics, including the law of universal gravitation, have yielded a variety of predictions and are still very useful (3). While explaining many phenomena, the laws themselves are simple and mathematically beautiful (4).

When we are explicitly aware of the nature of hypotheses as implicit premises, we understand that concepts and categories are also hypotheses. For example, when we see the fish shown in Figure 6.4 at a fish shop, we can easily see that the top fish is sea bass and the bottom fish is salmon. But we are only looking at these individual silver fish. We are not directly observing the concept of salmon or sea bass (1). We do not need to observe salmon and sea bass all over the world in order to obtain the concept of salmon and sea bass. In this sense, the concepts of salmon and sea bass are derived from incomplete information (2). However, by using the concept and the accompanying experience and knowledge, one can predict, for example, in the case of salmon, that "when cut open, the flesh would be pink" (3). Also, by simply saying the short word "salmon," we can conveniently exchange information about it with others (4). There is no need to say, for example, "A fish with pink flesh and silver skin that returns to the river in the fall."

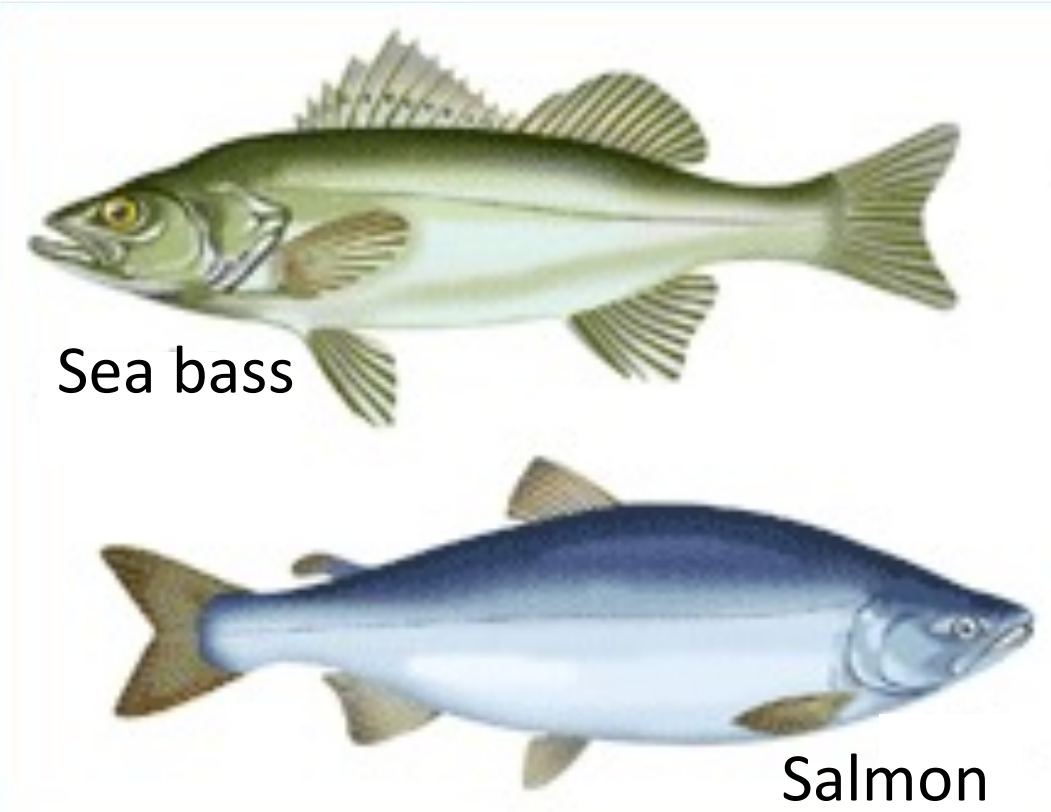


Figure 6.4. Silver fish of similar size. http://www.homemate.co.jp/useful/fishing_zukan/seasons/

Hypotheses Are not Given from the Outside

Satoshi Watanabe's The Ugly Duckling Theorem is a theorem that proves that hypotheses, i.e., concepts and categories that lump together various individuals, constraints that bring up three-dimensional images by corresponding a certain point in the right-eye image to a certain point in the left-eye image, and physical laws that explain various phenomena from falling apples to orbiting planets are not given from the outside world[6].

Usually, two individuals that are considered similar have a number of common characteristics, or categories that are lumped together. Identical twins, for example, have numerous similarities, from appearance to behavior. However, it should be noted that no matter how similar they may be, their bodies are separate and can be distinguished by whether they are on the right or left side of the face, etc.

Looking at the three birds in Figure 6.5. The two on the right are ducks and the one on the left is a swan chick. In Andersen's fairy tale, the swan chick was bullied for not looking like the others. However, the number of categories that distinguish one swan chick from one duck is the same as the number of categories that distinguish two ducks, i.e., the similarity is formally identical, which is the ugly duckling theorem.

For simplicity, here we consider only two categories. One is "gray wings" which caused bullying; let us call this category A. The other is "on the right side," which we call B. Because the two ducks are distinct entities, we can have these categories as well.

Look at the enclosure under the birds. The enclosures refer to categories. There are four categories that lump the two ducks together: the union set of regions a_2 and a_4 (regions a_2 and area a_4 combined), the union set of a_2, a_4 and a_1, the union set of a_2, a_4 and a_3, and the union set of a_2, a_4, a_1 and a_3 (that is, the whole set U). On the other hand, what about ducks and swan chicks? There are four categories that the lump duck (1) and the swan chick together: the union set of a2 and a3, the union set of a2 and a3 and a1, the union set of a2 and a3 and a4, and the whole set, while for ducks (2) and swan chicks (which we omit) there are also four. In other words, in terms of similarity, i.e., the number of distinct categories, the swan chicks are exactly the same as the ducks.

This is true even when the number of predicates increases and the number of regions enclosed by the lines shown in the figure is n. In other words, if any two things are somehow distinguished from each other, the number of categories that lump them together is constant at 2n-2. Satoshi Watanabe, the proponent of the theorem, asserts that the theorem fundamentally denies that categories exist. The hypothetical categories that lump together the various objects and phenomena are not given to us by the outside world but are created by us for our own convenience as we look at the outside world.

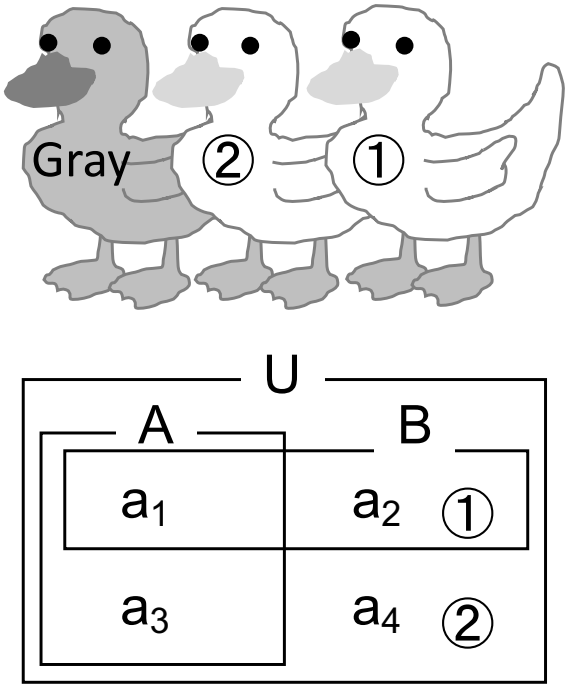


Figure 6.5. The Ugly Duckling Theorem.

III. Abduction - Voting by Wiring

Fine Wiring of the Brain

One neuron receives input from many other neurons (convergence), and sends the output to many other cells (divergence). The neural circuitry of the brain is composed of divergent and convergent structures of neuronal wiring (projections). The wiring structure is extremely complex, and experimental methods are still limited, so it is difficult to determine what kind of wiring structure a particular brain region or neuron nucleus has, that is, whether there is any regularity in the wiring or not, and what kind of computation can be performed by the wiring. We still do not know much about what kinds of calculations are made by wiring. When we try to model neural circuits in the absence of any guidelines or policies, we end up with wiring that is not arbitrary and easy to theorize between one group of neurons and another, i.e., random wiring, or its opposite, the assumption of total coupling.

However, as shown in the following BOX, the functional structure of the primary visual cortex (V1 area), which has been the most investigated area in the cortex, does not seem to have a meaningless wiring structure, such as random or total connections. We do not think that such structures are unique to V1. The brain is thought to perform sophisticated computations due to its elaborate wiring structure.

BOX *The Columnar Structure of the Visual Cortex*

It is no exaggeration to say that the primary visual cortex (V1 area), which often appears in this book, is the entry point to the cortex for visual information reflected on the retina. It is also located in a relatively easy-to-experiment location, i.e., spreads over a large portion of the cortex's occipital lobe surface. For these two reasons, it is one of the most investigated regions (areas) in the cerebral cortex.

There are various ways to examine them. For example, the figure on the left shows a brain sample taken immediately after continuous visual stimulation of only one eye and stained using a special method. The cortex is known to have a six-layered horizontal structure based on the distribution of cells. This staining suggests that information from the right and left eyes is input and processed in a regular comb-like pattern perpendicular to the layered structure. These functional units, which are organized vertically in the cortex, are called columnar structures.

Cells in the V1 cortex respond to line stimuli of a specific orientation presented in a spatial range called the receptive field. Cells that prefer similar orientations are vertically distributed.

Taken together, the functional structure model shown on the right side can be considered. Such an elaborate structure is thought to consist of an elaborate wiring structure between the neural circuits.

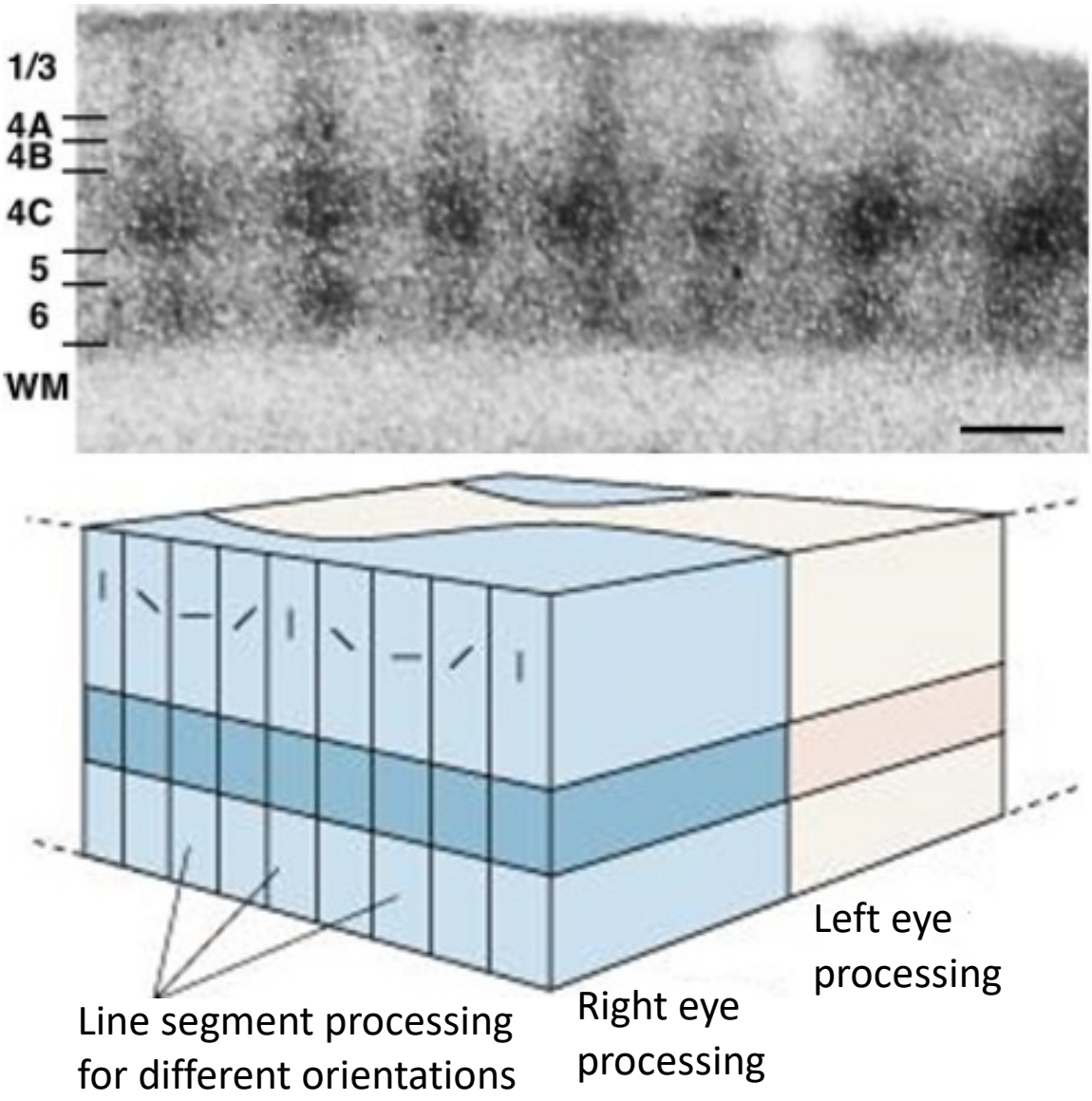


Figure. Up: Ocular dominance column structure of monkey V1 cortex by 2-deoxy-glucose staining. The numbers on the left are cortical layer numbers from the surface, WM is white matter (Ref. 7). Down: Model of the column structure of V1 cortex. It is believed that there is one functional unit perpendicular to the layer structure (from Ref. 8).

The Calculation Capability of Neural Wiring

There are various types of neural network models, but when the goal is not to realize a specific computational function, but rather to investigate the properties of the circuit itself, the wiring between cells is often assumed to be random, totally connected, or something else that is easy to theorize. However, actual neural circuits always perform some kind of computation, and wiring must serve to achieve that computational purpose. In contrast, there are neurons in the cerebral cortex, for example, that can only be assumed to have extremely sophisticated information-transforming wiring behind them, with responses so complex that it is difficult to imagine from the neuronal responses in the preceding stages.

For example, a visual area called the CIP cortex in the parietal lobe, discovered by Professor Ken-ichiro Tsutsui of Tohoku University, contains neurons that encode the three-dimensional orientation of a plane (in which direction the plane is tilted)[9]. Surprisingly, these cells show a strong response not only to binocular disparity (disparity between right and left eye images) but also to plane orientation inferred from monocular perspective cues (images seen with one eye appear smaller for distant objects and larger for near objects) (Figure 6.6). However, no such complex cellular response has so far been reported in the area called V3A, which is the main projection source to the CIP cortex and plays a preliminary role in visual processing[10]. The V3A cortex is known to respond to binocular disparity, but the only report on the response to monocular stimuli is that many neurons respond to line segments in a specific orientation presented in the receptive field (the area in the visual field to which neurons respond).

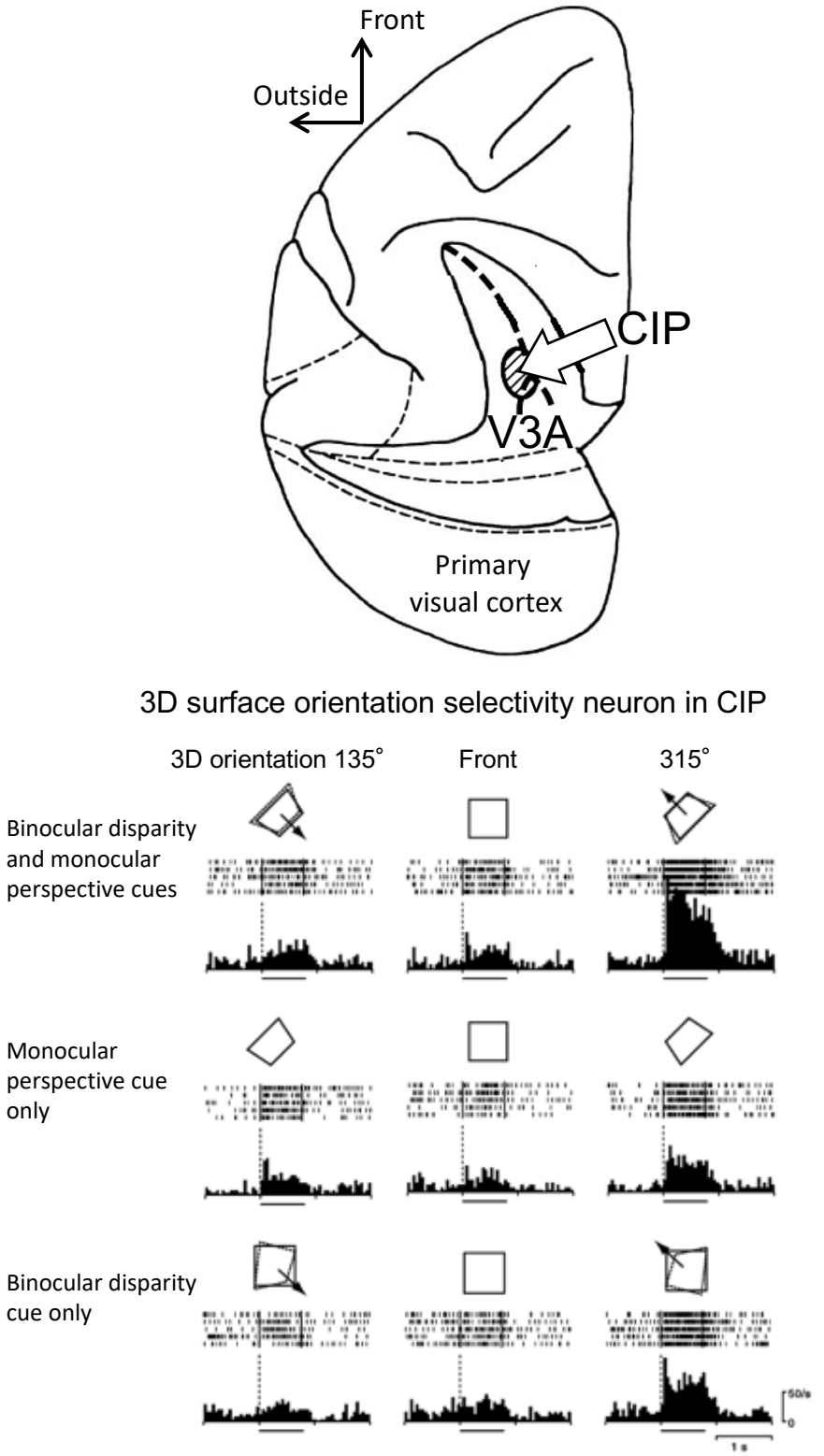


Figure 6.6. Location of the CIP cortex (top) and response of the CIP cortical neurons to the three-dimensional orientation of the plane (bottom). (Top) Top view of monkey cerebral cortex. The intraparietal and lunate sulci were open. (Bottom) Cells respond well to a three-dimensional plane orientation of 315°. They responded best when both binocular disparity and monocular perspective cues were included in the stimulus, but they also responded to each cue alone (from Ref. 9).

Hough Transform - Voting for Parameter Space

By viewing neuronal wiring not as random, but rather as a geometric transformation called voting into parameter space, Dr. Susumu Kawakami and his group constructed a neural network model for detecting moving objects[11]. Let us consider one of the core components of the model, a line detection method called the Hough transform, as an example to illustrate voting in parameter space.

The Hough transform facilitates the detection of interrupted straight lines[12]. It was first used to automatically detect the trajectory of particles in a device called a bubble chamber, to observe neutrinos and other particles.

Let point A in the image be represented as (x_A, y_A) using x-y coordinates (Cartesian coordinates). A perpendicular line is drawn from the origin to an arbitrary line passing through point A. Let ρ_A and θ_A be the length of the perpendicular line and the angle it makes in the positive direction of the x-axis, respectively (Figures 6 and 7, left). Then, for (ρ_A, θ_A) and (x_A, y_A), the relationship

$$\rho_A = x_A \sin\theta_A + y_A \cos\theta_A$$

is established.

Of course, infinitely many straight lines pass through point A. That is, (ρ_A, θ_A) also exists infinitely. However, it is constrained to pass through the point (x_A, y_A). If we now consider a new parameter space consisting of ρ and θ coordinates, the line passing through (x_A, y_A) draws a sinusoidal wave in the direction of the θ axis in the ρ–θ parameter space (curves A–A in Figure 6.7, right). In other words, the entire straight line that could pass through the point (x_A, y_A) will draw a curve in the ρ–θ space, which is called a "vote" from the point (x_A, y_A) to the ρ–θ parameter space. Similarly, point B in Figure 6.7 left votes for curve B to B in Figure 6.7 (right) in the ρ–θ space.

If the points are aligned in a straight line from our point of view, as shown in Figure 6.7, the line can be easily detected in the ρ–θ space. That is, (ρ_0, θ_0), which represents a straight line through all these points, can be detected as the intersection of the curves drawn by voting from each point (Figure 6.7, right).

Conversely, the transformation from ρ–θ space to x-y space, which is called the inverse Hough transform, is also a parameter space vote. Point (ρ_0, θ_0) votes for the line indicated by the dotted line on the right in the x-y coordinate system.

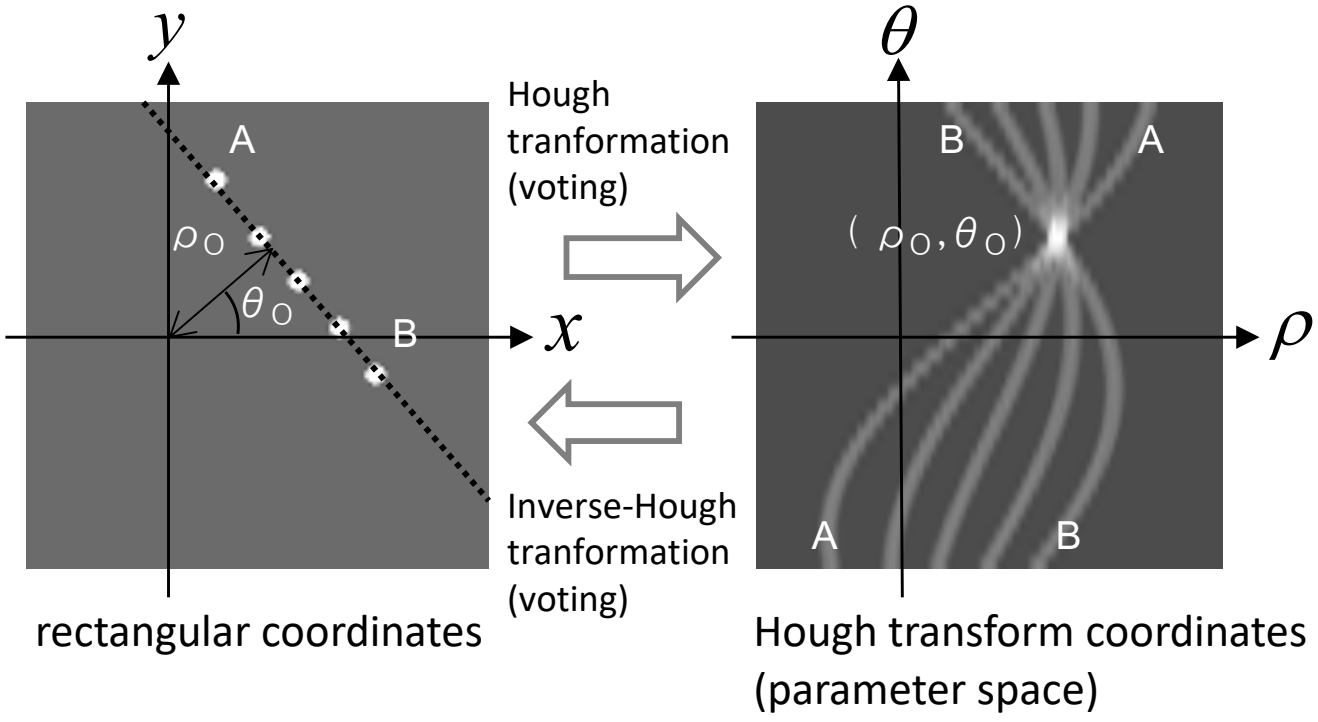


Figure 6.7. A typical example of voting in parameter space, the Hough transform. Although not discussed in the text, the inverse Hough transform, i.e., the transformation from a single point in Hough transform coordinates (ρ–θ space) to Cartesian coordinates, is also divergent voting.

View Orientation-Selective Neurons as Hough Transformers

Dr. Kawakami, in creating his neural model, viewed the divergent and convergent structures of neural circuits as voting in a parameter space; for example, orientation-selective simple cells in the primary visual cortex (V1 cortex) were viewed as performing the Hough transform described in the previous section in a certain spatial range.

Contours are important cues in visual perception. Correspondingly, neurons in cortex V1, the first major area in the cerebral cortex to receive signals from the eye, detect contour fragments or local line segments. That is, the neurons respond well when the line segment is presented in the neuron's preferred orientation in the receptive field (the spatial extent within which the neuron responds) (Figure 6.8A). In orientation selective simple cells, the response depends on where in the receptive field the line segment is presented. Some cells respond well when a segment is presented at the center of the receptive field, whereas others respond well when a segment is presented at the edge of the receptive field. These properties are referred to as spatial phase properties.

On the other hand, the neurons in the lateral geniculate nucleus (LGN) that relay the visual signal from the retina and project to orientation-selective simple cells have no orientation selectivity, and each cell only responds to light spots in the receptive field.

It is difficult to imagine that the responses of orientation-selective neurons emerge from random projections of LGN neurons that respond only to spot stimuli. There must be well-organized projections, as modeled by Dr. Kawakami (Figure 6.8B) :

He considered that a group of cells with receptive fields in the same spatial range, but with different orientation selectivity and spatial phase properties, constituted the local Hough transform coordinates. That is, he regarded orientation selectivity as θ and spatial phase property as ρ.

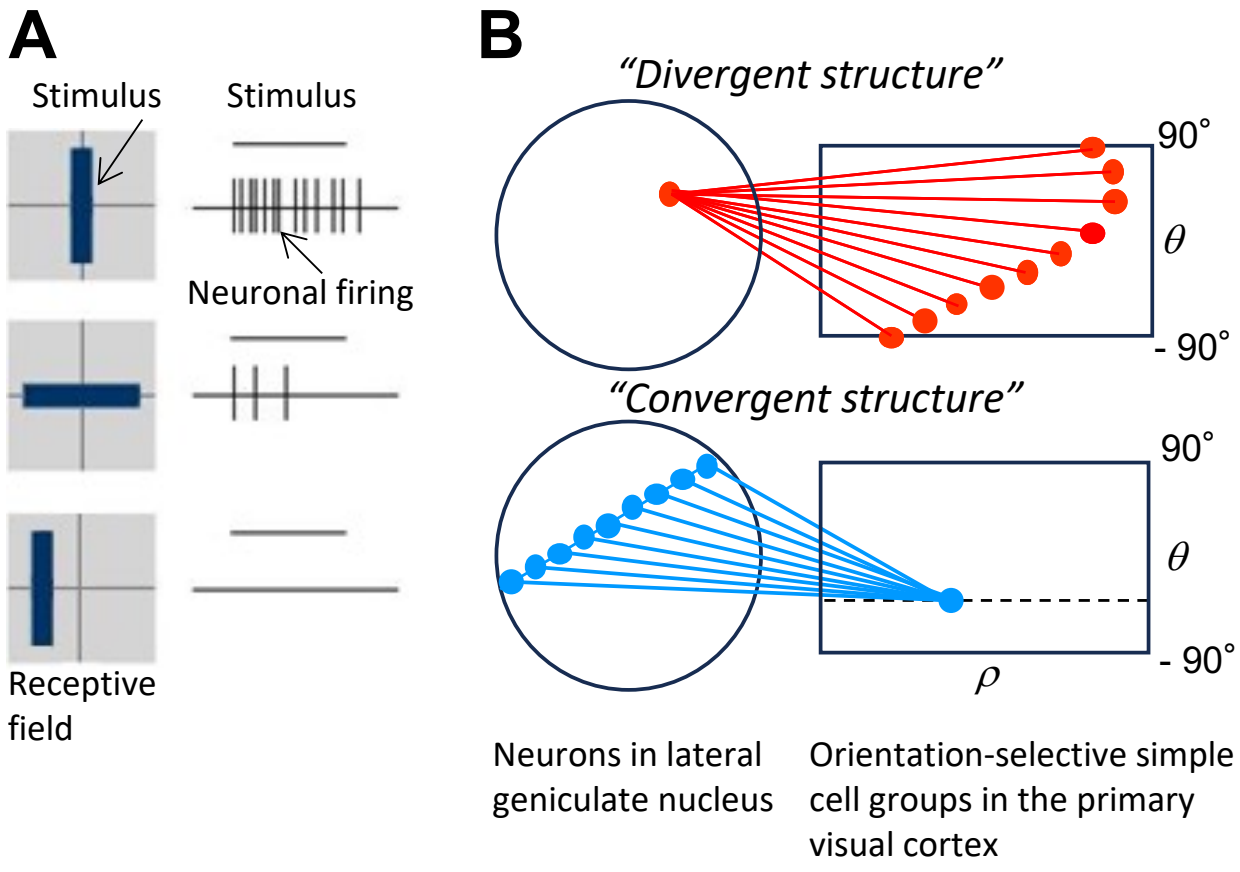


Figure 6.8. Orientation-selective simple cells in the primary visual (V1) cortex can be viewed as performing a Hough transform over a certain spatial range. A. Schematic representation of the response of orientation-selective simple cells in V1 cortex. This cell responds to the vertical line segment presented at the center of the receptive field, i.e., θ=0°and ρ=0. The response was reduced for different orientations and spatial phases. B. Response properties as in A are realized by divergent-convergent projection structures that were organized from the lateral geniculate neurons. Right, group of cells in the V1 field with a common receptive field range; each point in the ρ–θ space represents the response properties of each cell. Left, range of lateral geniculate cells projecting into ρ–θ space on the right; each point represents a single cell.

Similarity between Hough Transformation and Abduction

As mentioned previously, the divergence/convergence structure of neural wiring can be used to vote on the parameter space represented by the Hough transform. If you think about it carefully, there seems to be a strong correspondence between the properties of the result of the Hough transform and the properties of the hypotheses that are obtained as implicit assumptions.

(1) A hypothesis itself is not directly observable,
In the Hough transform, only individual points can be observed directly.

(2) It is obtained from incomplete information,
In the Hough transform, a straight line is estimated from a small collection of points.

(3) It enables various predictions,
As a result of the Hough transform, we can predict that the area without points would also be part of a straight line.

(4) It is simple and beautiful.
For example, a simple collection of points can be expressed as (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), and (x_5, y_5) in the Cartesian coordinate system, but as a result of the Hough transform, they are integrated and expressed as (ρ_1, θ_1), which is a neatly compressed simple representation.

Given these extremely good correspondences, you say that the straight line obtained by the Hough transform is a hypothesis.

Implementation of Abduction: "Part" and "Whole" Inversion

So can abduction be implemented using the divergence/convergence structure of neural wiring? To consider this, let us consider why, on the contrary, it is difficult to implement abduction.

Looking back on why it is difficult to implement the process of abduction, it seems that it is because a hypothesis is often about the harmonic relationship of the "whole." For example, the law of universal gravitation and Newtonian mechanics are about the "whole" of the universe, and in the example of the Hough transform mentioned above, the obtained (ρ_0, θ_0) is about the "whole" of the five points. However, the "whole" is vague and elusive. Therefore, at first glance, it seems unsuitable for implementation in neural circuit models that are suitable for local processing, such as lateral inhibition (inhibition of neighboring neurons).

Can we localize the processing of the whole to make it easier to implement in a neural circuit model? This seems to be exactly what Dr. Kawakami did with their neural circuit model, which implements the Hough transform in the divergence/convergence structure of the neural wiring. In the Hough transform, a point that is a "part" on the Cartesian coordinate plane becomes a curve on the Hough transform plane, and a line that is a "whole" on the Cartesian coordinate plane becomes a point on the Hough transform plane (referred to as the line/point duality). In other words, the whole and parts are inverted on the Cartesian coordinate plane and on the Hough transform plane. Using this inversion relation, it is expected that abduction that is about the "whole" can be realized by local processing of neural circuits.

However, Abduction through Voting Still Has Some Problems

So far, we have discussed the possibility of creating a hypothesis, an implicit assumption about the "whole," using the divergence/convergence structure of neuronal wiring, which reverses the "whole" and the "part" relationship of the represented information in the network. Namely, individuals and categories are not in a unidirectional inclusionary relationship, but are in a mutual inclusionary relationship through the inversion of the "whole" and the "part."

Consider, for example, the case of a silver object seen at a fish store that definitely belongs in the salmon category. Conversely, the concept of salmon is about the entire experience and knowledge of the silver fish. On the other hand, the individual object in front of us encompasses salmoniness or salmon-like qualities, but also includes qualities not limited to salmon-like qualities such as "cheap" or "freshness.

However, one should be careful about using parameter-space voting, such as the Hough transform, as the basic brain mechanism for abduction. It has something in common with Bayesian inference. The Hough transform infers the most likely straight line represented by (ρ, θ), when multiple points are observed. This is the exact Bayesian inference. In other words, one straight line (ρ_0, θ_0) is given as a prior distribution.

However, there are some important points in which the Hough transform differs from mere Bayesian estimation. One is that there is a spatial relationship between the "prior probabilities" in voting on the parameter space. A point near (ρ_0, θ_0) in the Hough transformation space represents a line similar to that represented by (ρ_0, θ_0). Hence, it is also possible to have an interaction between "prior probabilities." Kawakami's neural network model, which I mentioned briefly before, had a mechanism whereby the point with more votes wins through mutual inhibition between points in (ρ, θ) space. Introducing self-organized interactions among "prior probabilities" may generate "prior probabilities" and "categories" that did not exist before.

One may also find such an argument incompatible with what I have been saying. That is, one may feel that experience and categories can interact, even though I have said that hypotheses themselves cannot be derived directly from experience. I want to emphasize that categories and hypotheses definitely exist in our minds and can interact with experience; however, they do not come directly from experience. For example, one cannot look at the night sky through a telescope and find the laws of universal gravitation written down, nor can one catch the very concept of salmon, no matter how many silver fish catch running up the river in fall. Hypotheses are only what you create for yourself. They are useful for representing information, reducing the cost of processing it, and making predictions.

However, to reveal the neuronal mechanisms of abduction, we need to study them through specific and good examples. In the next chapter, I pick up the amodal completion problem.

References

1) Milner B. Effect of different brain lesions on card sorting. *Arch. Neurol.*, 9:90-100 (1963)
2) Cutting JE, Proffitt DR. The minimum principle and the perception of absolute, common, and relative motions. *Cogn. Psychol.*, 14:211-246 (1982)
3) Yonemori Y. *Abduction – Kasetsu to Hakken no Ronri*. Keiso Shobo, Tokyo (2007) in Japanese
4) Yagyu T. *Sekkei kara Mita Abduction*. In G*ijutsu-Chi no Honshitsu*. 135-158. University of Tokyo Press, Tokyo (1997)
5) Duda RO, Hart PE, Stork DG. *Pattern classification (2nd)*. John Wiley & Sons, New York (2001)
6) Watanabe S. *Ninshiki to Pattern*. Iwanami, Tokyo (1978) in Japanese
7) Lund JS, Angelucci A, Bressloff PC. Anatomical substrates for functional columns in macaque monkey primary visual cortex, *Cereb. Cortex*, 12:15-24 (2003)
8) Delcomyn F. *Foundation of Neurobiology*. WH Freeman, New York (1998)
9) Tsutsui KI, et al. Integration of perspective and disparity cues in surface-orientation- selective neurons of area CIP. *J. Neurophysiol.*, 86:2856-2867 (2001)
10) Nakamura H, et al. From three-dimensional vision to prehensile hand movements: The intraparietal area links the area V3A and the anterior intraparietal area in macaques. *J. Neurosci.*, 21:8174-8187 (2001)
11) Kawakami, S., Okamoto, H. A cell model for the detection of local image motion on the magnocellular pathway of the visual cortex. *Vision Res.*, 36:117-147 (1996)
12) Hough PVC. Methods and means for recognizing complex patterns. U.S. Patent 3069654 (1962)

Chapter 7. Inference of the Occluded Part: Amodal Completion and Abduction

The brain faces many ill-posed problems whose solutions cannot be determined with the given information alone. To solve such problems, constraints, which I called hypotheses in the previous chapter, are required. But if the brain is a truly autonomous system, hypotheses must also be generated. But, it is not easy to obtain the general theory of abduction.

As clues to solve this problem, the previous chapter described the implementation of a line detection method called the Hough transformation, which uses the divergence-convergence structure of neural networks, and showed examples where the apparent constraints change dynamically depending on the situation.

In this chapter, I will discuss the problem of amodal completion as a familiar and simple example of abduction. To solve this problem, I and Taichi Kumada developed a computational model including a method for curvature detection by extending the Hough transformation. Our model included the more important mechanism of voting and representing the input image in a high-dimensional feature space and compressing it in a specific dimension. The dimension of compression varied with the input image. This change corresponds to the apparent change in constraints. The model was inspired by the neurophysiological findings, and also implemented the Praegnanz Law of Gestalt psychology. The integration of this model with the distinctive mechanisms of complex systems will, I believe, lead to a deeper understanding of the brain mechanisms of creativity.

I. Abduction Can Be Studied through Specific Problems

Amodal Completion Problems Are All around Us

Our surroundings are full of occluded objects. Look around you. You will find many things partially hidden by others. Nevertheless, you can easily guess the whole shapes, even if it is unfamiliar. For example, unexpected objects are often unearthed at archaeological sites and fossil excavations. However, even non specialists can distinguish the buried object from a mere stone. This is because our visual system has the ability to complete the occlusion part from visible cues for any given shape.

Without such an ability, daily life would be difficult. You wouldn't be able to get a shirt out of a pile of laundry. You wouldn't be able to get a jar of jam out of the refrigerator. Amodal completion of arbitrary objects is an indispensable ability for field robots and helper robots to come. One might say that one can infer from the object's memory. But our world is full of the unfamiliar and the unknown, much like an archaeological site or a fossil excavation site.

Consider Occluded Shapes

Shirts in a pile of laundry or jars in the refrigerator are not suitable for the first step to revealing the mechanism of amodel completion or the neural processing that completes the arbitrary occluded shapes based on clues from the unoccluded part. This is because natural images contain various factors, and it is not possible to clearly separate which factors influence the completion process and how.

The abstract figures shown in Figure 7.1 were used in our study. These are part of huge number of figures created by Manabu Kato, a graduate student of our laboratory, to explore factors affecting amodal completion. These figures provided us with a deeper understanding of the problem. Namely, we recognized the importance of factors such as how the white and gray areas contact each other, whether the contours are smooth, and whether the unoccluded areas can be interpreted as part of a line symmetric or rotational symmetric figure.

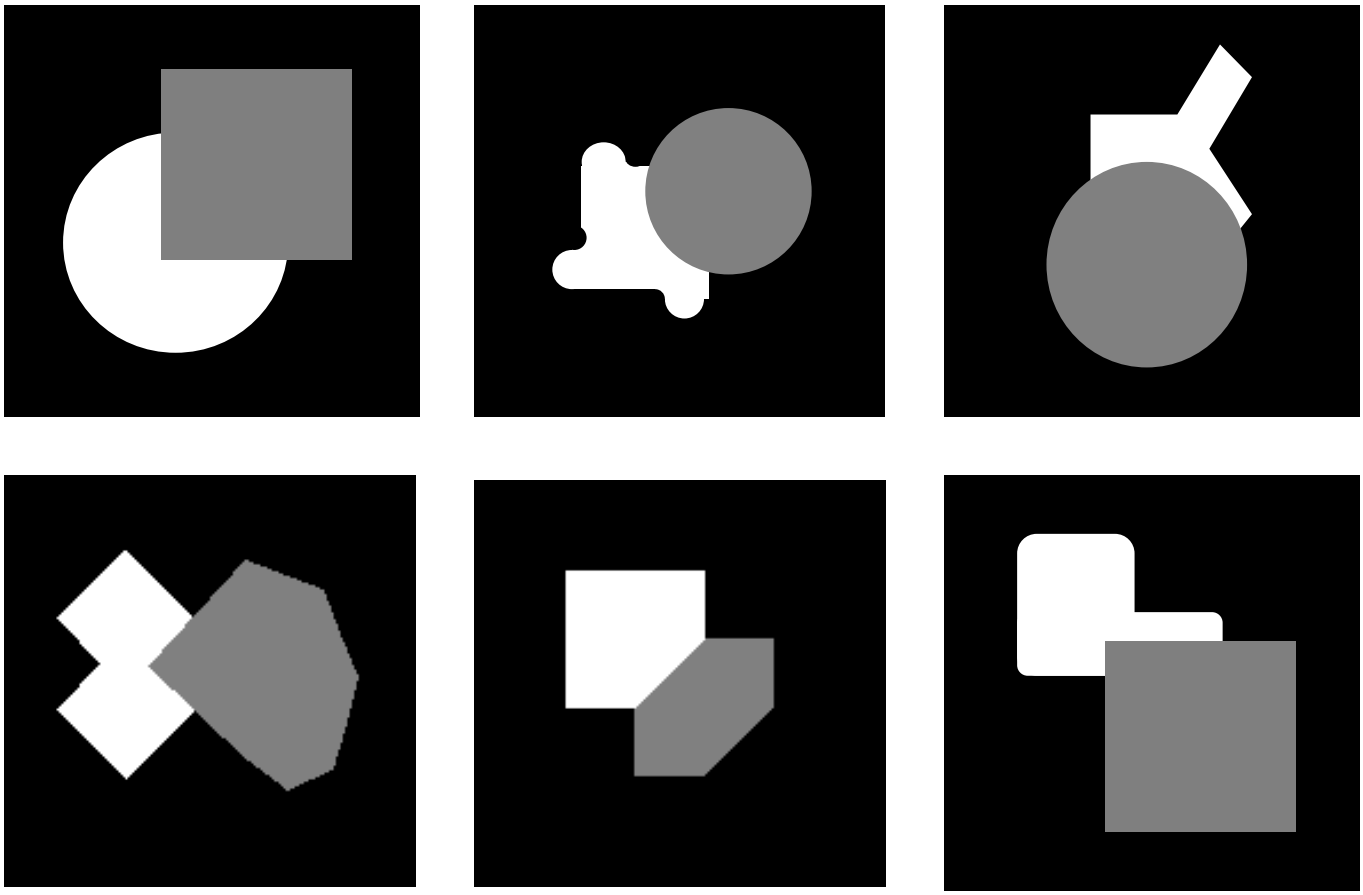


Figure 7.1: Examples of amodal completion figures.

Amodal Completion Does not Have a Unique Solution

A figure that includes shapes overlapping each other is called a occluded figure. How to perceive such figures cannot be uniquely determined from the physical properties of the shapes. In this sense, the interpretation of occluded figures is an ill-posed problem. This problem is divided into two subproblems.

One problem is the uncertainty of the occlusion relationship, i.e., which shape occludes the other (Figure 7.2, middle row). This problem includes the case where two regions do not overlap but are in contact with each other like a mosaic (Figure 7.2 middle row, second from the right). Of course, it is also possible to understand the two regions not as two areas or objects, but as a single mass (the rightmost in the middle row of Fig. 7.2). In the case of this figure, however, the gray circle will appear to most readers to be in the foreground and the white region in the background (middle leftmost in Figure 7.2).

The other problem is the multiplicity of contour completion. There are infinite possibilities for completion, including the case of no completion (Fig. 7.2 bottom row, second from the right). Even odd cases, as indicated in the rightmost of Fig. 7.2 bottom row, are allowed. In practice, however, one would feel that the two cases illustrated on the left side of Fig. 7.2 bottom are reasonable. The leftmost is the case where the shape is completed depending on its global feature, specifically, in this case, partial rotational symmetry. In contrast, the second from the left is the one where the completion is based on the local continuity of the contour.

In the case of this figure, many people may feel that the completion based on the global property of overall symmetry is more reasonable than completion based on local continuity of contours. However, for some figures, completion based on local contour continuity may be dominant. For example, in the upper right of Figure 7.1, completion based on line symmetry with a horizontal symmetry axis is possible, but completion based on local continuity that produces a scoop-like shape looks more natural.

These observations show that the rules (constraints) that seem necessary to solve the ill-posed problem of amodal completion change dynamically depending on the figure.

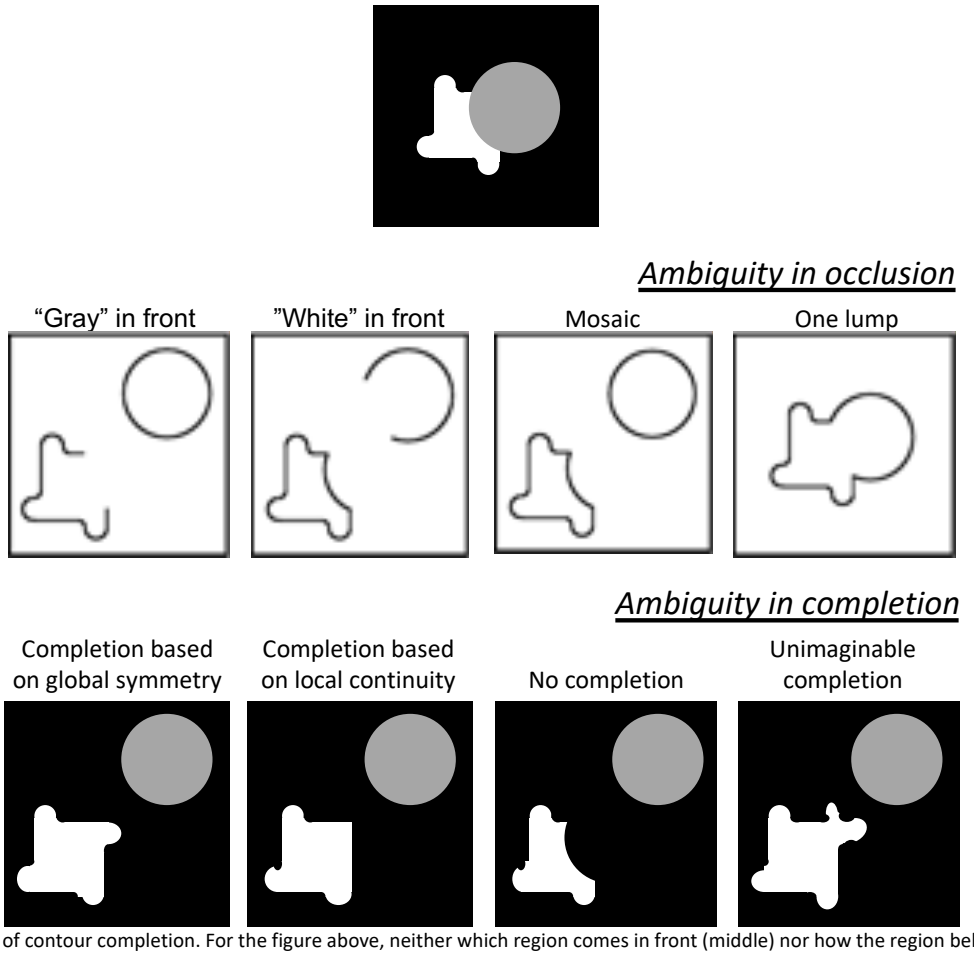


Figure 7.2. Multiplicity of contour completion. For the figure above, neither which region comes in front (middle) nor how the region behind is complemented is uniquely determined (from Ref. 1).

Amodal Completion Is a Good Example of Abduction

The amodal completion problem, in which the hidden part of an occluded shape is speculated, cannot be solved uniquely from the given figure. In this sense, it is an ill-posed problem that requires constraints as implicit assumptions in order to be solved.

However, the author believes that the amodal completion problem is not just an ill-posed problem, but a good example of abduction, because it contains important aspects as described below.

Let's look back again at the properties that should be provided for the hypotheticals listed by Satoshi Watanabe, and compare them to amodal completion.

(1) A hypothesis itself is not directly observable,
The completed shape cannot be directly observed.

(2) It is obtained from incomplete information,
The occluded part is inferred from the incomplete information of the visible part.

(3) It enables various predictions,
The occluded part can be predicted in detail.

(4) It is simple and beautiful.

There are infinite ways to complete the occluded part, but a simple and beautiful shape is preferred as discussed in the following sections.

After examining many figures including occluded shapes, I can raise two major constraints, or hypotheses, for amodal completion.

One is the local-continuity-of-contour constraint, i.e., a constraint by which the hidden contour is generated so that it is smoothly connected to the contour of the visible portion. This computation can be implemented based on the local curvature of the visible part adjacent to the occluding region. In fact, a neural network model for this purpose has already been proposed[2)].

The second is symmetry-based completion. This is a completion based on the incomplete symmetry of the contour of the visible area, i.e., the shape of the unoccluded part can be regarded as part of a line symmetrical figure or a rotationally symmetrical figure. This is a constraint that takes the entire shape into account.

The problem is that these two binding conditions are not always compatible. If the occluded shape is a square or a circle, they do not conflict, but for many shapes, one or the other is dominant depending on the figure. This fact implies that beyond these constraints there is a mechanism that decide which is preferable. Such mechanism is the very mechanism related to abduction we are now seeking. And the mechanism should have something to do with the nature of hypotheses that “it is simple and beautiful.”

What Does "Simple and Beautiful" Mean?

Our vision can infer and complete the invisible parts of a shape from the visible parts, even if parts of the shape are hidden. However, the correctness of the completed figure itself can never be known unless the obstruction on the shape is removed. In this sense, the completed shape is only a hypothesis. The preferred hypothesis is undoubtedly a simple and beautiful one. "Simple and beautiful" is an important law in Gestalt psychology that flourished in Germany in the early 20th century, known as Pregnanz's law (Japanese for "simple"). However, "simple and beautiful" does not fit in with science.

In this respect, I appreciate Van Lier's approach (Fig. 7.3)[3]. In amodal completion problem, they symbolized the features of the shape and expressed it by the symbols. If the shape had symmetry, it was used to compress the representation. Furthermore, they conducted a psychological experiment to determine whether partial symmetry-based or local continuity-based completion is preferred, and found that the shapes with shorter representation were preferred. In other words, they replaced the scientifically undesirable concept of "simple and beautiful" with the scientific concept of length of representation.

Their approach, however, cannot be performed automatically or mechanically, since the detection and symbolization of features is done manually. In addition, they target only figures composed of straight lines. A method to detect features automatically, i.e., bottom-up, from curved lines as well as straight lines of contours has not been realized.

Details are illustrated in the subsequent sections, but we have developed a novel computational model by considering the physiological findings and extending the Hogh transformation described the previous chapter. The model extracted not only line features but also curve features from the contour in the image in a bottom-up manner, allowing for both partial symmetry-based and local continuity-based completion. The candidates for completed shapes were assigned a representation amount. The representation amounts of symmetrical figures were compressed based on their symmetry. Then, the amounts were compared among all candidates, including those that did not needed to be completed. All of these candidates were output, but shapes with smaller representation amount were prioritized.

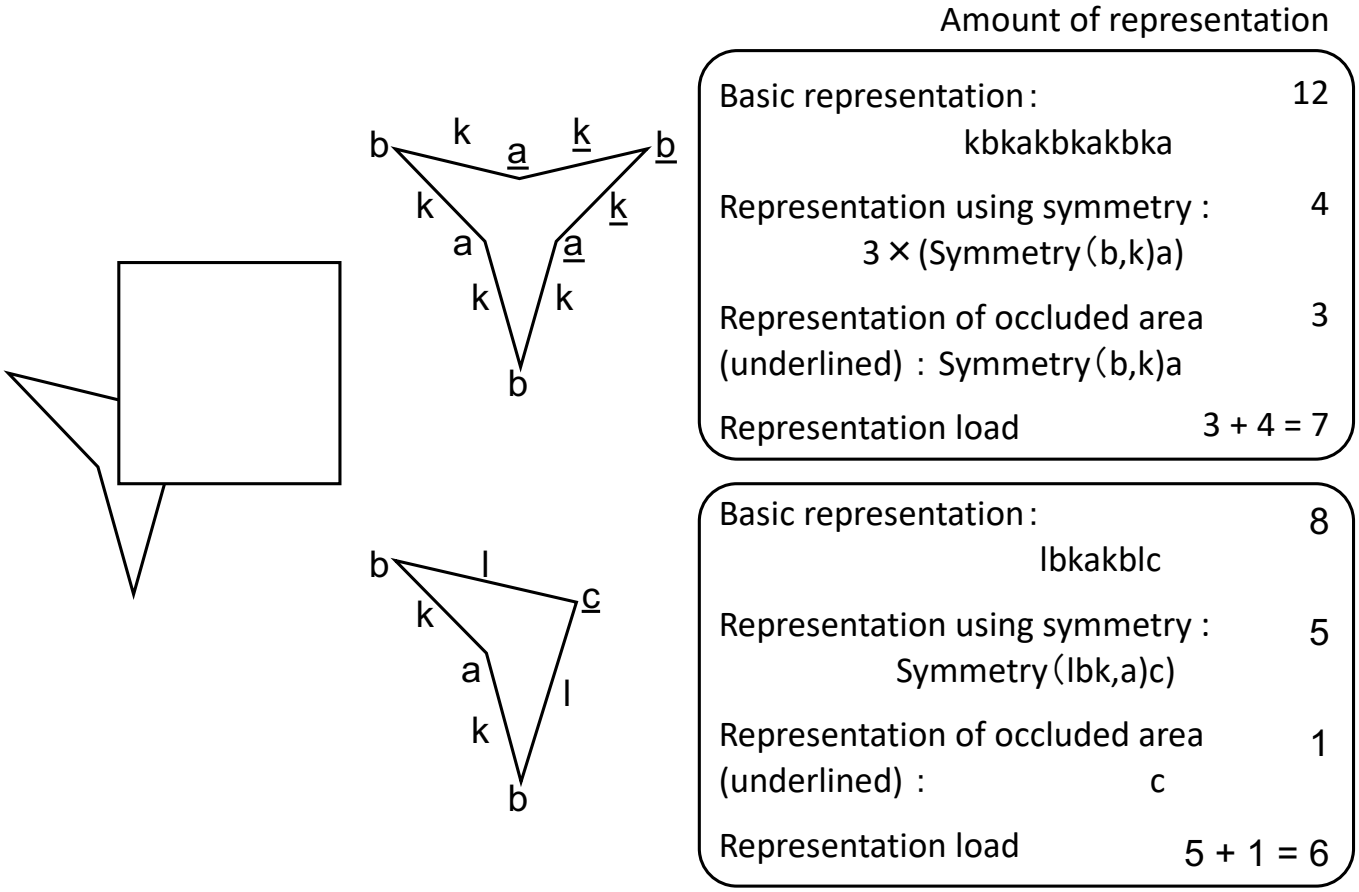


Figure 7.3. Occluded shapes and the shape representation amount. There are two major interpretations (center) for the occluded shape (left). The edges and corners are expressed by symbols, and the representations are compressed based on rotational or linear symmetry. Considering that complex representation for the occluded part (underlines) is not preferred, the proffered amount of representation is decided (based on Ref. 3).

Our Amodal Completion Model Explicitly Addresses the "Simple and Beautiful" Law

In the part II, we have discussed various concepts such as abduction, the Hough transformation, voting and amodal completion etc. Here I illustrate the relationships among them for review before going into the details of the model.

Hypotheses refer to implicit assumptions that lump different things together. To solve many of the ill-posed problems that brain faces such as binocular stereopsis, problems that cannot be uniquely solved from given cues alone, hypotheses or constraints are needed to establish consistent relationships between different elements or factors. In probabilistic pattern recognition, it is discussed which category each observation probabilistically belongs to. However, this problem also requires a prior probability as an implicit assumption. The process of obtaining a hypothesis from limited experience or incomplete information is called abduction. Hypotheses are inherently unobservable, but they are simple, beautiful, and allow for a variety of predictions. However, the problem of how to create hypotheses is quite huge and ambiguous. Therefore, we have to consider a specific and focused problem to find the fundamental principles of abduction.

The Hough transform is a well-known algorithm designed to robustly detect interrupted straight lines. A straight line can be represented by two parameters: the distance and angle of the line perpendicularly drawn from the origin to the line of interest. A single dot has countless straight lines passing through it. However, the two parameters that represent a straight line have a specific relationship. Then, all sets of the two parameters that represent any possible lines passing through the dot are "voted." Given a set of linearly aligned dots, the very parameter set that collects votes from each dot represents the line of dots. The "straight line" detected by the Hough transformation has the properties of a hypothesis: it is obtained from incomplete information, a line of dots; the line is simply expressed with only one set of parameters; you can predict that the areas between the points are also the part of the line. Therefore, I believe that the divergence-convergence structure of the nervous system can be used as a specific method of "voting" in abduction. However, as mentioned above, the Hough transform has some similarities with induction or Bayesian inference. In addition, simplicity of representation is not an overt issue in the Hough transform. Therefore, we focused on the amodel completion problem.

The amodal completion problem is the problem of inferring occluded shapes. This problem is an excellent example of abduction in the sense that the occluded portions of a contour that are never observed must be inferred from non-occluded portions or incomplete information in a simplicity-first manner. In our computational model, we also used a new method for detecting curved portions of contours by extending the Hough transform to the extraction of global features of contours, which was done manually by Van Lier et al. Furthermore, in contrast to the straight line detection problem with the Hough transform, the amodel completion problem requires addressing the simplification problem. To do so, the extracted features must be mapped to a higher dimensional space and a lower dimensional subspace must be found to compress the shape representation. Details are given in the following sections.

II. Theoretical Model Based on V4 Curvature Neurons

Curvature Neurons in the Visual Area V4

When developing a computational model for solving the amodal completion problem, in which the occlusion portion of a contour is inferred from the visible portion of a shape and completed, especially when developing it based on linear or rotational symmetry of the shape, it is necessary to compute global features from a given image in a bottom-up fashion. If the contour of a figure consists only of straight lines, straight parts and corners can be easily detected and defined, but what if it contains curves? Here, we focused on the neuronal activity in a region of the cerebral cortex called the V4 area.

The V4 cortex is located at an intermediate stage of the visual object processing pathway in the cortex (see BOX). We have referred to the experiments of Pasupathy and Conner as a clue to the intermediate stage of contour processing in V4 [4,5]. Neurons in the visual cortices have an area called the receptive field in which neuronal activity is excited or inhibited by the stimuli presented. Neurons do not respond to any stimuli presented in the receptive field. They showed that V4 neurons selectively respond to stimuli with a specific curvature and an orientation.

The significance of these neuronal activities become apparent when visual shapes are used as stimuli: Pasupathy and Connor presented a number of shapes containing various curvature features in V4 neurons' receptive fields (Figure 7.4). These neurons responded selectively to specific curvatures in specific directions (i.e., up or to the right of the shape, for example), regardless of the differences across the whole shapes.

As Pasupathy and Conner also attempted, visual shape may be represented by the collective activity of these V4 neurons.

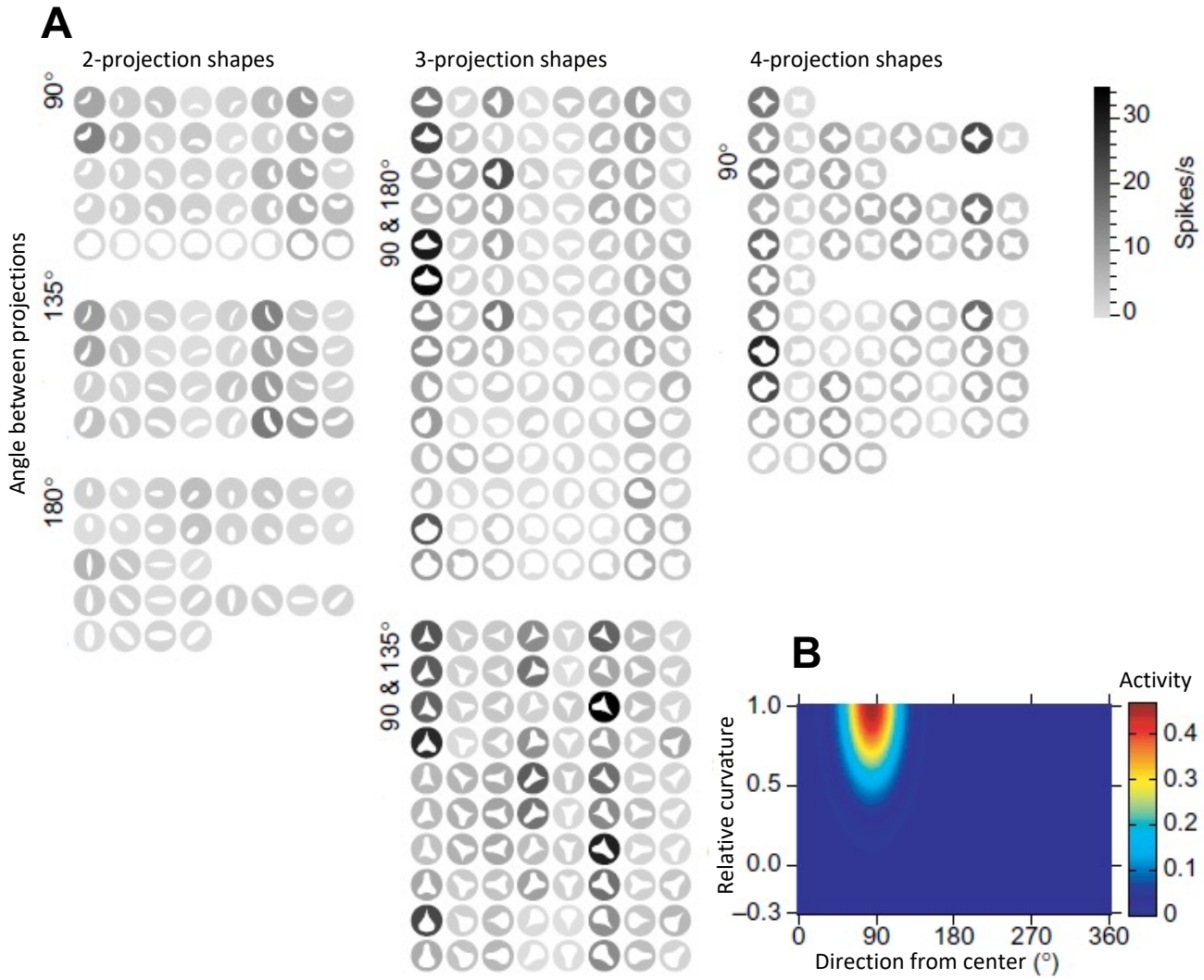


Figure 7.4. An examples of V4 curvature cells. a. The shapes within each circle indicate the stimuli presented to the receptive field, and the gray scale indicates the strength of the response to each stimulus. The cells prefer when there is a high curvature feature on the top of the shape (B) (from Ref. 4).

BOX *The Occipitotemporal Pathway for Visual Shape Processing*

Visual information projected onto the retia enters the primary visual cortex (V1) via the lateral geniculate nucleus, a relay nucleus in the thalamus. The V1 undergoes a variety of preprocessing for advanced processing. In terms of shape, especially contour processing, V1 contains neurons that selectively respond to line segments of a particular orientation presented in the receptive field, a small area of the visual field with a viewing angle of 0.2 to 0.5 degrees (in the vicinity of the central fossa). As mentioned above, these neurons are referred to as orientation selective neurons. These neurons represent/encode the orientation of the line segment in the receptive field by their firing activity.

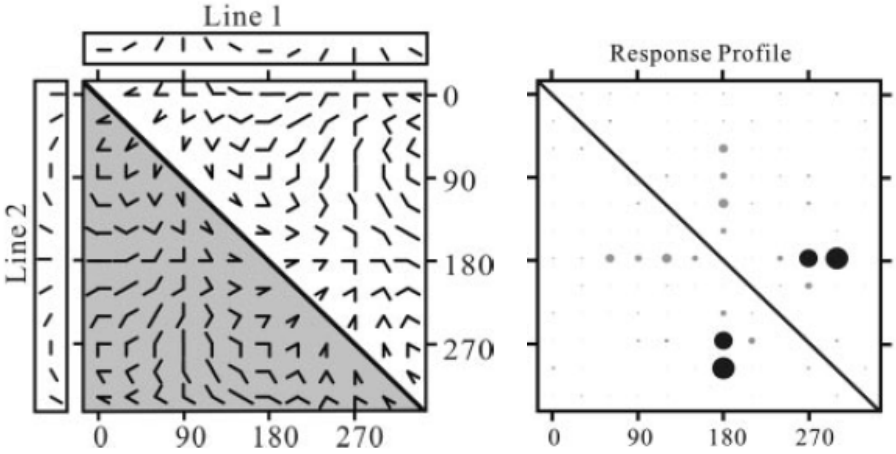


Figure 1. Angle-selective cells in area V2. (Left) Angular relationship of two line segments (Line 1, 2) presented in the receptive field. (Right) Response of an angle-selective cell in V2 to each stimulus. The size of the black circle indicates the strength of the cell's response (modified from Ref. 7).

Area V1 sends information directly to higher visual areas. Among these, the adjacent area V2 is strongly innervated and performs the next level of processing. In contour processing, each V2 neuron integrates the information encoded by the orientation-selective V1 neuron and responds to specific stimuli combining different orientations that are presented to a receptive field slightly larger than that of area V1 (Figure 1). These responses are thought to be involved in the representation of contour curvature.

Area V4 receives many inputs from area V2 and performs higher-level processes. As I said above, each neuron has a larger receptive field than the V2 neuron. They respond strongly to a shape presented in the receptive field, including its preferred curvature in its preferred direction (for example, the upper part of the shape). The outputs of the V4 neurons are sent to the IT (inferior temporal) area.

Although the IT cortex is divided into several subregions, neurons in the IT generally respond to complex stimuli with specific features (Figure 2A). Receptive fields extend over a wide area of the visual field, from a dozen to several dozen degrees. The neurons also fire to the stimuli they prefer, regardless of where they are presented within the receptive field and regardless of their size (Figure 2B). These properties are called position invariance and size invariance, respectively.

In summary, the feature to which a neuron responds becomes more complex in the visual shape processing in the occipitotemporal pathway, while the size of the receptive field, the area in which the neuron responds to its preferred stimulus, becomes larger. In particular, the response property to the preferred stimulus presented anywhere within the receptive field is prominent in IT neurons.

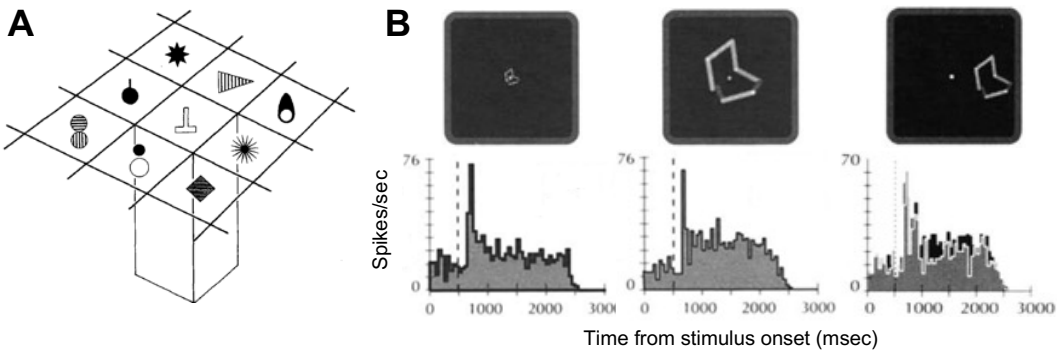


Figure 2. Response properties of IT neurons. A. Schematic diagram showing that neurons in each region of the IT cortex (a square of the grid) have their own preferred complex feature. Neurons in the longitudinal cortical layer structure have similar properties (from Ref. 8). B. Example of neuronal response (bottom row). The preferred shape (top row. The white dots indicate the fixation point) causes response independent of size and position. The black part in the rightmost histogram shows the response when the same figure is presented in the center (from Ref. 9).

BOX *Cortical V2 Area And Object-Centered Coordinate System*

An object-centered representation of a visual feature is a representation of a feature based on the center position of an object. For example, take a look at Figures A and B below. Let us assume that the black rectangles represent the entire field of view and that there is a white square within it. In Figure A, the square is to the right of the visual field center, and in Figure B, it is to the left. Consider expressing the boundary between black and white within the area enclosed by the gray dashed line. With respect to the center of the field of view, the two-dimensional locations of the two boundaries of A and B can be expressed as (right, center) and (center, center), respectively. In the early stages of visual processing, immediately after an image is projected onto the retina, the features of the image are inevitably represented in the retinocentric coordinate systems.

However, the eye is constantly moving. The retinal image changes dizzyingly accordingly. On the other hand, objects often remain still. Given this, it is more convenient to represent the location of object features with respect to the object's center. For example, the following expression is possible: "Can you pick up that teacup over there with the top chipped off?" As mentioned in the previous BOX, it is desirable to represent an object form invariant to position, that is, independent of its spatial location in the visual field, but the spatial position of features within the form should be preserved. Even a chipped teacup needs to be distinguished whether it is missing the top or the bottom. There is great merit in representing features based on their position from a reference point within the object, such as the center of an object.

So from what stage of the brain can this feature representation based on the object-centered coordinates be found? At least, we can see its beginnings in the secondary visual cortex (area V2), where Zhou and von der Heydt found what they called border ownership cells (BO cells)[10]. Since area V2 is the second stage of visual processing in the cerebral cortex, the receptive fields are still small and information is still represented based on visual field center coordinates. However, BO cells fire as if they know the direction the visual stimulus presented to their receptive fields.

Now consider the cases in Figures B, C, and D, where the area enclosed by the gray dashed line represents the receptive field of a single BO cell. If this were a neuron in the primary visual cortex or V1, it would be activated for B and D, but not for C. On the other hand, if it is a BO neuron, it will fire for B and C, but not for D. In other words, its activity depends not on the black-and-white pattern presented in the receptive field, but on the direction of the rectangle to which the boundary in the receptive field belongs, as if the neuron knew the location of the rectangle outside the receptive field.

I and Naoki Chiba proposed a model that calculates the properties of BO cells in area V2 from those of V1 neurons[11]. In the model, we focused on the fact that the activity of orientation-selective cells in area V1 is modulated when an orientation stimulus is presented outside but near the receptive field.

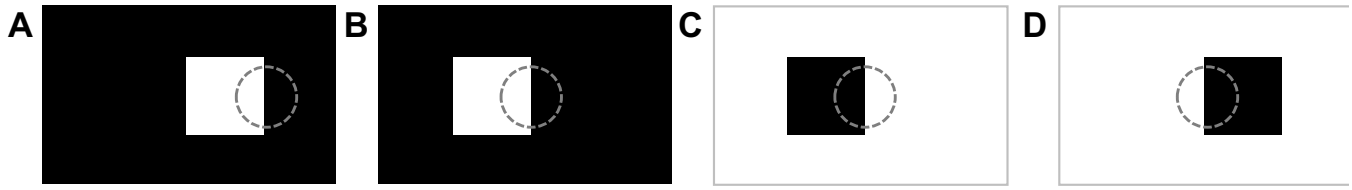


Figure . If a square is presented in the visual field (rectangle), how would you represent the contour feature enclosed by the dashed line? A and B differ in the position of the circle in the visual field, B and C differ in the black-and-white orientation within the circle, and C and D differ in the position of the square.

Detecting Contour Curvature

Our goal is to automatically decompose the contour of a shape from a given image into certain features. Contours include curves. For example, the shape shown in Figure 7.5A should be decomposed into coarse features with constant curvature as shown in Figure 7.5B.

First, let us consider detecting a part of the contour with constant curvature κ. A part of constant curvature κ is a part of a circle of radius 1/κ centered at point c, as shown in Figure 7.6a. Note that if we draw a circle of radius 1/κ from each point on this contour of constant curvature, all circles will intersect at point c (Figure 7.6B). The number of intersecting circles increases with the length of the section of constant curvature.

Conversely, if you want to detect a section of arbitrary curvature, simply draw circles of various radii from each point on the contour (Figure 7.7A). If a section of a certain curvature exists, it can be detected as the point where several circles of the same radius intersect (Figure 7.7B). Considering a circle as a vote, this intersection can be regarded as a collection of votes. No more than three circles of any other size can intersect (Figure 7.7C). This allows for the mechanical detection of sections of constant curvature in the contour.

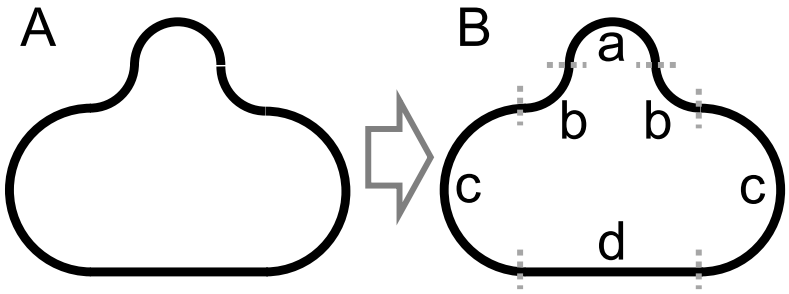


Figure 7.5. Our goal is to decompose the contour including curves (A) into rough features with constant curvature (B).

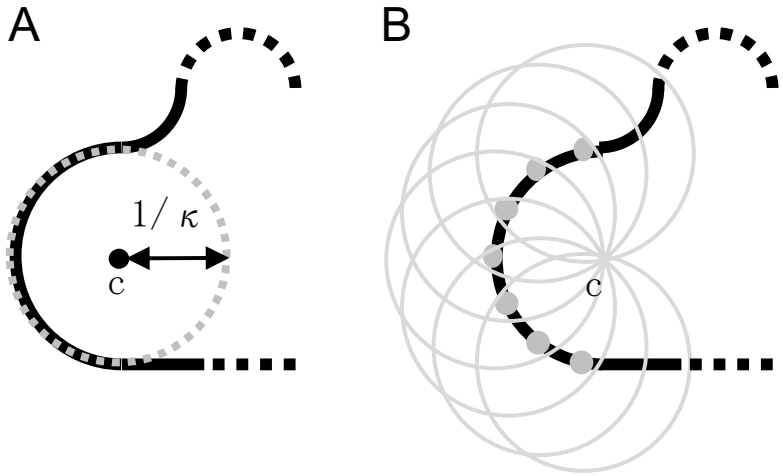


Figure 7.6. When circles of radius 1/*κ* are drawn from a point on the contour of constant curvature *κ* (A), they intersect at a point *c* (B).

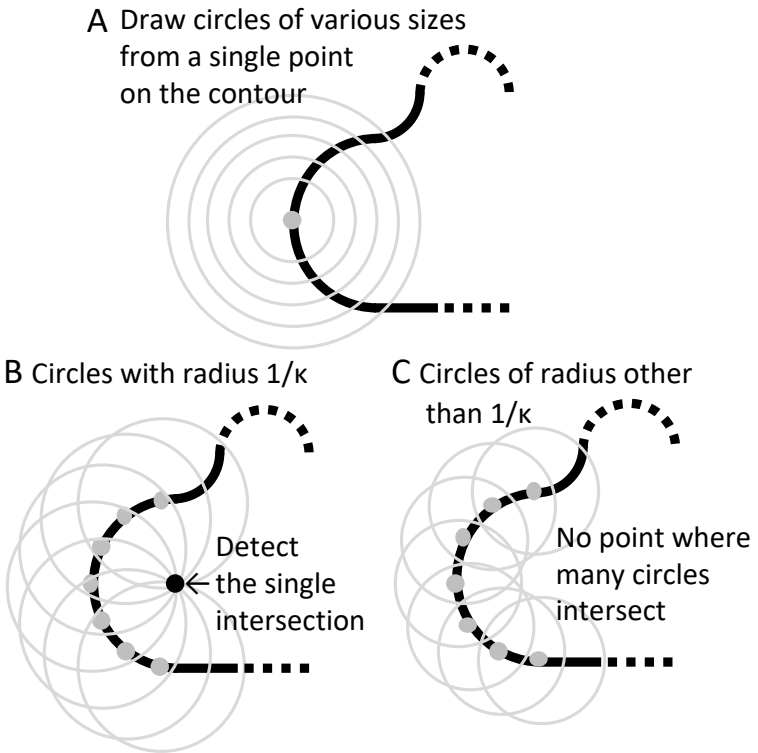


Figure 7.7. The principle in Figure 7.6 can be used to detect rough features of the contour = areas of constant curvature. The number of intersecting circles is proportional to the length of the contour.

Spherical Geometry, Unified Treatment of Lines and Curves

The previous section explained how to detect portions of constant curvature on a contour by drawing circles of various sizes from each point on the contour. Using this method, if there is a constant portion of curvature κ on a contour, drawing circles of radius 1/κ from different points on the portion will intersect at a certain point (this is called voting). By detecting this point, it is possible to determine what curvature portion exists on the contour. Using the location information of the original point from which those circles were drawn, we can also obtain information about the length, orientation, and location of that constant curvature section.

So what should we do if the contour contains straight portions, i.e., portions with zero curvature? One could use the Hough transform described earlier only for straight lines, but this seems not very elegant. A comprehensive method is preferable. To address this issue, we took advantage of the properties of spheres. That is, we used the fact that the radius of a circle tangent to a line drawn on a plane is infinite, while the radius of a circle tangent to a line drawn on a sphere is finite.

A point on the sphere has a corresponding great circle that cuts the sphere in half (Figure 7.8A). This point is called the pole. If we draw two great circles from the two poles, the two circles intersect at two points (Figure 7.8B left). Conversely, if we draw another great circle from one of the intersections, this circle connects the original two poles (Figure 7.8B, right).

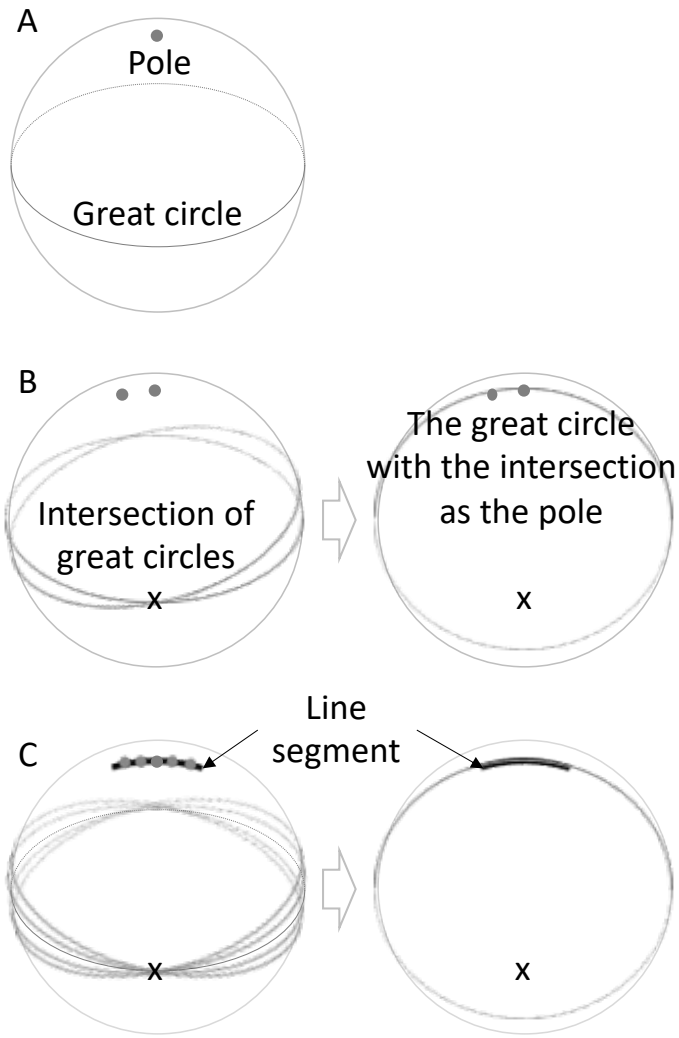


Figure 7.8. The relationship between poles and great circles is used to detect line segments projected on the sphere. A. One point on the sphere (a pole) has a corresponding great circle. B. Great circles drawn from two poles intersect at two points (left), and the great circle drawn from one of the intersections passes through the original two points (right). C. Great circles drawn from each point on a line segment (black line) projected on the sphere and intersects at two points on the sphere (left). If we draw a great circle from the point, it passes through the projected line segment.

Now consider the case where a line segment is drawn on a sphere. If we draw a great circle from each point on the line segment, they will intersect at two points on the sphere. The number of great circles that intersect at the points of intersection, i.e., the number of votes that come together, is proportional to the length of the line segment (Figure 7.8C, left). Conversely, if you draw a great circle from the intersection, it will pass through the original line segment (Figure 7.8C, right). In other words, a circle tangent to a line on the sphere is a great circle.

Using this method, the straight parts of the contour can be detected in a unified manner with the curved parts.

Segmentation of Contours by Great Circle-Small Circle Transformation

Let us summarize what I have discussed so far about the general characteristics of the contour of a shape, specifically, how to divide it into portions of constant curvature (hereafter referred to as curvature segments), which also include straight lines. First, project the contour of the shape onto a sphere sufficiently larger than the shape. From each point on the contour, draw circles of various radii, including a great circle. Then, find intersections where many circles of the same size intersect. The number of intersecting circles, or votes, corresponds to the length of the contour with the corresponding curvature. We called such a method the great circle-small circle transformation (in the actual model, we used local line segments of contours detected by orientation-selective cells in the primary visual cortex, like the brain, to save wiring in the neural circuitry, and memory in the actual program. Further wiring and memory savings can be made by using cells that represent combinations of different orientations in the secondary visual cortex, as described in BOX).

Furthermore, if we keep information about which point on the contour the circles were drawn from, we also know which part of the contour the detected intersection point corresponds to. Using this information, the point representing the curvature segment is moved from the point where the large or small circle intersects to the midpoint of the curvature segment (Figure 7.9A). The concave part must have negative curvature, which can also be determined by the position of the representative point in relation to the intersection of the circles. The number of intersecting circles (correctly, the central angle of the curvature segment) is also represented by the shading of the points. The representative points of these curvature segments were plotted based on their direction from the center of the entire shape (the center of gravity), following Pasupathy et al. (Fig. 7.9B).

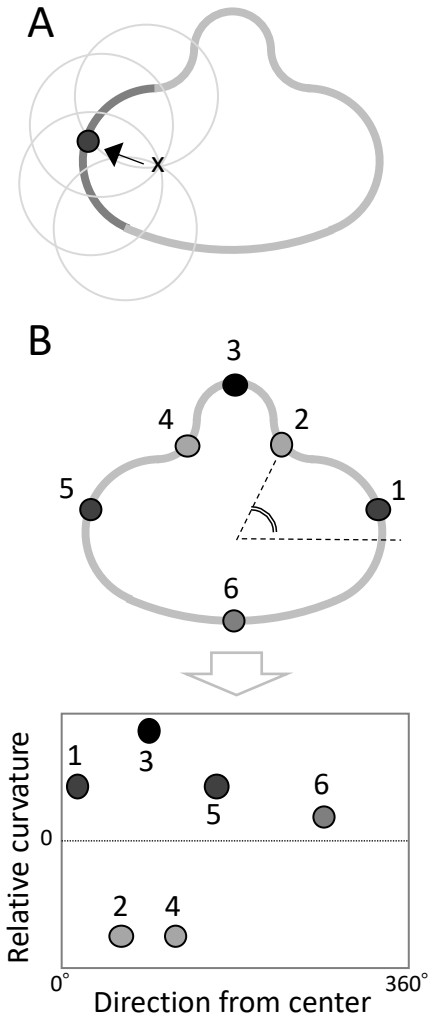


Figure 7.9. Plot of detected curvature segments in the direction from the shape center.

The model reproduced well the characteristics of neuronal responses in V4 cortex. It was also able to represent all curvature segments by the cell population and reconstruct the original shapes (Figure 7.10). In addition, the model detected global curvature features for a variety of shapes, including squares with noisy contours and interrupted circles, which were not tested in the physiological experiment of Pasupathy et al.

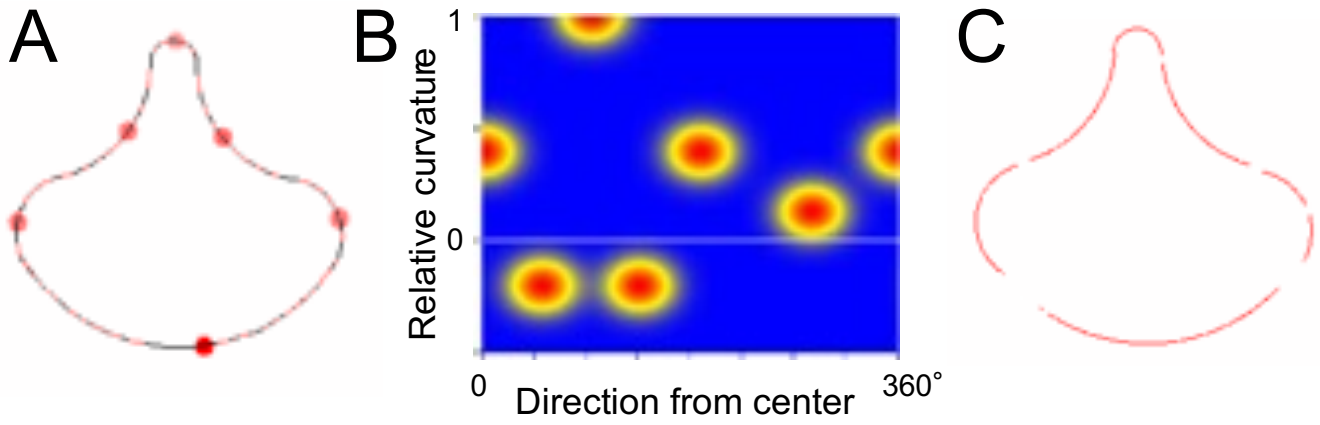


Figure 7.10. Authors' V4 field model. A. Input figure (black line) and representative points of detected curvature segments. B. Representative points of each curvature segment expressed in terms of direction (horizontal axis) and curvature (vertical axis) from the center of gravity of the entire contour. C. Input figure reconstructed from the curvature segments (modified from Ref. 1).

Symmetry Evaluation Using Properties of V4 Neurons

So far, based on physiological findings, we detected the global features of a shape contour, specifically, to divide the contour into sections of constant curvature. How, then, can we assess the symmetry of the shape, which is necessary for the complementation of the occluded figure? A clue to this question was also obtained from the experiments of Pasupathy et al.

We have already mentioned that Pasupathy et al. found that neurons in the cortical area V4 respond to specific curvatures present in certain directions of a shape. For example, some cells fire when a figure with a portion of intermediate curvature on the right side is presented in the receptive field, the area of the visual field to which the cell responds. The team further investigated these properties of V4 neurons and made a very interesting discovery.

They found that each neuron has a preferred orientation and curvature, but that its activity is modulated by other factors. For example, they found that the neuron shown in Figure 7.11 has a primary preference for the high curvature (i.e., projection) on the left side of the presented shape, but it is most active when the shape has a gradual negative curvature (i.e., concavity) on the counterclockwise side of its most preferred high curvature feature, which is not the case for the other curvatures (Figure 7.11A). The neurons were also unaffected by the features the clockwise side of the preferred feature. They classify neurons into two types: those that are affected by features on the clockwise side and those affected by features on the counterclockwise side[5].

Is it possible to use the properties of these V4 cortex neurons to assess symmetry? Sasaki et al. presented symmetric visual stimuli to humans and monkeys and used fMRI to measure which areas of the brain were activated. The results showed that the area around V4 was most activated[12]. This led us to investigate whether the above-mentioned properties of cells in V4 could be used to assess line and rotational symmetry.

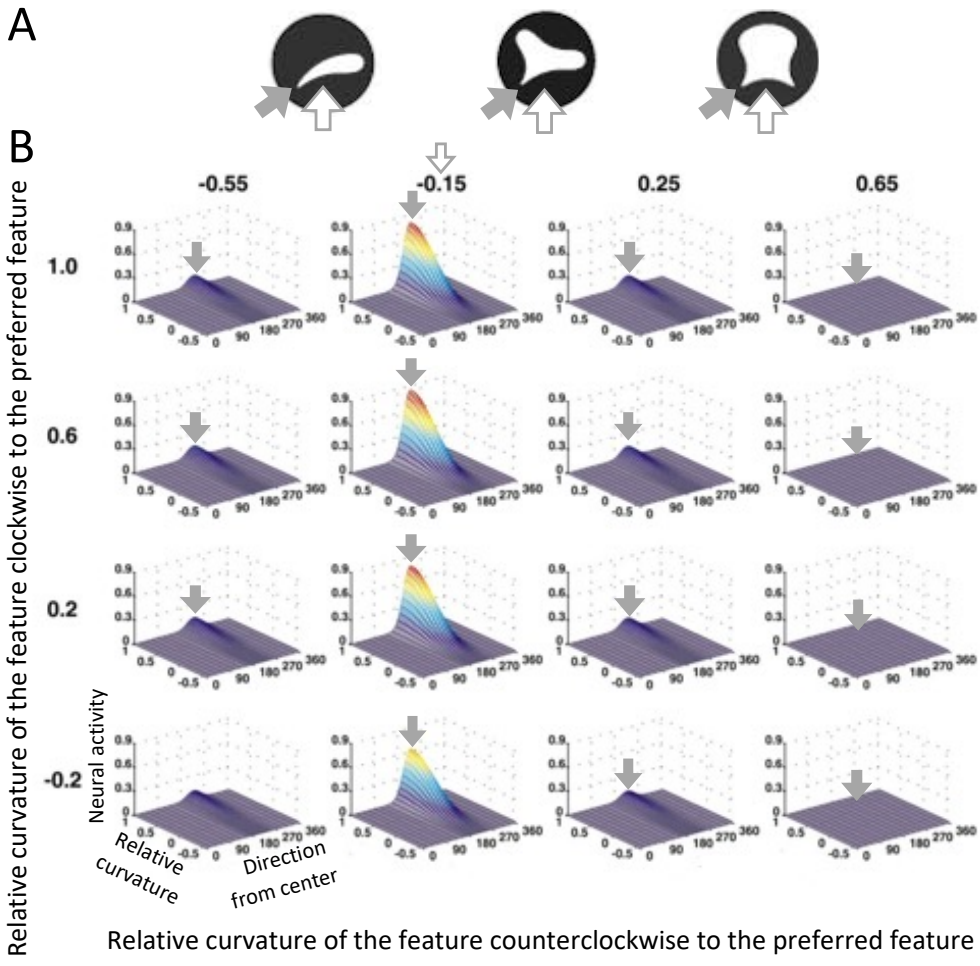


Figure 7.11. An example of how the activity of neurons in V4 is influenced not only by the preferred direction and curvature but also by the adjacent curvature. A. Stimuli to which this cell responds well. This neuron prefers the feature indicated by (→), but its response was highest when the feature indicated by (⇒) coexists. Each graph shows the neuronal activity for the direction, curvature of the feature within a shape. The cell responds best when there is a high curvature (1.0) in the 230° direction and a gentle negative curvature (-0.15) next to the counterclockwise side. Not affected by features on the clockwise side neighbors (modified from Ref. 5).

First Idea - Rotational Symmetry Detection by Compression

In the previous section, we described that neurons in V4, one of the visual areas of the cortex, are activated by the preferred curvature (degree of sharpness of a curve) and its direction in a presented shape, but their activity is modulated by adjacent curvature features, clockwise or counterclockwise. On the other hand, a transformation called the great circle-small circle transformation was used to detect and represent the constant curvature portion of the contour (called the curvature segment). How can these transformations be combined to successfully evaluate partial line and rotational symmetries and, based on these symmetries, complement incomplete shapes?

The fact that the neuronal response to a feature of a given direction and curvature is also affected by neighboring features indicates that the curvature segment feature is represented in a three-dimensional space of at least the direction (in shape-centered coordinates), the curvature that the cell primarily prefers (principal curvature), and the adjacent curvatures on the clockwise or counterclockwise side. Our primary idea was to determine if a figure has rotational symmetry by projecting this 3D space onto the two dimensions of principal curvature - adjacent curvature.

Now consider the rotationally symmetric figure shown in Figure 7.12a. Devide the contour is divided into curvature segments and label them as x,y,...(Figure 7.12B). Finally, plot them in a 3-dimensional space consisting of direction - principal curvature - adjacent curvature (Figure 7.12C).

There are a total of 12 curvature segments on the contour of this shape. This number remains the same when the adjacent curvature dimension in 3D space is compressed and the features are plotted in two dimensions: direction from the shape center and principal curvature (Figure 7.12D). However, if the figure has rotational symmetry, the number of features becomes three, as shown in the figure, when the directional dimension from the shape center in 3D space is compressed and the features are represented in the principal curvature-adjacent curvature space (Figure 7.12E).

We used the difference in the number of features depending on which dimension is compressed to evaluate partial rotational symmetry, i.e., the "goodness" of the shape due to the aforementioned Van Lear-style compression.

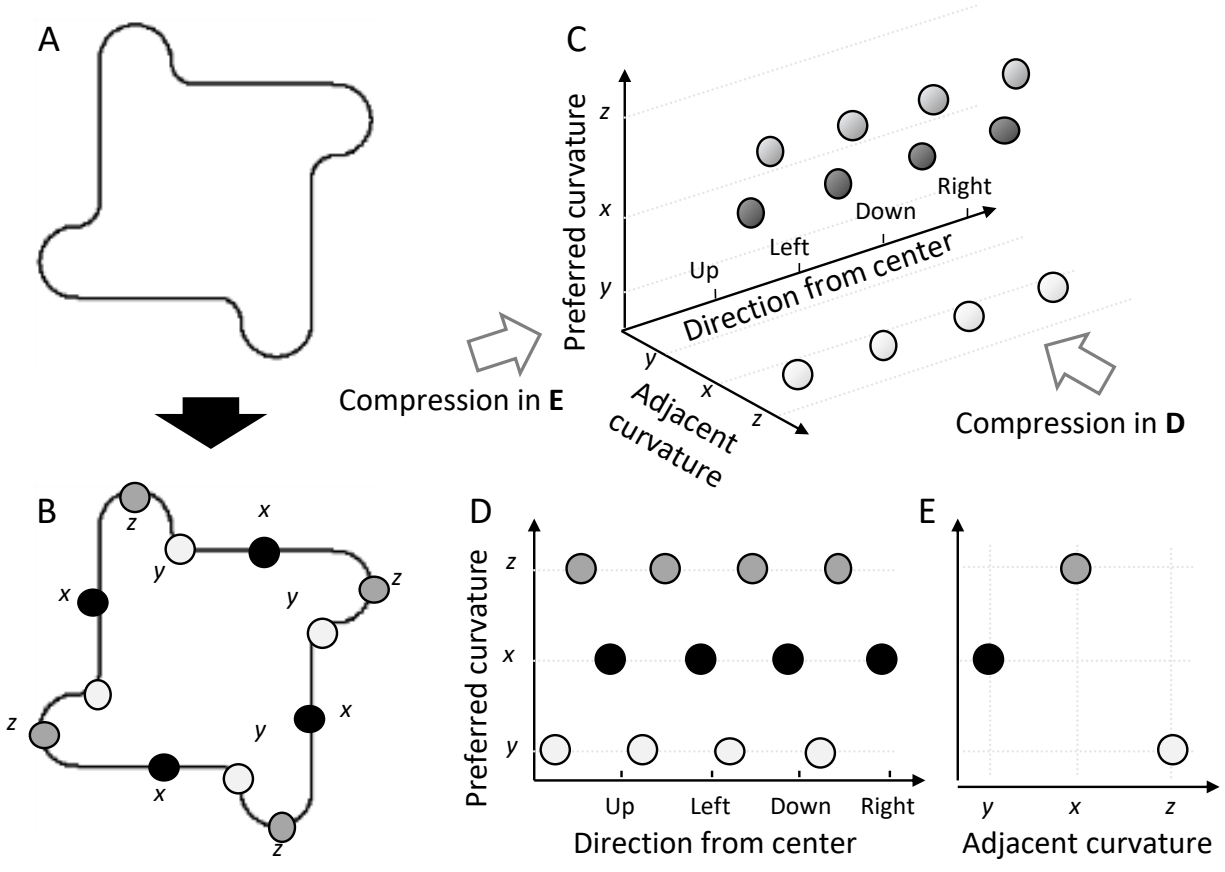


Figure 7.12. Evaluation of rotational symmetry. A. Original figure. B. Divided into curvature segments. C. Curvature features contained in a rotationally symmetric figure represented in three-dimensional space of direction from the shape center - principal curvature - adjacent curvature. The number of features is 12. D. The number of features is 12 when the direction of adjacent curvature is compressed. E. The number of features is 3 when the direction from the shape center is compressed.

Second idea -- evaluation of line symmetry by counterclockwise and clockwise adjacent curvature

On the other hand, how can we evaluate line symmetry? The picture in Figure 7.13 has 10 features. Representing these features in the principal curvature-neighbor curvature space does not reduce the number of features, unlike in the case of rotational symmetry. But once again, let us recall that neuronal activity in cortical V4 area is modulated by the neighboring "clockwise" or "counterclockwise" curvature feature. This means that there are two principal curvature-adjacent curvature spaces. That is, there is a clockwise space and a counterclockwise space. What does this mean?

Figure 7.13 shows a representation of the features of the figure in two-dimensional space of "clockwise type" and "counterclockwise type". We noticed that the feature placement patterns of the line-symmetric figure are identical in these two spaces. We use this to evaluate the partial line symmetry of the occluded figure, which is our second idea.

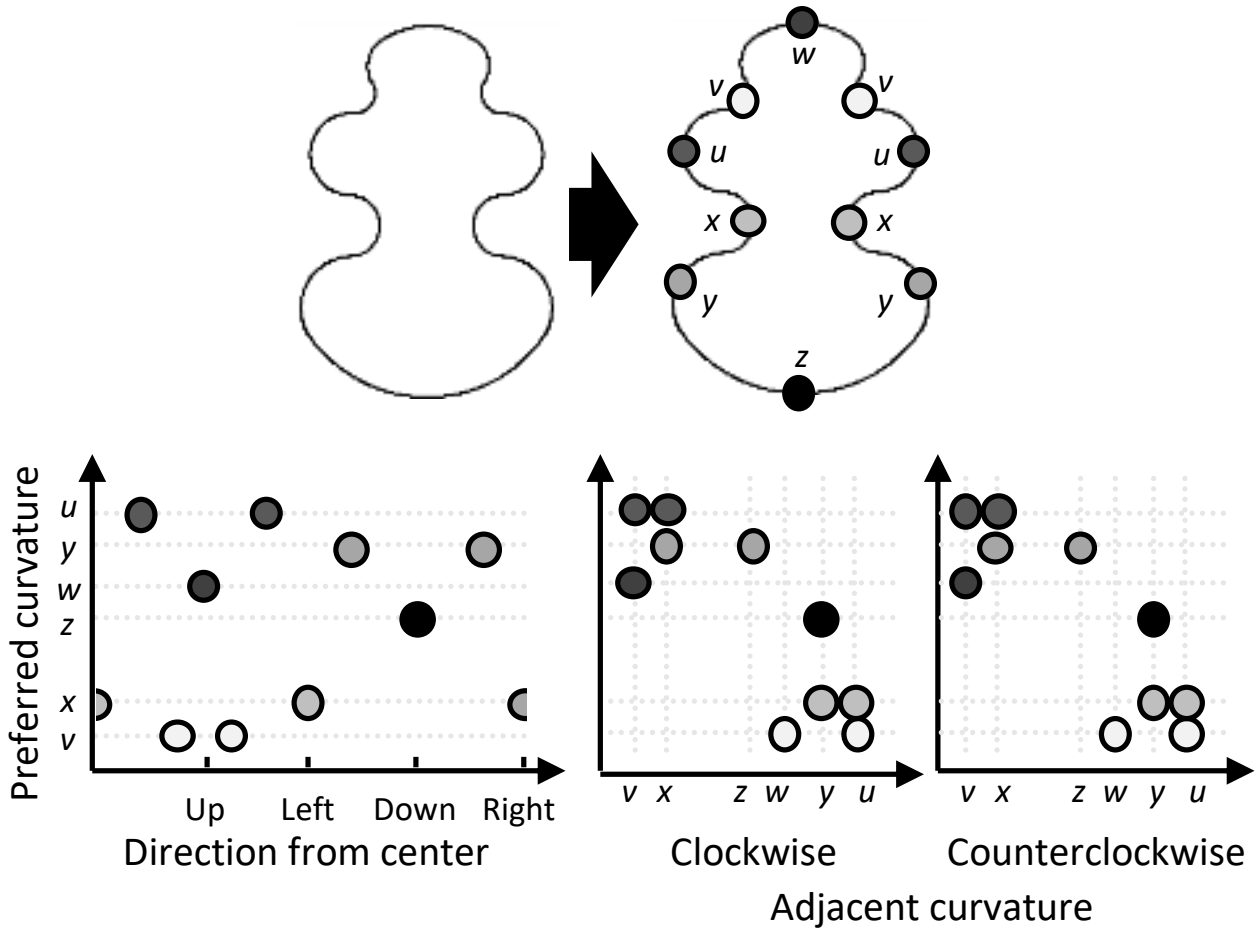


Figure 7.13. Curvature features in a line symmetrical figure are represented in the principal curvature-adjacent curvature space, the pattern is the same whether the adjacent curvature is in the clockwise or counterclockwise direction.

Evaluate Partial Symmetry and Complement

The current computational model only allows completion based on local continuity of contours. Based on the ideas described above, we have constructed a computational model that also allows completion based on the symmetry of the entire shape.

The process flow is shown in Figure 7.14. Without going into details, contours are first extracted from the input image. Next, using the T-junction feature of the contour as a cue, all possible contour segmentations are performed (divided into *i*, *j*, and *k* as shown in Figure 7.14 to obtain all possible combinations such as *i* + *j* and *j* + *k*). Then, the constant curvature portion of the contour, the curvature segment, is detected by the previously described great circle-small circle transformation. The detected curvature segments are simply drawn in Figures 7.14 for reasons of space, but in reality, as described above, they are represented in a higher-dimensional space that includes the dimensions of direction from the shape center, principal curvature, and adjacent curvature (clockwise and counterclockwise) and are complemented under the decision described in the next section.

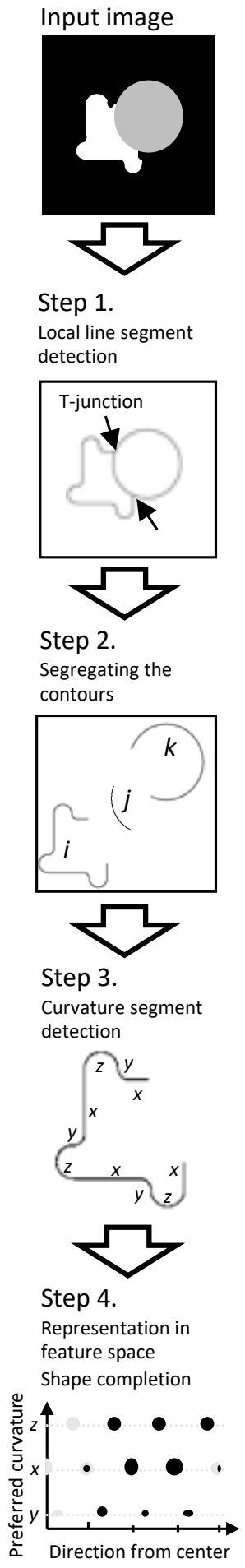


Figure 7.14. Process flow of amodal completion. Gray spots in step 4 are the completed features.

Computational Models that Allow for Multiple Interpretations

Even if part of the contour of a figure is missing, its partial symmetry is evaluated and complemented. Our goal is to construct a computational model that allows this. For this purpose, we divide the contour into segments of constant curvature. These features, which we call curvature segments, are plotted in a three-dimensional space that includes their curvature (principal curvature), their direction from the shape center (orientation), and the curvature segment next to the principal curvature (clockwise or counterclockwise adjacent curvature).

The computational model assumed that a figure has partial rotational symmetry if there are two or more rotational symmetry repetitions in a partially missing contour. In the example in Figure 7.15A, ten curvature segments were detected in the given incomplete contour, but in the principal curvature-adjacent curvature subspace, the number appears to be three. This means that there are more than three rotational symmetry repetitions in the given incomplete contour. Based on this judgement, the contour is complemented.

Partial line symmetry was determined by taking the logical product (AND) of the principal curvature-adjacent curvature space of the clockwise type and that of the counterclockwise type, and if there was even a slight overlap, partial line symmetry was determined to be present. The completion result was easy to obtain by taking the logical OR of the two (Fig. 7.15B).

On the other hand, completion based on local continuity of the contour was also obtained by extending the curvature segment until the contour closes (Figure 7.15C).

Whether completion based on contour symmetry or completion based on local continuity appears stronger depends on the shape. In some cases, one or the other is extremely dominant, while in other cases, the two completions compete with each other. Our model outputs both. Their relative strength is based on the method of Van Lear et al. described previously. That is, we assign a symbol to each curvature segment of the complemented figure and represent the entire figure by the symbols. If there is symmetry, the symbolic representation is compressed. The shorter this symbolic representation is, the simpler the figure is represented, leading to a dominant percept. Our model can perform this calculation automatically.

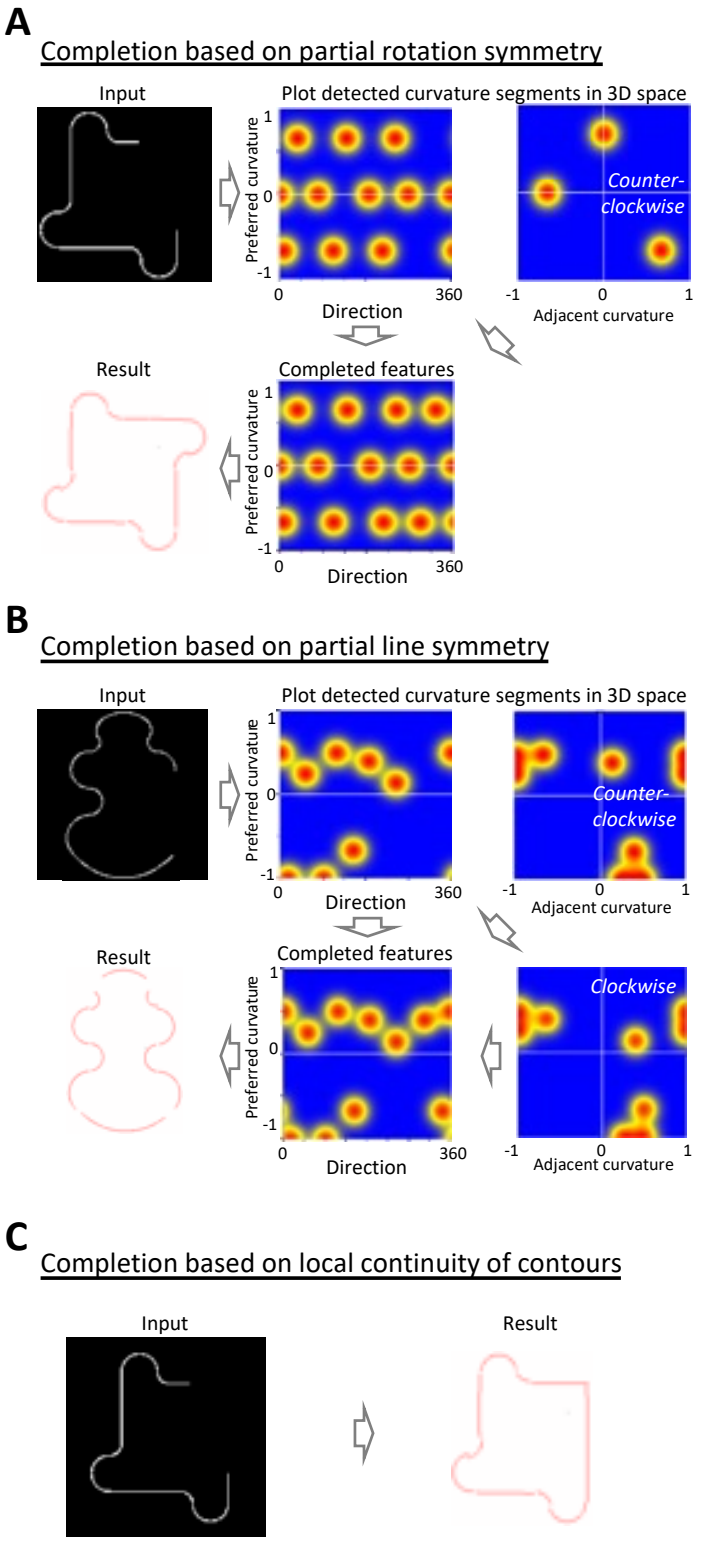


Figure 7.15. Example computation of amodal completion with our model. Here, curvature segments represented in the three-dimensional space of principal curvature-direction-adjacent curvature are shown in the two dimensions of the principal curvature-adjacent curvature plot; note the reduced number of curvature segments in A and the overlap between the represented curvature segments between the "clockwise" and "counterclockwise" plots in B. (modified from Ref. 1).

III. What Does the Computational Model of Amodal Completion Tell Us?

Creative Clues Come from Multiple Perspectives

In this chapter, we have described a computational model for solving the amodal completion problem, in which a part of the shape is occluded but is estimated and complemented from the unoccluded part. (i) A kind of voting method for the parameter space was used to detect rough features in the visible contour, i.e., parts of the contour with constant curvature (curvature segments). (ii) The curvature segments were represented in a three-dimensional space of segment curvature (principal curvature), direction (in which direction the segment lies in the shape), and adjacent curvature (curvature of the segment next to the principal curvature). (iii) When that higher-dimensional space is projected onto a particular subspace (in this case, two-dimensional), any overlap of segments and any collection of votes can be considered as partial symmetry in the contour, and only then is symmetry-based completion performed (completion based on local contour continuity is always performed).

The three stars of Orion have long been recognized in both the East and West, and are an easy pattern for anyone to notice. But why do they stand out so prominently when they are only three stars among the vast number of stars in the night sky? The Orion Triplets are second magnitude stars, not first magnitude, so they don't stand out just because of their brightness. To think concretely, let us imagine a machine that captures the three stars without any knowledge of the constellations.

The author considered the three dimensions shown in Figure 7.16. The three dimensions are the relationship of luminosity, distance, and orientation between any two stars. The relationship between the three stars in Orion, the two on the right and the two on the left, should plot very close together in this three-dimensional space. In each of the lower dimensions, there are many binary relationships that plot in similar positions. However, as the number of dimensions increases, it becomes rare for even just two points to overlap. On the other hand, if we consider the Big Dipper, the orientation relationship is not so constant, but in the subspace of luminosity and distance relationships, the plots gather in fairly close positions, which can be used. Although this constellation detection machine is only a thought experiment of the author, it is similar to the computational model of amodal completion in terms of representing and finding overlaps in higher dimensional space.

When we judge things from multiple perspectives and look at them from various directions, we sometimes find unexpected coincidences. These are clues to novel ideas and creativity, an idea we realize as we go through our experiences in life. One of the royal roads to increasing the number of axes of judgment is to consider relationships, especially change and continuity. The computational model of amodal completion presented in this chapter implements this intellectual aspect of ours.

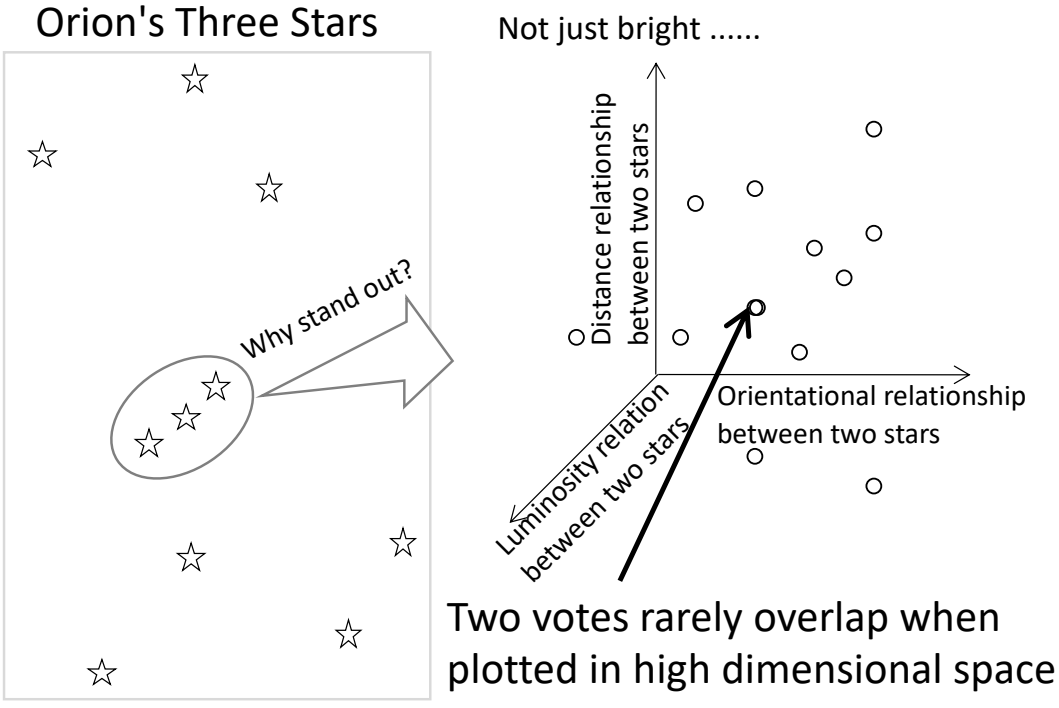


Figure 7.16. How do we detect three stars in Orion?

Does the Amodal Completion Model Perform Abduction?

Hypotheses can make a variety of predictions even if they cannot be directly observed. Hypotheses are obtained from incomplete information based on "beauty," "goodness," "simplicity," etc.

In our computational model of amodal completion, figures that cannot be directly observed are also obtained from incomplete contours as predictions. The simplicity of the figure was used as the criterion for completion.

Amodal completion, on the other hand, is a ill-posed problem whose answer cannot be uniquely determined from the given conditions. To solve it, the solver must make assumptions (constraints) about how the completion should be done.

It is difficult to know whether the completed shape is a hypothesis or whether the constraint for completion is a hypothesis, so let's make it clear here. The completed shape is a rule. It is a constraint to establish a consistent relationship between the features detected from the presented image. Abduction (hypothesis generation) is set up on the fly.

As a basis for abduction, our model represents the detected features in a high-dimensional space. By finding a dimension in the higher dimensional space that provides a simple representation and completing accordingly, we have consistently accounted for the apparently varying constraints depending on the presented figure, i.e. completion based on local continuity of the contour and completion based on an evaluation of the overall symmetry.

As mentioned in the previous section, such an approach would allow for flexible processing depending on the situation. The reason is that if you represent it in a higher dimensional space, there is a possibility of finding simplicity in unexpected dimensions. For example, the figure I considered with Manabu Kato (Fig. 7.17A) has no symmetry. But Fig. 7.17D looks more plausible than B and C, where corners and curves that do not exist in the visible contour are created in the completed area. If the local continuous constraint condition of the contour is applied in a straightforward manner, completion results like B and C will be obtained, but D takes into account the possible types of curvature, which probably contributes to its natural appearance. This is not possible with our model at the moment, but I think such a completion is achievable.

The architecture implemented in our amodal completion model, where features are represented in higher dimensional spaces and the discovery of dimensions in which the representation can be compressed, also means that the rules or constraints are hierarchical. That is, above the in-situ constraints of the complemented contour, there are specific modality-independent 'simplicity' constraints (which can be called meta-rules) that can be compared to the Pregnanz law of Gestalt psychology.

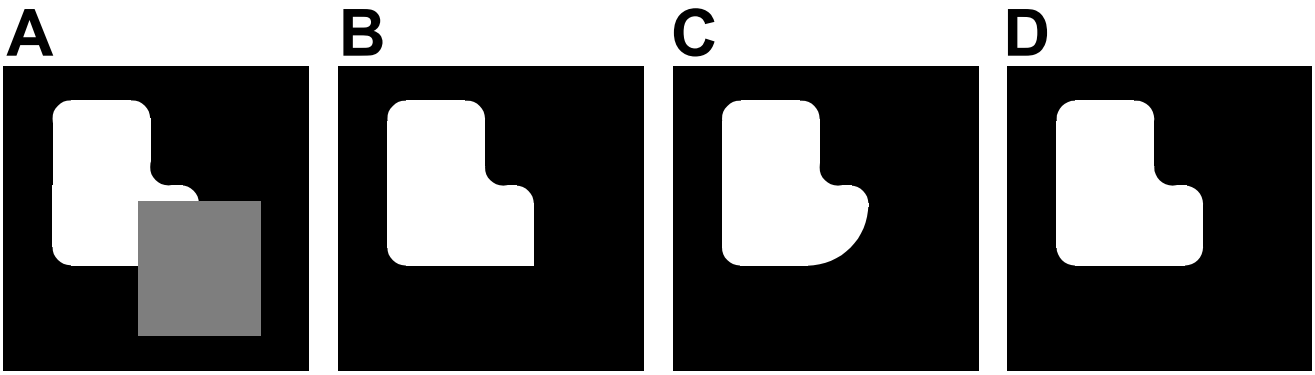


Figure 7.17. The Kato-Sakamoto occluded shape. A. The white area is perceived behind. B-D. How the white area is completed. B. Completion by straight lines; C. The curvature of the hidden areas is smooth and constant. D. The completed corner has the same curvature as the other corners.

Integration of the Amodal Completion Model and Complex Systems Theory

Neither the computational model for amodal completion discussed in this chapter, nor the line detection model using Hough transform discussed in the previous chapter, contain any dynamics of complex systems discussed in Part I and in the Appendix. How can we integrate the arguments discussed in Part II with complex systems theory? For example, if we were to use non-linear oscillators in the authors' computational model of amodel completion, what would the use be?

Non-linear oscillator is a general term for oscillators described by nonlinear differential equations. Under certain conditions, oscillators oscillating at different periods and phases (positions within a period) may synchronize through interaction. This is a self-organization phenomenon called mutual entrainment. Let's consider again what the advantages of using this phenomenon in computational models of the brain are.

The author believes that there are advantages to using phases. Once again, consider the stability or instability of a quantity, as shown in Figure X, using the metaphor of a marble. If the marble is at the bottom of a trough, the effects of any external action on the marble will eventually cancel out. That is, if the marble is at the bottom of a trough, if some external action is added to it, the effects of that action are ultimately cancelled out and the marble remains stationary at the bottom of the trough. In this sense, the state of the marble at the bottom of the trough is stable but insensitive to external action. In contrast, if the marble is placed on a flat surface, the marble is sensitive to external action but wanders on the flat surface and is not stable. Now consider placing the marble on a perfectly flat star. Preferably, the star should be like a ring in the two-dimensional world. If we apply an action to the marble from the outside here, the marble will roll all the way in one direction because there is no valley. But this time, because it is on a ring-like star, the marble will move over the star and return to its original position. In other words, it is in a sense stable, in that it is sensitive to action from the outside but cannot escape from the star.

If the interactions between the non-linear oscillators are carried out over such a ring, then one would expect that, under certain conditions, each would settle into a good overall phase relationship without being trapped in meaningless troughs or going to infinity and beyond. Of course, the actual calculation will not work out so desired, but at least we are trying to use the advantages of the oscillator that are not available elsewhere. If the interaction between the non-linear oscillators were to take place on such a ring, one would expect that under certain conditions they would settle into a good phase relationship as a whole, without each being trapped in a meaningless trough or going to infinity and beyond. Of course, the actual calculations would not work out so well, but we are trying to exploit this unique oscillator property.

The computational model created for amodal completion detected rough features of the contour as curvature segments and represented them in a three-dimensional space of curvature, direction from the shape center and adjacent curvature. However, the actual computational difficulties were in the areas between curvature segments, where the curvature varied continuously. When trying to connect segments smoothly, we had to explicitly use dimensional quantities such as the distance and direction between curvature segments and the orientation and size of the curvature segments themselves. We believe that if we can successfully use non-linear oscillators, we can smoothly connect curvature segments and reconstruct the shape in a consistent manner without having to explicitly express these dimensional quantities.

What Are the Mechanisms of Creativity?

This book, a slightly quirky brain science book for the general public, comes to a close here.

We are by no means a predetermined being. When faced with situations involving the unknown to varying degrees (what we call the indefinite environment in this book), we can say 'Yes!' and cope with them with creative ideas, large and small, and manage to live with them. The aim of this book was to see how far we could get into the creativity of humans and other living creatures. Was this book able to capture even part of the brain mechanism of creativity?

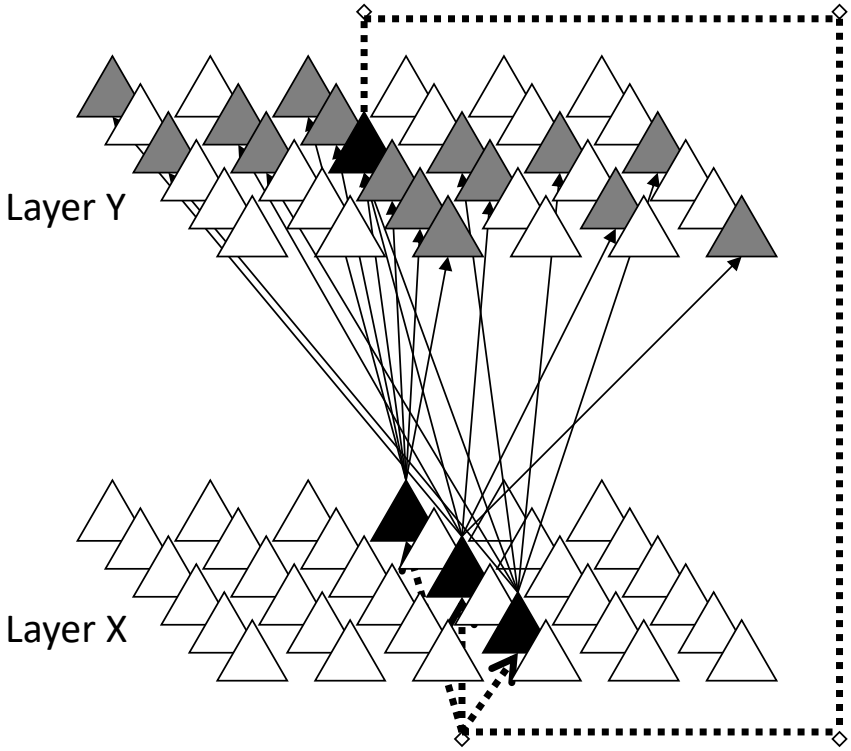


Figure 7.18: Image of the simplest neural circuit mechanism of creativity.
Triangles represent neuron-like units. Wiring between units has been omitted to make the schematic diagram easier to read. Each unit encodes different higher-dimensional information, like the neurons in the prefrontal cortex and V4 cortex discussed in this book.

The units can oscillate as required and synchronize each other. Mutual entrainment ensures that the consistency of the information encoded by each unit is pursued without being trapped by false local minima. As a result, information encoded in high-dimensional space is represented in low-dimensional space, as discussed in the section on the amodal completion problem. In the process, units that do not encode much information are also activated or incorporated into synchronous clusters to compensate for missing information. Even when units are synchronized, there may be phase gradients. In such cases, as in the case of slime moulds, each unit can 'know' its position in the entire synchronized cluster.

The wiring from the units of layer X to layer Y is a vote to the parameter space. There is a part-whole inversion relationship between the information encoded in layer X and the information encoded in layer Y. Elements in layer Y, where votes are collected, vote back to layer X. Through the loop between X-Y, a higher level of consistency between part and whole in the represented information is pursued. The loop structure itself may behave as an oscillator, causing an entrainment between individual loops; even in the voting from Y to X there is an 'inversion of part and whole'. As a result, layers X and Y act as constraints on each other.

When I was in the first grade of primary school, I loved Toyotomi Hideyoshi, who rose from the lowest class to a great success and ended a long period of civil war. Looking back on why I loved him, I think it was because he solved various difficult problems with wisdom that others could not imagine. Take, for example, the overnight castle of Sunomata. In order to attack neighboring Mino, a frontline base was needed. However, a suitable location for a frontline base is always exposed to enemy attack. To clear this difficulty, half-assembled building materials were poured from the upper reaches of the river and built on-site at once. The episode of solving a difficult problem with an idea that could be regarded as a forerunner of the modern prefabricated construction method was poignant even to a child's mind.

Good problem-solving, not only in the episodes described above, is a product of creativity. Problem-solving situations usually involve seemingly incompatible requirements that have to be met. For example, a forward base needs to be built, but when you try to build it, you are attacked. A good solution meets all requirements neatly. The solution of pouring prefabs down the river and building them all at once fulfils both the requirement to build a forward base and the requirement not to be attacked by the enemy.

I have been studying the brain from the perspective of complex systems science to address this major problem. Complex systems science, also known as self-organization theory, deals with how ordered states such as spatiotemporal patterns are generated autonomously. Considering the brain and the various problems that the brain solves from this perspective, I expected to be able to elucidate the mechanisms by which the solutions are generated autonomously in the brain.

Specifically, many problems, such as the problem of how to understand the three-dimensional world from the two-dimensional image of a retinal image, do not have enough information to determine recognition and behavior based on externally given cues alone. The main message of this book is that the brain and nervous system as a complex system has mechanisms to generate this missing information itself. One of the mechanisms considered is a phenomenon in complex

systems called mutual entrainment of non-linear oscillators. This phenomenon was expected to produce a coherent relationship between immediate demands. Indeed, in order to compensate for missing information in a self-organizing manner using non-linear oscillators and other elements, a set of rules, or constraints (also known as hypotheses), is needed to guide the desired relationship between oscillators and other elements. However, a serious problem arises when considering these constraints.

First, there is the question of whether local rules are sufficient. Let us look back at the problem of amodal completion, which completes the invisible part of a contour. Let us consider a rule where the orientation of the contour segment changes smoothly between the units encoding the local segments of the contour. In fact, looking at various forms, there are many cases where this rule/constraint holds, i.e. all segments are smoothly connected. However, there are also many cases where symmetry-based completion is preferred. Symmetry is something that can only be seen by looking at the shape as a whole. The symmetrical shape of the Goryokaku castle in Hakodate, Japan, can only be seen on the ground by surveying the castle's perimeter, but is immediately apparent on aerial photographs. This suggests that a coherent relationship between elements is not simply a series of good local relationships. Holistic constraints are inherently necessary. We believe that the method of voting in parameter space, which inverts the part and the whole, is one of the reasons why the computational model worked so well. A more serious problem was the problem that essentially the constraints themselves had to be created as well. In fact, there are a number of ill-posed problems in which the constraint changes depending on the situation. The amodal completion problem included completion rules based on local continuity of contours, rotational symmetry and linear symmetry, the dominance of which varied depending on the figure. Looking back at everyday experience, there are all sorts of relationships that are overall orderly and consistent. However, it is not possible to find good relationships when one is struggling. Creative inspiration may be about discovering where the consistent relationships are. Our amodal completion model enables symmetry-based completion by plotting feature segments in a high-dimensional feature space and evaluating sub-dimensions such as repetitive structures that provide a compressed representation of the entire shape.

In our model, we used a pre-defined process because the subspace in which symmetry is assessed was known in advance. If there were a mechanism that could autonomously bring up hidden structures in this unexpected subspace of higher dimensional space, it would be an important part of the creativity mechanism.

Figure 7.18 brings together all that has been discussed in this book and provides a schematic, albeit very crude, representation of what I imagine to be the neural circuit mechanism of creativity.

This is as far as the author, a scientist, can responsibly discuss the brain mechanisms of creativity. However, I feel that something important is still missing. In the final chapter, I will consider this not as a scientist but as a human being and discuss issues that cannot be contained in a scientific paper at this time.

BOX. *Can the Mirror-Image Confusion Cells in the Area IT Be Explained by the Properties of The V4 Neurons?*

The computational model of amodal completion constructed by the authors is based on the neuronal properties of cortical V4 areas and represents rough features/curvature segments (contour areas with constant curvature) in a curvature-direction-adjacent curvature space (three-dimensional space of curvature of each segment, direction from the shape center and curvature of adjacent segments). There are two types of neurons: those affected by clockwise neighboring curvature and those affected by anti-clockwise neighboring curvature. The model therefore has two curvature-direction-neighboring curvature spaces, clockwise and anti-clockwise. The model used the overlap of curvature segment representations between these two spaces to assess line symmetry. If such a process is carried out in the real brain, it implies that these two spaces are integrated in some way.

By the way, children often show mirror-image confusion when reading and writing letters. For example, they confuse the letters ‘b’ and ‘d’. Some neurons in the inferior temporal cortex, which is the higher area of the area V4, have responses that reflect this mirror-image confusion. Such responses may be the result of a misintegration of the rough features of the shapes represented within the curvature-direction-adjacent curvature space of the two clockwise and anti-clockwise types.

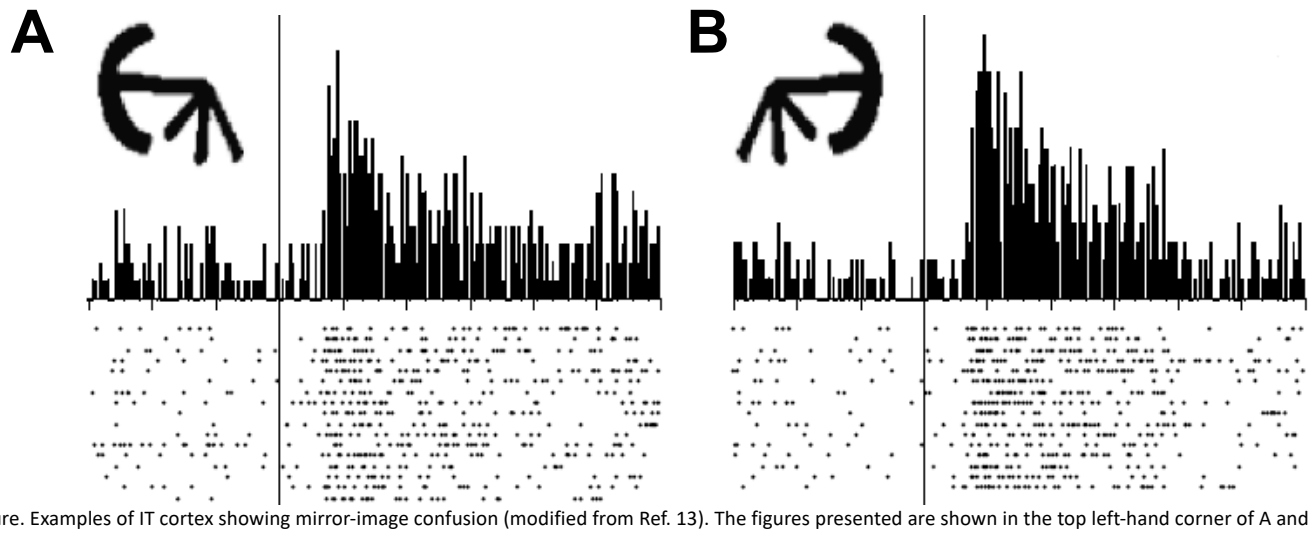


Figure. Examples of IT cortex showing mirror-image confusion (modified from Ref. 13). The figures presented are shown in the top left-hand corner of A and B respectively. The bottom line shows each trial and the points within each line show the firing of the neuron. Above them are the histograms of firing. The vertical line is the timing of the presentation of the diagram, the horizontal axis scale is 200 ms.

BOX *How to Pack Higher Dimensional Information into the Three-Dimensional Structure of the Cortex?*

In the authors' computational model presented in this chapter, what we have called curvature segments as rough features of shape contours were represented in a high-dimensional space with curvature, direction, neighboring curvature and, for computational requirements, several more parameters. However, can such a higher dimension be packed into the three-dimensional structure of the neocortex of the brain?

Although the functional structure of the cortex is not yet detailed, the primary visual cortex (V1 cortex) is almost the only area whose functional structure has been well studied. Advances in imaging techniques have revealed considerable detail about the spatial structure of the V1 cortex and the types of stimuli to which it responds. Methods that use electrodes to examine the orientation selectivity of neurons (i.e. which directional line stimuli they prefer) have only revealed a horizontal arrangement of cells in the V1 cortex with close receptive fields (areas of the visual field to which cells respond) but slightly different orientation selectivities. In contrast, imaging shows at a glance how these cells line up horizontally in the cortex. However, orientation is not the only stimulus to which neurons in V1 cortex respond. Well-known ones include which eye is stimulated, left or right (ocular dominance), which wavelength of light they respond to (wavelength selectivity) and which spatial frequency (i.e. the repetition period of the grid stimulus) they respond to (spatial frequency selectivity). Of course, there is also the two-dimensional location of the stimulus in the visual field. The higher dimensional space composed of these is cleverly embedded within the three-dimensional structure of the cortex.

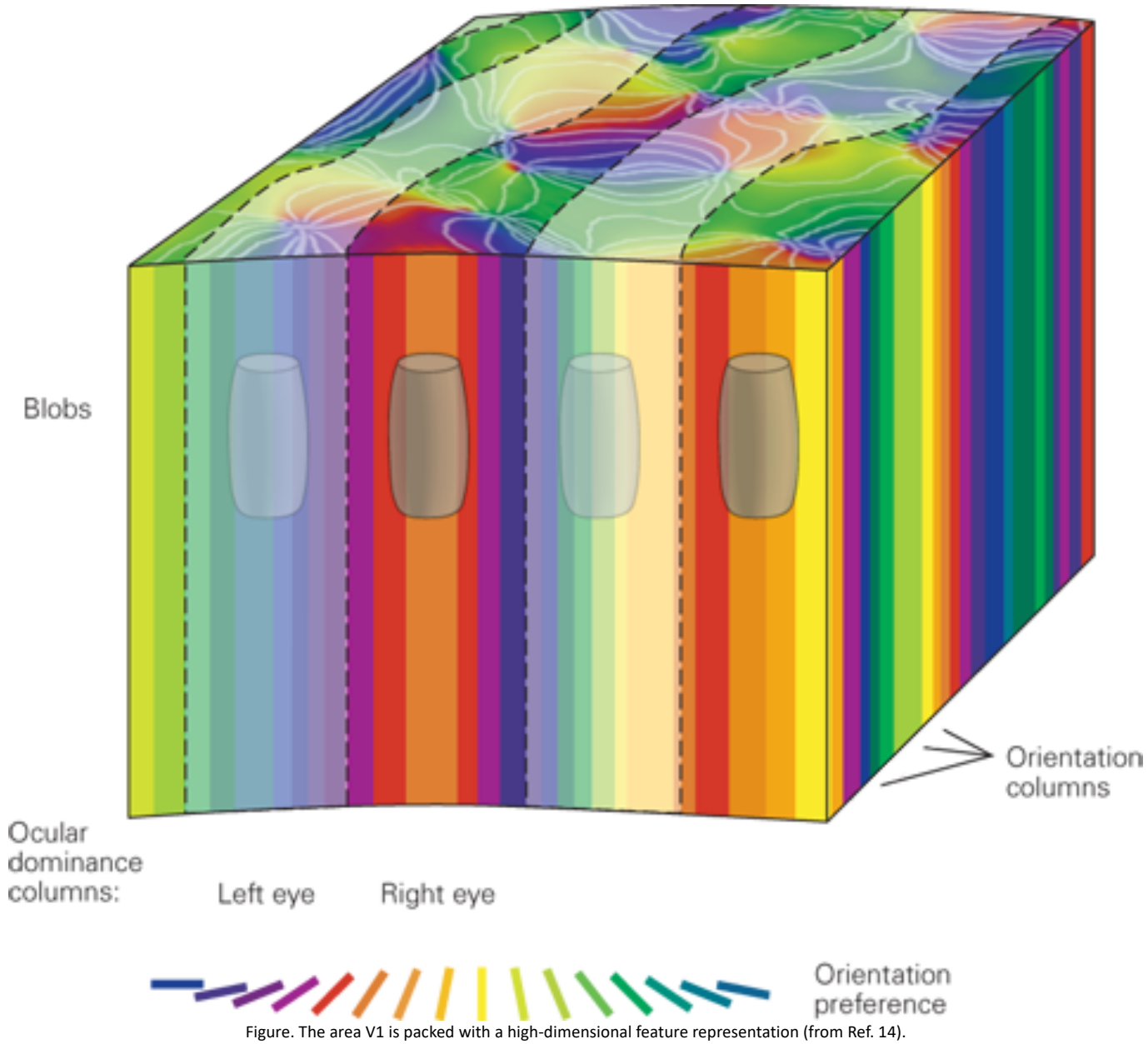


Figure. The area V1 is packed with a high-dimensional feature representation (from Ref. 14).

References

1) Sakamoto K, Kumada T, Yano M. A computational model that enables global amodal completion based on V4 neurons. *Lecture Notes Comput. Sci.*, 6443:9-16 (2010)
2) Fukushima K. Neural network model for completing occluded contours. *Neural Networks,* 23:528-540 (2010)
3) Van Lier R, Van der Helm, P, Leeuwenberg E. Competing global and local completions in visual occlusion. *J. Exp. Psychol.*, 21:571-583 (1995)
4) Pasupathy A, Connor CE. Population coding of shape in area V4. *Nat. Neurosci.*, 5:1332-1338 (2002)
5) Pasupathy A, Connor CE. Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.*, 86:2505-2519 (2001)
6) Ito M, Komatsu H. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J. Neurosci.*, 24:3313–3324 (2004)
7) Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathways of the macaque cerebral cortex. *J. Neurophysiol.*, 71:856-867 (1994)
8) Logothetis MK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.*, 5:552-563 (1995)
9) Zhou H, Friedman H, von der Heydt R. Coding of border ownership in monkey visual cortex. *J. Neurosci.*, 20:6594–6611 (2000)
10) Sakamoto K, Chiba N, Yano M. A physiological model for bottom-up object-centered representation in the visual cortices. The 15th Annual Conference of the Japanese Neural Network Society, 95–96 (2005) in Japanese
11) Sasaki Y et al. Symmetry activates extrastriate visual cortex in human and nonhuman primates. *Proc. Natl. Acad. Sci. USA*, 102:3159-3163 (2005)
12) Rollenhagen JE, Olson CR. Mirror-image confusion in single neurons of the macaque inferotempral cortex. *Science*, 287:1506-1508 (2000)
13) Kandel ER et al. *Principals of neural science (5th)*. McGraw-Hill, New York (2012)

Final chapter. Seeking Further Sources of Creativity

So far in this book, the emergent and creative aspects of the brain's information processing have been reviewed from the perspective of complex systems theory of biological systems. In complex environments, determining cognition and behavior from the slightest cues is not easy. In many cases, it is an ill-posed problem, i.e. one that cannot be uniquely answered. This is where complex systems theory of biological systems has room to become involved. In complex systems, ordered states such as patterns can be generated autonomously, and applying this to information processing in the brain could lead to new techniques for creating what is missing in ill-posed problems. What actually needs to be created is often a coherent relationship between the things involved. Coherent relations appear as a phenomenon, as a synchronic order.

In this book, I have shown that coherent relations emerge as a synchronous order when something is generated, from simple organisms such as slime molds to action planning in the prefrontal cortex of monkeys. However, coherent relations vary from situation to situation. There needs to be some binding on what aspects are coherent. We call them constraints. There are different levels or hierarchies of constraints, but when considering the true autonomy of biological systems, we must consider the mechanisms that generate lower-level constraints according to higher-level constraints. The computational model of visual amodal completion was oriented towards this problem, albeit rudimentary. In the final chapter, I go beyond my position as a scientist to discuss what the top-level constraints or norms (which might also be called meta-rules) look like.

The Ultimate Robot That Turns against Humanity

When I was a kid, there was a superhero anime called "Neo-Human Casshern." Casshern's father was a talented robotics scientist. One night, lightning struck his home and laboratory, and a robot that was almost complete accidentally started up. The robot got up and came from the laboratory to the scientist's bedroom. When the scientist ordered, "Go away!", the robot calmly retorted.

"I don't obey human orders, you humans have created and used machines. But this time, it's different. Now I will use you. I will rule over you humans!"

Figure. Bryking Boss.

This robot is so amazing that it cannot be broken even when the army fires cannons at it. Instead, it instantly wiped out the army with its beams. Not only that, he called himself Bryking Boss, created robots to serve him (he even built a robot factory), and organized a troop called the Andro Army. Then, He set out to conquer humanity with his army. Casshern volunteered to become an android himself to defeat the Andro Army. This is how the story began.

This robot is, in a sense, the ultimate robot. It has many capabilities that are far beyond the reach of current technology, in recognition, motor control, and memory. So what went wrong?

There is no doubt that Bryking Boss feels anger and dissatisfaction to the point of plotting to conquer mankind. Since the robot behaves in a very autonomous manner, he must have an internal desire to take action on his own. On the other hand, even though he started to malfunction due to lightning, he was created by the doctor to be useful to human society in some way, Apparently, there is a problem with either the processing to meet this demand from human society, the processing to satisfy Bryking Boss's internal desires, or the processing or rules to make a good compromise between these two.

In the stories that follow, we do not see Bryking Boss gradually calming down and settling the spears against humanity. It seems that there is no natural reconciliation between the demands of society and his internal needs. Nor is there any indication that he himself tried to come to terms with humanity. Even if the lightning had altered the processing of human society's needs and internal needs, he himself might have made some effort if the higher norms or meta-rules for compromises between them were preserved.

Such a story may seem like a pipe dream, but these days we have cleaning robots that start charging themselves. In the future, there may even be robots that shoplift batteries at the supermarket when the battery is about to run out. It may not be long before the problem of reconciling the demands of human society with the internal needs of robots becomes apparent.

The "Three Laws of Robots"

Living organisms face many ill-posed problems in their daily lives, in which the answer cannot be uniquely determined, in terms of cognition and movement, and solve them using some kind of rule (constraint). When it comes to the ultimate robot that rebelled against humanity, introduced in the previous section, it seems that such ill-posed problems can be solved without difficulty. His problem seems to be a problem of processing or rules to meet the demands of this human society, to satisfy his own internal needs, or to find a good compromise between these two.

The most famous rules that robots should follow are the Three Laws of Robotics, proposed by science fiction writer Isaac Asimov in the year 2058 (!)[1)].

The First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
The Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
The Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Asimov's Three Laws of Robotics are a good idea. They prioritize the rules so that there is no conflict between the demands of human society and the robot's own internal needs. However, it may not always be the most desirable to simply follow these three laws.

For example, let's say that care robot A is caring for Mr./Ms. B, a human. If robot A does not provide care, B's health condition will worsen and he/she may die. One day, because of B's worsening condition, he/she begins to try and destroy A. To stop the destruction, it needs to do a lot of harm to Mr./Ms. B. What decision will A make in this situation?

If we were to superficially apply Asimov's Three Laws of Robotics, robot A would be destroyed at the mercy of B. This is because A cannot stop B's destructive behavior, as the first law states that robots must not harm humans. However, A will not run away from B in order to comply with the third law, because if A runs away, B may die. The first law also does not allow A to overlook the danger to the lives of its human users, and it takes precedence over the third law.

In the above example, it seems like there is a better solution. However, as long as the rules are fixed, it does not seem easy to get out of this impasse. Such an impasse can always occur. No matter how broadly applicable a rule is, there will always be situations where it cannot be applied. This is because we cannot predict everything that may happen. In this book, we have referred to such an inherently unpredictable environment as an indefinite environment. There is no definitive answer for how to deal with an indefinite environment. However, we can list some things that are necessary.

To Face Indefinite Environments

The term “indefinite environment,” which appears frequently in this book, refers to the real world in which we live as an environment in which we do not fully know in advance what will happen. Note that the indefinite environment here is not the same as a dice problem. Certainly, unless it is a cheating dice, you cannot predict which number will come up. However, we do know that there is a one-in-six probability that a number from one to six will appear. The totality of what can happen is known. But, this is not the case in the real world. We are constantly recruiting our knowledge and experience to predict and act. However, in the real world and real environments, unexpected things often happen.

Where should we look for clues to the basic principles and meta-rules that make these abilities possible? Here, I would like to seek a clue to the so-called “I - Thou” problem, which has been deeply felt and pondered by humankind since ancient times.

BOX *"Something Is Wrong" Ability of Creatures*

In recent years, artificial intelligence has been attracting attention for the third time. Unlike the past two times, this latest boom is likely to have a major impact on the society. There is no shortage of topics such as winning against masters in Go and Shogi, which are said to be extremely difficult. This was made possible partly due to advances in software such as deep learning, but of course also due to significant advances in hardware, such as processing speed and memory. Their computational power far surpasses that of humans in some aspects. However, in terms of the quality of information processing, I strongly feel that there is a huge gap between artificial intelligence and living things. In this book, I have written a lot to point out exactly that, but I think the difference can be summarized in the way creatures deal with indefinite environments.

In a recent game between an AI and a Go master, the master somehow managed to get revenge. I do not know anything about Go, but I heard that the master played a surprise attack, i.e., a move that is not standard. If the opponent had been a human Go master, he would have noticed this "anomaly" immediately, but the AI did not "notice" it until much later.

I enjoy mountain stream fishing. Trouts are very sensitive. In the ever-changing natural environment, they are sensitive to the slightest sign of anglers. When I go up a beautiful mountain stream, I am keenly aware that not only trouts, but also insects and birds have an extremely high ability to sense "something is wrong." Their information-processing abilities, in some aspects, are probably not far behind today’s artificial intelligence. However, the "something is wrong" ability of creatures seems to be completely different from that of the current AI.

This may indicate how organisms cope with indefinite environments. They may not think philosophically like humans do, but it seems to imply that they always process information on the assumption that something unknown can happen.

The "I-Thou" Problem

The "I-Thou" problem may be unfamiliar to you. This is a question of how to face the world and the environment. The opening line of Martin Buber's book "I and Thou"[2] is impressive:

To man the world is twofold, in accordance with his twofold attitude.
The attitude of man is twofold, in accordance with the twofold nature of the primary words which he speaks.

He says, in other words, that there are two ways in which the self or subject face the world.

One is the "I - it" approach, where "it" is something that is defined for “I.” In science, things and concepts need to be defined,, so science takes "I-it" approaches to the world.

On the other hand, "thou" is an entity that is essentially undefined, or cannot be described by a finite set of attributes. In the sense that it is essentially undefined, it can also be said to be an entity that can never be known as "it." Also, the fact that it cannot be described by finite attributes means that it is irreplaceable.

The fact that entities that can be described by a finite number of attributes can be easily replaced was something I learned from my own job hunting experience in recent years. Naturally, there are several requirements that must be met for the person being recruited, such as having a PhD or being able to teach linear algebra. Anyone who meets those requirements well is a good candidate.

Yet, the way we face the world is two-fold: we interact with it in an “I-it” way as well as in an “I-thou” way.

My father, born in 1940, is left-handed, square-jawed, and 178 cm tall, has a straight nose and wide eyes, all of which are similar to Sadaharu Oh, the record holder for the most home runs in professional baseball. However, my father is still irreplaceable to me.

Such an entity, "thou," exists only when "I" calls out and asks questions. This point reminds me of the words of my high school teacher, Brother André LaBelle, who once said, "You probably don't think there is a God. God is not some old man with a white beard living in a cave on Jupiter or somewhere. If you call on Him, He exists; if you don't call on Him, He doesn't exist.” Hearing this, I thought to myself, "I see. God exists in a different way than other beings." Buber's words penetrated me so deeply, thanks to Bro. LaBelle, I suppose.

When we think about it this way, "Thou" seems similar to the unlimited environment, but also different. Starting from these subtle differences, we will discuss "Thou" from the perspective of complex systems theory of biological systems.

"Give" to "Thee"

Biological systems live in an environment where events can occur that are even probabilistically unpredictable. This is what we called in this book an indefinite environment. On the other hand, we have also mentioned that "thou," which has long been discussed in philosophy and religion, is essentially something that cannot be defined. The following is just an impression I get from these words, but a certain positivity is felt in "Thou". "Thou" is associated with irreplaceability. Because "thou" is irreplaceable, there are many episodes in which "I" actively sacrifices itself. This is something that is never spoken of in connection with an indefinite environment. From the perspective of an indefinite environment, I, who has been exposed to this term for many years, feels a slightly negative nuance: "Sure, it's unpredictable what will happen, but if possible, I would like to keep the environment predictable. It would take time and cost to fundamentally change the way we have been doing things because of unpredictable events." You cannot expect anything in return for working on something that cannot be defined or inherently unpredictable. In order to expect a return, the target of the action must be defined and predictable, i.e., "If I do this, he or she will do that." This is not possible since the target is essentially indefinable. Therefore, "I" must be prepared to be one-sided when working on something that cannot be defined. This is why "giving" is often discussed in the "I-Thou" issue. So where does positivity come from?

"Love" - The Responsibility Rhat Arises from a Relationship

What we give to "thee" can be rephrased as "love." The Japanese word "love" is a difficult word to use. One reason is that, socially, the moment you start talking about it, it sounds like the doctrine of a fake new religion. Another reason is that this Japanese word itself is a mixture of agape and eros. Love as a natural desire, including its sexual aspects, is called eros in philosophical terms. The issue I want to address here is not eros, but the word called agape.

It seems to me that love = "giving to thee" is a bit unsatisfactory in terms of the source of positivity and irreplaceability. In this regard, I really like the following part of Father Akio Awamoto's book[3)]:

Love is "the fulfillment of one's obligations arising from a relationship." Of course this is not meant to be the best definition of what love is. It's a bland expression that has almost nothing to do with love.

When love is viewed in this way as your obligation and responsibility arising from a relationship, the act of giving becomes proactive, rather than something you do because you have no choice. "Arising from a relationship" is also important. If you think, "This has nothing to do with me," you won't take action. On the other hand, if you think, "In light of the relationship, I deserve to do this (even if it is hard)," you will act proactively and you will not expect anything in return.

The Meta-Rule That Robots Should Have

Biological systems face many ill-posed problems, that is, problems where recognition and behavior cannot be uniquely determined by clues obtained from the environment alone, and constraints are required to solve them. Constraints are hierarchical, with lower-level ones created from higher-level ones. What is the highest level of constraint? In the last chapter, I started by considering the ultimate robot, and pointed out that pointing out that even the well-designed Asimov's Three Laws of Robotics can be deadlocked because they are fixed rules. The meta-rule that should come above these three principles is "love" given to "thee," that is, to work without asking for anything in return as a responsibility that comes from a relationship with a being that cannot be defined. As a reason for this, let us consider what this meta-rule brings about as follows. It may seem a little odd to argue about the reward for not asking for anything in return, but it is not a bad idea to have such a discussion, since living beings have laid the groundwork for it over a long period of time and humanity has sought its importance on the basis of much blood and tears. First, it is a necessary condition for breaking away from fixed and prescribed relationships. Giving and not expecting anything in return can be viewed more broadly than giving to others or not seeking the fulfillment of natural desires. In other words, it can be expanded to include: performing some action or effort even if you do not expect to get a certain result; knowing that all actions may not return the expected result; "knowing" that there are some things that are essentially unknowable; continuing to ask "why?" about the real world, which cannot be defined. If you think about it this way, "giving" brings about the necessary condition for breaking away from fixed and prescribed relationships with the world when the situation changes, and provides the opportunity to solve problems.

Second, it will bring the ability to make autonomous decisions in an indefinite environment. Since all living creatures live in a harsh environment where they must do their best to survive, those who strictly pursue the fulfillment of their desires, the reward for their actions, and their own interests would be able to survive more robustly. However, the opposite seems to be true in many cases, especially in human society. I like novels and dramas based on Japanese history, especially the long period of civil war in the 16th century. When I read or watch them, they remind me of how harsh the period was. Various benefits are offered create traitors among the enemy. In many cases, however, those who benefited were deprived of their freedom of action and judgement, and ultimately manipulated and destroyed. On the contrary, those who sacrifice themselves for righteousness rather than for interests are trusted, and even if they unfortunately die, they are able to make autonomous and independent decisions until the end. In the case of robots, it corresponds to the ability of a robot with a specific mission to avoid being lured by others. By "giving," in turn, autonomy of judgment and thus essential freedom of mind is ensured.

At the end of the previous chapter, I outlined what I believe to be the mechanism of creativity, which I believe is strongly supported by the "love" that “I” give to "thee," as described here.

Neurocoaching

So far, I have considered how "love" (as the act of working on the indefinable " thee" without expecting anything in return) can be understood from the perspective of biological systems theory. However, some readers may be skeptical and wonder, "Is such a thing even an issue in my lifetime?" Certainly, I have not seen or heard of such discussions at any of the neuroscience or other societies in which I participate, let alone at philosophical societies. However, I believe that these discussions, and the techniques that may be gained by deepening them, are potentially needed in society more than we can imagine. I feel that the time is fast approaching when this will be actively studied as a new academic field, which I have arbitrarily named neurocoaching.

So what would neuro-coaching be like? For the purposes of this article, coaching is understood as a technique that facilitates the growth and development of the person being coached (hereafter referred to as the student) through communication. A more theoretical version of coaching, one that theorizes it to the point where machines can coach, is what I imagine neurocoaching. Remember Doraemon? He watches over Nobita and encourages his development, being swept away by his unexpected demands and actions. Doraemon's artificial brain should definitely have a built-in neural circuit to guide Nobita.

Specifically, what kind of mechanism should such a neural circuit have? First, it would have to understand what the student currently needs to learn. Then, we need to build a model of the student, that is, a model of how the student understands what he or she is trying to learn, through their behavior. We also need a mechanism to correct and update the model when the model does not correctly predict the student's behavior. What seems difficult is a mechanism to generate alternative teaching methods or sub-goals when the student is not learning well. Even more difficult is a mechanism to guide the student in the direction he or she should go after he or she has achieved the goal to learn or when he or she shows different abilities or talents in the learning process. Such a mechanism, if realized, is expected to provide a theoretical basis for the pros and cons of various educational methods, as a complement to statistical data from psychology and social sciences.

In this light, it is easy to understand how neurocoaching relates to the issues of 'thou' and 'love' discussed in the previous section. In the coach's mind, the understanding of the student is defined by a model. The relationship between the coach and the student here can be said to be what Martin Buber calls an "I-it" relationship. However, due to the student's growth and unexpected behavior, the model often needs to be discarded and reconstructed. But a good coach must never abandon the student and go beyond the model to stay connected to the student as an indefinable entity. This relationship can be called an "I-thou" relationship. To stay connected to the student and to try to reach a higher level is love as a responsibility that comes from the relationship. In this direction, scientific and theoretical research on 'thou' and 'love' is possible.

References

1) Asimov I. *I. Robot*. (1950)

2) Buber M. *I and thou*. (1937)

3) Kurimoto A. *Kekkon Suru Futari e*. Joshi Paulo Kai, Tokyo (1993) in Japanese

Appendix: Introduction to Complex Biological Systems Theory

Appendix A: Self-Generating Order in Complex Systems

What is the state of being alive? The starting point for research into answering this big question based on physics, rather than reductionism or spiritualism, is considered to be the book "What is Life?" by the famous quantum physicist Erwin Schrödinger. He pointed out that biological systems seemingly violate the second law of thermodynamics, i.e. the law of increasing entropy.

The word entropy refers to the degree of randomness. Most systems develop towards randomness over time. This tendency is called the law of increasing entropy and is the equivalent of the proverb "there is no use crying over spilt milk." Your room will not clean itself.

On the other hand, living organisms live by eating something and excreting something. The existence of such a flow is thought to reduce disorder and create orderly states and patterns such as body structure and movement. However, the autonomous generation of temporal and spatial patterns is not found only in living systems. Not a few are found in nature. Scale clouds are an example of this.

Self-organization theory, or complex systems theory, deals with how these orderly phenomena and states emerge. I would like to make this theory one of the major foundations behind creativity. Below, I will provide an overview of the theory in a little more detail than the main text, although it is still very basic.

I. Deviation from Equilibrium and Pattern Generation

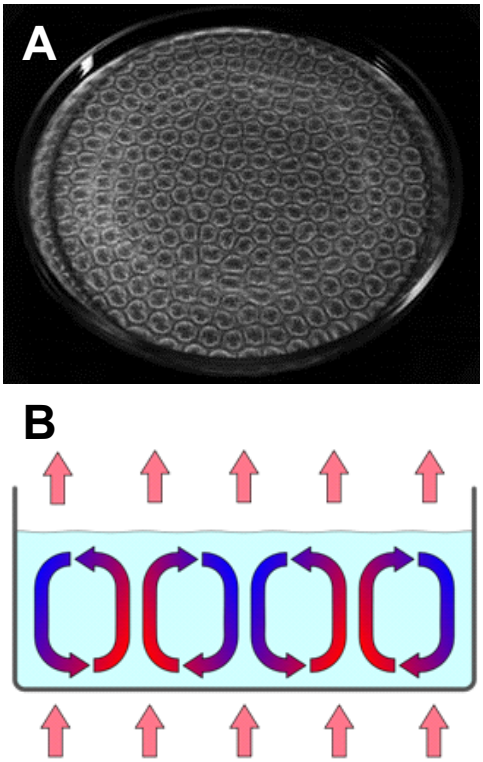


Figure 1. Benard convection. a. Example made with silicon oil (from Reference 2). b. Schematic of convection structure. Arrows indicate direction of heat.

Self-Organizing Phenomena Generating Spatio-Temporal Patterns

Seeing sardine clouds high in the autumn sky, some of you may feel refreshed and at the same time wonder how such patterns are formed. I don't think many people sip miso soup without any ingredients, but some may have found some patterns in such a slightly desolate miso soup. Such a spatial structure is called Benard convection in complex systems science, after its discoverer, French physicist Henri Benard. If you make it well experimentally, you will get a splendid patterning like the one shown in Figure 1A. Such patterns do not always occur. They occur when heat escapes rapidly from one direction and a large temperature flow or gradient above a certain threshold is generated in the air or liquid. In the case of sardine clouds, heat is released into space; in the case of miso soup and Benard convection, heat is released into the air. If the temperature gradient is below a threshold, the heat propagates evenly in the liquid or gas, but if it exceeds a certain threshold, a rolled convective structure like that shown in Figure 1B is generated.

Although it is not a natural phenomenon, Figure 2 shows a chemical reaction called the Belousov-Zhabotinsky reaction (hereafter BZ reaction). It is so called because it was discovered by Russian scientists Boris Belousov and Anatole Zhabotinsky.

When the reagent is prepared and stirred with a stirrer, the color of the entire solution changes periodically and oscillatingly. In other words, a temporal pattern is generated. On the other hand, without stirring, local reactants in the solution diffuse slowly to the surroundings. This causes spatial structures in the progress of the reaction, specifically colored ripple patterns in the form of concentric circles, spirals, or vortexes, to appear, and these spread over time. This type of system is called a reaction-diffusion system. The starting points of the patterns are considered to be dust, scratches on the bottom of the petri dish, impurities in the solution, etc., but the patterns varies from trial to trial. When ripple patterns from different starting points collide, unlike waves on the surface of water, they do not overlap but rather disappear together like a wildfire, but the wave with the shorter period gradually pushes forward and eventually swallows the wave with the slower period. I have experienced this myself, and it was very moving. If you have the opportunity, I hope you will try it out.

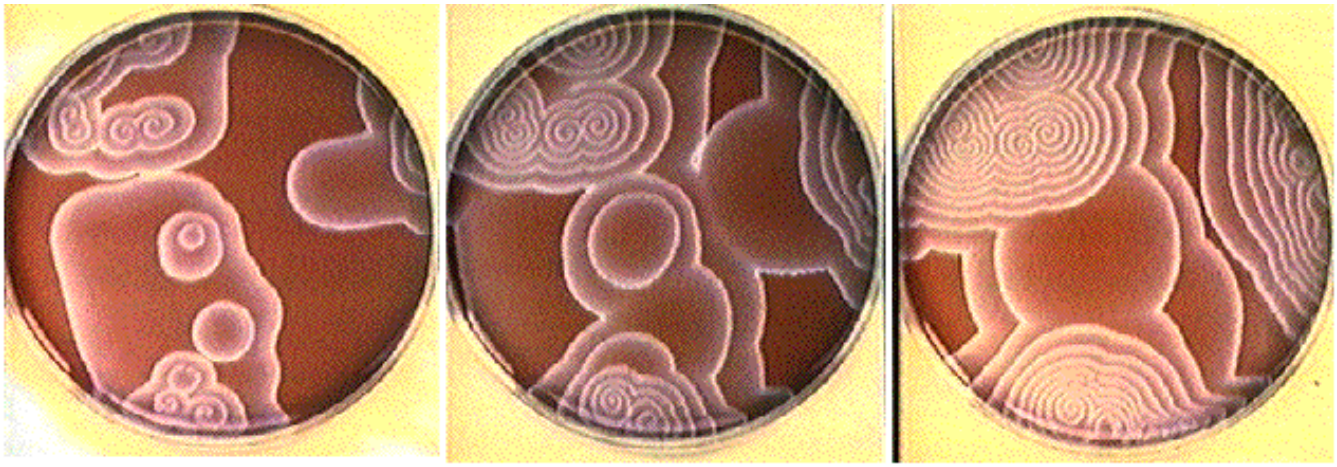


Figure 2. BZ reaction. Reaction proceeds from left to right (from Ref. 3).

Prigogine's Dissipative Structures[4)]

Figure 3. Ilya Prigogine

The scaly cloud-like structure called Benard convection, which we saw in the previous section, appears when the thermal gradient is greater than a certain threshold, and disappears when the gradient disappears and the heat is balanced (this is called the equilibrium state). Patterns in miso soup disappear when the temperature of the miso soup balances with room temperature. In a chemical reaction called the BZ reaction, spiral or circular wave spreading and color oscillations with time are observed until all reactions are completed and equilibrium is reached. The temporal and spatial structures that occur away from these equilibrium are called dissipative structures. Dissipative structure may be an unfamiliar term to readers, but its opposite is equilibrium structure, or even crystalline structure. Dissipative structure is dynamic, occurring when there is a flow of heat, energy, chemical reactions, etc., whereas in crystalline structure, the interaction with the outside world is calm, i.e., equilibrium has been reached. For proposing this concept of dissipative structures, Ilya Prigogine (Fig. 3) was awarded the Nobel Prize in Chemistry in 1977.

Mathematically, the change of things over time is generally expressed by an equation called a differential equation. For example, the rate of change of one quantity X is expressed as an equation that depends on the magnitude of X at the moment or the magnitude of another quantity Y. By considering the differential equation of the quantity X, we can determine whether the equilibrium state of X is stable or not. For simplicity, let's assume the case where the equilibrium value of X is zero and its rate of change is described as $-X$. When X is positive, the rate of change is negative, i.e., X decreases and moves toward the equilibrium point of zero, while when X is negative, the rate of change is positive, i.e., X increases and returns to the equilibrium point of zero. The reason why X returns to its original value even if it deviates from the equilibrium point zero is that the rate of change is proportional to -1, or negative value, of X. If this value is positive, X cannot return to its original value even if it deviates slightly from zero.

Next, let's consider the case where the rate of change of the quantity X follows, for example, $X - X^3$. Suppose again that X is in equilibrium with a value of zero. Now, if X deviates slightly from zero, say 0.1, the rate of change is -0.101. This rate is not much different from -0.1, the value obtained by simply approximating $-X$, ignoring the X^3 term in the equation of change $-X - X^3$. Again, consider the case where the deviation from equilibrium is 10. In this case, the rate of change is -1010, which is far from the value -10 obtained by approximating the expression of change as $-X$.

In general, a system is linear if the rate of change can be expressed or approximated by a linear sum of quantities X and Y (each multiplied by a constant and added, e.g., $-X + 2Y$); all others are called nonlinear systems. The larger the deviation from equilibrium, the less linear approximation is possible, as we saw earlier. A dissipative structure is a temporal and spatial structure that can appear in situations where large thermal and energy gradients, rapid chemical reactions, etc., that is, when the system deviates greatly from equilibrium, and the nonlinearity of the rules (i.e., differential equations) that govern the change cannot be ignored (such a system is called a non-equilibrium nonlinear system).

How Can It be Structured? - Bifurcation Theory

As described above, the phenomena of spontaneous formation of spatiotemporal patterns, such as the Bénard convection, BZ reaction or pulsations, are called self-organization phenomena, and such patterns are called dissipative structures. Dissipative structures are dynamic structures that appear when chemical reactions and exchanges of heat and energy with the outside of the structure are far from an equilibrium state, and the equations describing the changes are nonlinear, that is, when the effects of terms that are not the sum of the quantities involved (e.g., the product of quantities X and Y or the square of X) cannot be

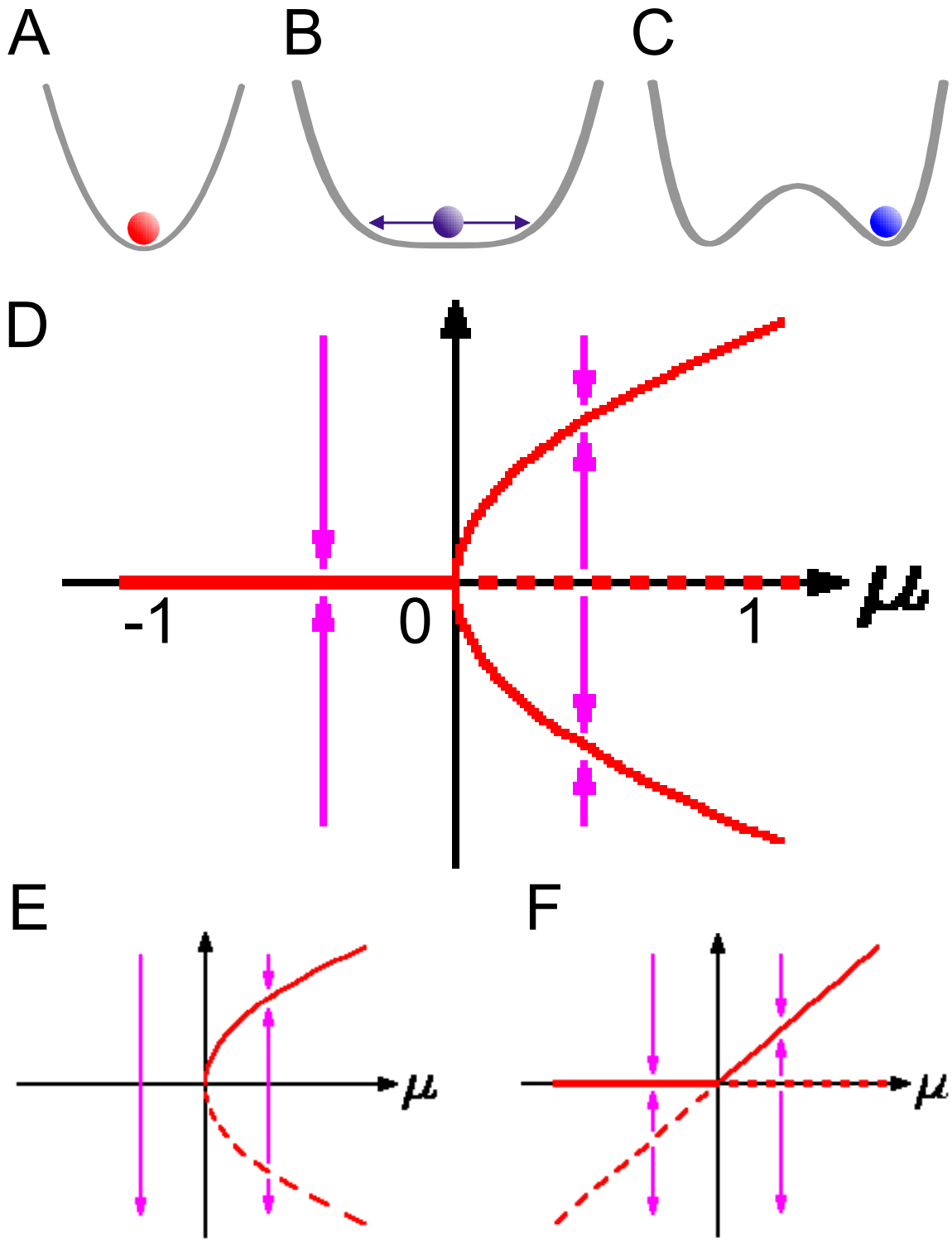


Figure 4. Schematic diagrams of bifurcation. The number of singularities and their stability/instability change rapidly as the bifurcation parameters change. A-D, An example of pitchfork bifurcation. A, B, C. Potential shapes when the bifurcation parameters are -1, 0, 1, respectively. Balls indicate the current state of the system. D. Change in singularities (peaks and valleys of the potential) as the bifurcation parameter changes. E, F. Changes in singularities for saddle node bifurcation and alternating bifurcation, respectively. The basic form of the equation for the change of variable X in the saddle node bifurcation is expressed as $\mu - X^2$, and for the alternating bifurcation as $\mu X - X^2$. Stable singularities are indicated by solid lines, and unstable singularities are indicated by dashed lines. Arrows indicate the direction of change when deviating from the singularity.

ignored. A theory called bifurcation theory in nonlinear mechanics provides the basis for how structures and patterns form in these non-equilibrium nonlinear systems.

As an example, consider the case where the rate of change of a quantity X follows $\mu X - X^3$, which includes the example in the previous section. μ is a constant, but is a parameter that we can freely set.

Now, to help readers get a clearer picture of bifurcation, let us consider "potential." Potential cannot be defined for all differential equations. But fortunately, for $\mu X - X^3$, it can be defined as the inverse of the integral with respect to X, $-\mu X^2/2 + X^4/4$. By imagining this potential, we can intuitively understand which values of X do not change and whether they are stable.

First, let's consider the case where μ is -1. The equation for the rate of change is $-X - X^3$, and its potential can be written as $X^2/2 + X^4/4$. The shape of the potential in this case is shown as a curve in Figure 4A. The ball on the curve represents the state of the system. Metaphorically, suppose this ball is driven by "gravity". You can immediately see that the rate of change is zero. That is, where $-X - X^3 = 0$ (such a point is called a singular point, stationary point, or singular point), is a simple cubic equation, so you can easily see that $X = 0$. This corresponds to the central valley in the potential curve in Figure 4A. How the ball is attracted to the valley corresponds to the slope of the curve where the ball is located. The steeper the slope, the stronger the attraction to the valley. The gradient of the potential at point X_0 is $-X_0 - X_0^3$. Looking at the shape of this potential, we can intuit that the singularity $X = 0$ is stable. That is, if the ball is shifted to the positive side, it moves back to the negative side, and if it swings to the negative side, it moves to the positive side. Such a point is called a stable singularity, or an attractor, in the sense that if it is displaced, it will be attracted back.

So, what if μ is 1? In this case, the singularity increases to three ($X = 0$ and ± 1) because the rate of change is zero, i.e. the solution of $X - X^3 = 0$. The shape of the potential in this case is as shown in Figure 4C. That is, there is a peak at $X = 0$ and valleys at $X = \pm 1$. You can easily see that the valleys are stable. On the other hand, the peak at $X = 0$ is unstable. Sure, if X is exactly zero, X does not change, but if X deviates from zero even slightly, it will roll off the peak.

Figure 4D shows the number of singularities and their stability or instability when this parameter μ is changed from negative to positive. When it is negative, there is only one singular point, zero, and it is stable. However, when μ is positive, the number of singularities suddenly increases to three (zero and $\pm\sqrt{\mu}$), and zero is unstable.

Let us also review the change in the shape of the potential when μ is changed from negative to positive. When μ is negative and has a large absolute value, the potential takes the shape of a deep vase. In this case, even if it deviates from the stable singularity of zero, it returns to the singularity immediately.

As μ approaches positive, the potential gradually changes to a shallower form, and when μ reaches zero, it becomes completely flat in the very vicinity of the singularity zero (Figure 4B). In such a case, it does not easily return to zero even if it deviates from zero.

When μ becomes positive, the singular point of zero becomes unstable, and two stable singular points $\pm\sqrt{\mu}$ appear. As μ increases, the two stable points move away from zero, the valleys of these points become deeper, and the peak at zero becomes steeper.

A sudden change in the number of singular points or their stability, depending on a parameter (in this case, μ) of the equation of change (differential equation), is called a bifurcation, and a phenomenon explained by bifurcation is called a bifurcation phenomenon. There are a relatively limited number of basic types of bifurcation patterns. The bifurcation shown here is called a pitchfork bifurcation, since Figure 4D shows a fork shape. Parameters such as μ are called bifurcation parameters.

Other typical ones with a single bifurcation parameter include saddle node bifurcations (bifurcations that result in one stable and one unstable singularity from a singularity-free state. Figure 4E), alternating bifurcations or trans-critical bifurcations (bifurcations where the number of singularities is two, but their stability/instability alternates. Figure 4F), and hop bifurcations, which generates oscillations (described in detail later).

Dissipative structures are the patterns created by bifurcations in non-equilibrium non-linear systems.

'Fluctuations' Drive the System and Predict Bifurcations

Again, consider a system described by a variable X whose change follows $X - X^3$. This system has three singular points, namely zero and ±1, where $X - X^3 = 0$. The point of zero is unstable in the sense that the value of X moves away from zero if it deviates slightly from zero, whereas ±1 are stable, i.e., even small deviations from these points disappear.

What if X is completely zero? Although zero is unstable, as long as it is completely zero, the quantity X does not change, so X stays at zero. But is this possible in our everyday world? Even in a closed room, gas molecules move around unless the temperature is absolute zero. Even if you are not playing a CD, if you turn up the volume, you can hear the noise. The real world in which we live is filled with noise or disturbance, which we call fluctuations.

For the above reason, the state where X is completely zero cannot exist. That is, a small noise or fluctuation will cause X to fall into a trough on either side. In this sense, the system is driven by fluctuations. Once it falls into the valley, it never climbs the mountain to the other side. A small fluctuation determines the fate of the system.

Sudden and significant changes can sometimes be a serious problem. Earthquakes, major depressions, epidemics, cancer, epilepsy, etc. are all caused by the complex interaction of different factors. Therefore, it is useful to consider them as bifurcation phenomena in nonlinear dynamical systems, because it allows to detect their early warning signals.

A stable point can be either highly stable or not so highly stable. That is, there are points where deviations and turbulences are quickly canceled out, and points where it is susceptible to disturbances and noise and takes time to cancel out their effects. It is particularly noteworthy that before the bifurcation, that is, before the stable point becomes unstable, it passes through a stable state, which is stable but not very strong. In such a pre-bifurcation state, the system is sensitive to perturbations, so that some observed quantities show large fluctuations or variances (critical fluctuation) or very long recovery time to the singular point (critical relaxation). Such phenomena have been known for a long time, but have recently attracted renewed attention. It would be wonderful if theories related to these phenomena could be developed (e.g., reference 5)) to predict major catastrophes in advance.

The following is an aside. When I look at serious social problems, I sometimes think, "How helpless I am." On the other hand, looking back at history, there are countless examples of how the decisions and actions of a few heroic (or not so heroic) individuals can have a decisive impact on subsequent societies. Where does this gap come from? Bifurcation theory teaches us the following: Many heroes appear during periods of social instability. An unstable state transitions to another state with a slight fluctuation. Just like an upside-down tumbler rolls with a little force. Heroes, through various circumstances, may find themselves in a position where they can push the upside-down tumbler that is society. In this world, your turn may come around. We must train ourselves to push the tumbler in the right direction, and we also need to cultivate an eye for those who can do so.

Haken's Synergetics

Prigogine's theory of dissipative structures provided a direction for understanding how structures and patterns self-organize. However, the real world is much more complex than the chemical reactions that he specifically studied. He did not address the question of how the many elements of a complex system behave and integrate, and which part of the system holds the cast for the whole system.

Hermann Haken offered a perspective on this question. Through his research on the fundamental theory of lasers, he realized that even in a system composed of many factors, there are a small number of factors that can well describe the state of the entire system. He called such parameters order parameters.

Bifurcation theory deals with differential equations that describe the time evolution of a system. The theory takes a singular point where the value does not change, and considers how a small deviation from that point behaves, i.e., whether it returns to the point or leaves it, or, if it returns, whether it moves quickly or slowly. Harken focused on this point. That is, he divided the parameters describing the system into those that return quickly to the singularity and those that do not. Small perturbations return quickly to the singularity, so quick parameters can be ignored. His idea was that the behavior of the entire system could be reduced to and described by relatively slow parameters. The self-organizing integration principle, the parameter that changes relatively slowly and represents the overall behavior of the system, and Haken's theory of self-organization based on these ideas are called the slaving principle, order parameter, and synergetics, respectively.

The concept of synergetics has paved the way for continuous discussion of macroscopic systems and microscopic levels such as molecules and atoms. At the same time, however, it has strongly warned against the elemental reductionist viewpoint to which many scientists are still trapped. Synergetics suggests that the state of the universe is not necessarily explained by the behavior of microscopic elements. In other words, the order parameters that represent the state of a system can be macroscopic, not necessarily microscopic. For example, locusts that have swarmed due to drought are different in color and shape from normal locusts, but the outbreak is a matter of macroscopic balance in the ecosystem. The root cause of the outbreak is not the secretion of a substance that causes morphological changes from the brains of locusts.

In recent years, the number of depressed patients has increased rapidly, and the examples around me make me realize that prescribing antidepressants does not necessarily solve the problem. Depression is, of course, an innate disposition of the individual patient, but the main problem is the social environment in which the patient lives, and abnormalities in the brain's neurochemicals are secondary to the problem.

Synergistics puts a stop to our excessive reductionist thinking.

II. Nonlinear oscillation

Among the bifurcation phenomena described in the previous section, there is one in which a stable singularity becomes unstable and an oscillatory state becomes stable. This type of bifurcation is called a Hopf bifurcation.

There are various oscillatory phenomena in nature. The washing machine I used when I was a student would rattle and vibrate when the spin cycle was finished, annoying the neighbors. My father used to drive an old used car that would shake a lot at about 100 kilometers per hour, but mysteriously the shaking would disappear if he went faster than that. These are all nonlinear oscillation phenomena. In the first place, linear oscillations do not occur unless the conditions are met, so most of the oscillatory phenomena we see in our daily lives can be considered nonlinear.

An oscillator is a term used to refer to something that oscillates. If the equation describing the oscillation is linear, it is a linear oscillator; if it is nonlinear, it is a nonlinear oscillator. The following explanation is based on the image of a spring pendulum (Figure 5). Let x be the variation of the spring from its natural length, and let positive and negative denote the state in which the spring is extended and contracted, respectively. The spring generates a force proportional to the deviation from its natural length, in the direction that cancels the deviation. If the proportionality constant is c and the mass m of the ball attached to the spring is 1 for simplicity, then according to Newton's law, mass × acceleration = force, so the oscillatory motion of the spring can be described by the differential equation

$$\frac{d^2x}{dt^2} = -cx \,.$$

The right-hand side is a first-order function of x, so this is a linear oscillator. Since the time variation of position is velocity and the time variation of velocity is acceleration, the left side is the second derivative of x.

This oscillator is a linear oscillator without friction. The kinetic energy of the balls and the potential energy of the spring (deviation from zero) are mutually exchanged. In other words, when the potential energy is zero (i.e., the ball passes through the zero point), the kinetic energy is maximum (the ball's speed is the highest), and when the motion stops, the spring has the maximum amplitude and the largest force is applied to the ball.

Actual springs are subject to friction. Friction is generally proportional to the velocity. In this sense, it is also linear. If we set this proportionality constant to c_1 (and at the same time replace the above constant with $c \rightarrow c_2$), the equation is,

$$\frac{d^2x}{dt^2} = -c_1\frac{dx}{dt} - c_2x \,,$$

which is also linear in this sense. In this case, the oscillation gradually decreases and finally stops at zero. The potential energy is irreversibly dissipated by friction into the heat of the floor, *etc*.

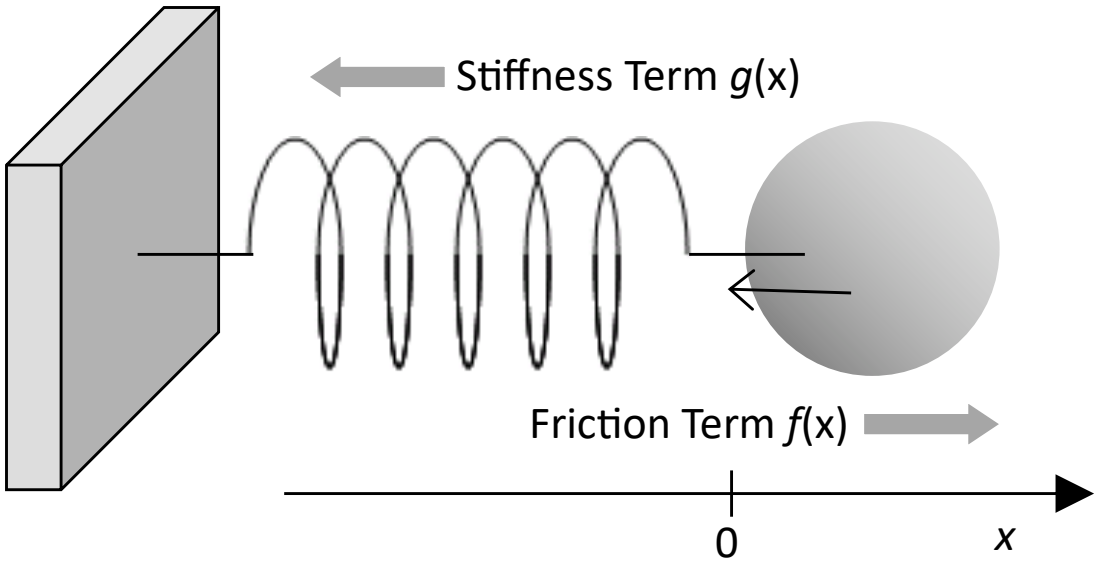


Figure 5. Image of a spring pendulum. Let the natural length of the spring be the reference, i.e., zero, and make an equation for the deviation X from it. In a linear oscillator, the terms g(x) for the spring stiffness and f(x) for the friction are constants.

Frictionless Linear Oscillation and Initial Value Preservation

In this section, we will look somewhat more carefully at the properties of a frictionless linear oscillator, analogous to a spring pendulum that receives no friction from the floor.

The amplitude of a frictionless linear oscillator depends on the initial value, i.e., from which position it started to oscillate, and it persists. If the oscillation starts when the deviation x from the natural length of the spring is 5 (i.e., if the initial value is 5), it will continue to oscillate with an amplitude of 5, and if the initial value is 2, it will continue to oscillate with an amplitude of 2 (Figure 6A). Also, if two springs of the same stiffness start oscillating at the same time and the initial values are different, the amplitude difference remains constant.

To make the following discussion easier to understand, let us consider a phase space. In Figure 6A, time is plotted on the horizontal axis and x on the vertical axis. In this phase space, x is plotted on the horizontal axis and the first-order time derivative dx/dt of x is plotted on the vertical axis. Then, each of the two oscillations in Figure 6A can be represented by circles of different radii, as shown in Figure 6B. The scale of the vertical axis depends on the stiffness of the spring, or spring constant (c_2 in the previous equation, the value that determines the speed, or frequency, of the oscillation), but it can be arbitrary, so here we have represented them as circles. Using phase space, it is convenient to express the position of an oscillation at a given time as the angle of a point on a circle. This angle is called the phase. There are also oscillations of the same amplitude that are temporally shifted (Fig. 6C). Using phase space, such a shift can be expressed in terms of angular differences in orbits, or phase differences (Figure 6D). If the springs have the same stiffness, the phase difference will not disappear. This is the same as the distance between people who start running from different positions on the track of an athletic field if they are running at the same speed.

As we have seen above, in a frictionless linear oscillator, when the starting amplitude, amplitude difference, and phase difference continue to remain the same, we say that the initial values are preserved. To be a bit more technical in preparation for later, the area consisting of a given amplitude difference and phase difference in phase space (we call this the phase volume, not the phase area, in preparation for extensions to phase spaces of three or more dimensions) always retains that area (the meshed part in Figure 6E).

Thus, a linear oscillator will continue to oscillate in the same way if nothing is happening, but it is vulnerable to disturbances. If the amplitude is disturbed, it remains the same; if the phase is modulated, it remains the same. Keep this in mind as well.

In a linear oscillator with friction, the phase volume gets smaller and smaller. This is because with friction, even if the initial values are different, the oscillation will eventually be attracted to the singularity and the oscillation will stop.

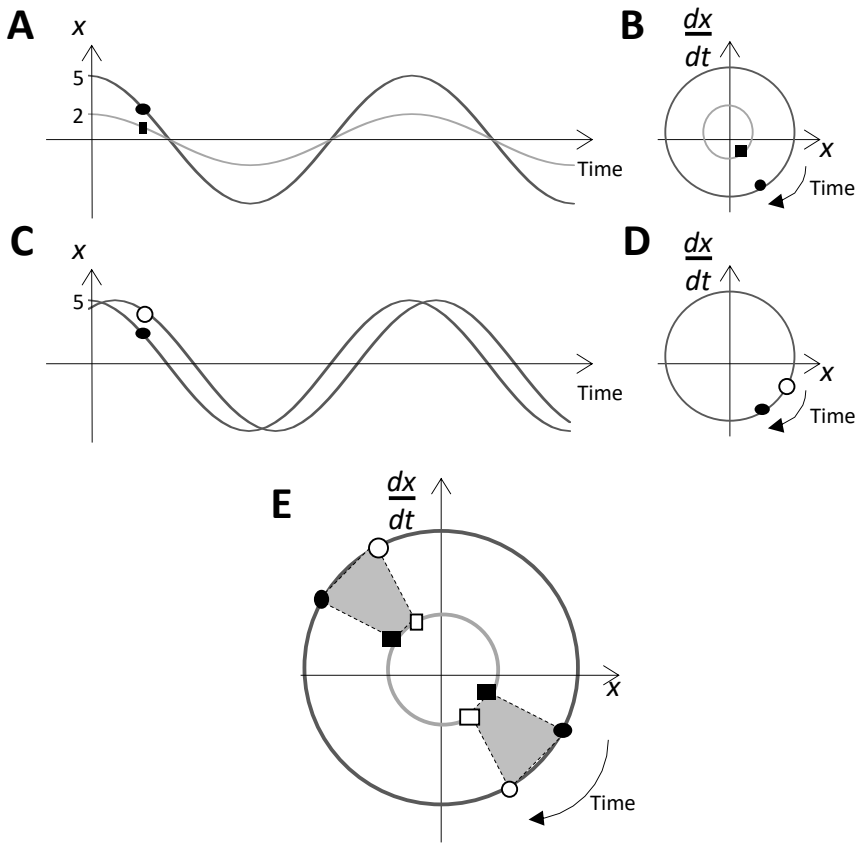


Figure 6. Properties of linear oscillators. For details, see text.

Many Oscillations are Nonlinear

A nonlinear oscillator is an oscillator in which the differential equation describing the oscillation is nonlinear, i.e., contains terms other than a linear expression of X. The most well-known nonlinear oscillator, and one that best possesses these properties, is the van der Pol oscillator (hereafter referred to as the VdP oscillator), which was discovered in 1927 by Dutch electrical engineer Balthasar van der Pol in an electrical circuit. It is expressed by the following equation:

$$\frac{d^2x}{dt^2} = -(a_1x^2 + c_1)\frac{dx}{dt} - c_2x$$

.

The friction coefficient $-c_1$ of a linear oscillator is replaced by a so-called nonlinear equation that includes the square of x i.e. $-(a_1x^2 + c_1)$. For oscillation to occur, a_1 must be positive and c_1 must be negative. This causes the value of $-(a_1x_2 + c_1)$ to be positive or negative, depending on the value of x.

The VdP oscillator does not preserve the initial values. That is, no matter what the initial values, it eventually settles down to a constant form of oscillation (such oscillation is commonly referred to as a limit cycle). This is a major difference from linear oscillators, which oscillate with an amplitude corresponding to the initial values.

Figure 7 shows the oscillation when a_1 is set to 1, c_1 to -1, and c_2 to 1. Figures 7A and B show the waveforms for the initial values of x of 2.5 and 1, respectively. It can be seen at a glance that both cases have settled into the same amplitude and the same waveform oscillation, regardless of the initial value.

This can be better understood by plotting the phase space in Figure 7C, where the initial value 2.5 is outside and 1 is inside.. The initial value of 2.5 is outside and 1 is inside.

Due to this property, the area of a certain range consisting of a collection of points with different initial amplitudes and phases (area since we are considering two dimensions now, but generally this is called phase volume) also shrinks rapidly. In Figure 7C, you can see that the area of the gray area surrounded by ●○□■ on the right side has almost no area after half a cycle (the area surrounded by ●○□■ on the left side). Unlike a linear oscillator with friction, the VdP oscillator continues to oscillate even when the phase volume approaches zero.

Because of this property, VdP oscillators return to their original waveform even if the oscillation is temporarily disturbed by external disturbances. This also differs from a frictionless linear oscillator. This nonlinear oscillator property plays a crucial role in the natural maintenance of homeostasis in living organisms.

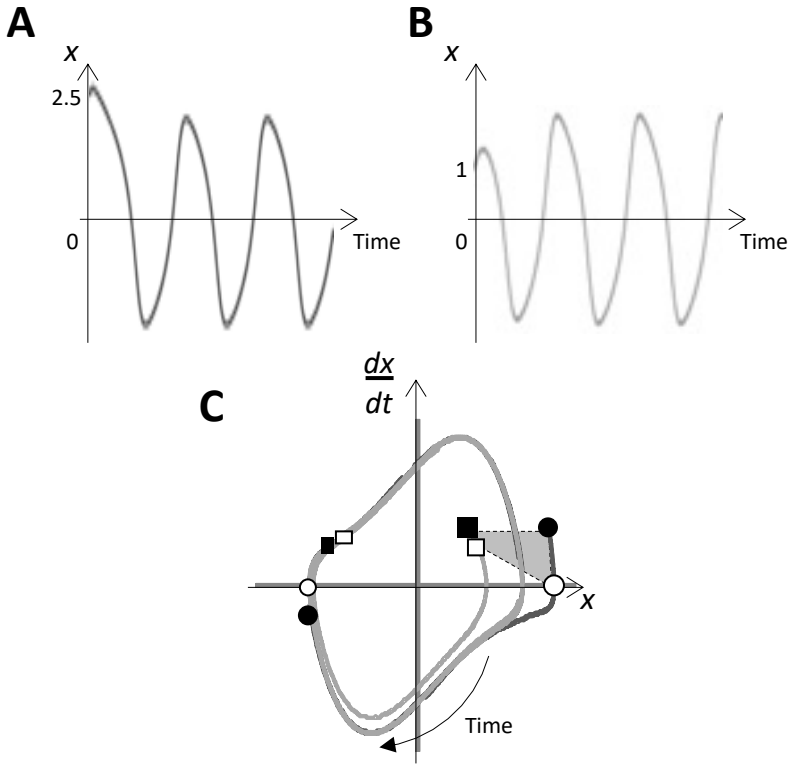


Figure 7. Properties of van der Pol oscillator.

Why Does the Van der Pol Oscillator Oscillate?

Why does the typical nonlinear oscillator, the van der Pol (VdP) oscillator, seen in the previous section, oscillate? The differential equation that represents the VdP oscillator is as follows:

$$\frac{d^2x}{dt^2} = -(a_1x^2 + c_1)\frac{dx}{dt} - c_2x$$

.

The inverted sign of $-(a_1x_2 + c_1)$ multiplied by dx/dt on the right side, i.e. $(a_1x_2 + c_1)$, is the friction (resistance) term. For the VdP oscillator to oscillate, a1 must be positive and c1 must be negative, as described below. There are other ways to explain the mechanism by which the VdP oscillator oscillates, but here, to help readers intuitively understand, we will focus on this friction term and explain it by analogy to the movement of the ball in a spring pendulum, which was used previously.

Now let a_1 be 1 and $c1$ be -1. If we look at the sign of the friction term $(a_1x_2 + c_1)$, it is negative in the range from -1 to 1, and positive otherwise (Figure 8).

Consider the case of an initial value of 2. The singularity of this oscillator, i.e., the point where it does not change in time, is the point at which x is zero. Also, since there is a positive spring constant c2, the ball starts its motion toward zero; until x reaches 1, the value of friction is positive, that is, the ball is in the positive resistance region (on the positive side), so the force is applied in the opposite direction of travel. However, the force from the spring continues to act on the ball, so it does not stop.

When x becomes less than 1, it enters a region of "negative friction". It's hard to imagine, because in the real world you never encounter a floor or floor with negative friction, but in this region the ball receives a force in the same direction as its direction of motion, so it continues to accelerate. As a result, it also passes the zero singularity, passes the point of -1, and enters a region of positive friction on the negative side, which is less than -1.

This may sound confusing, but if the ball enters a region of positive friction on the negative side, it will stop somewhere. Even though the ball stops for a brief moment, it is still receiving a positive force from the spring, so it will begin to move back in the positive direction accordingly. Then it crosses the -1 line, enters the negative friction region again, passes the singularity, crosses x = 1 and enters the positive friction region again on the positive side, and this continues forever.

This is how the VdP oscillator continues to oscillate.

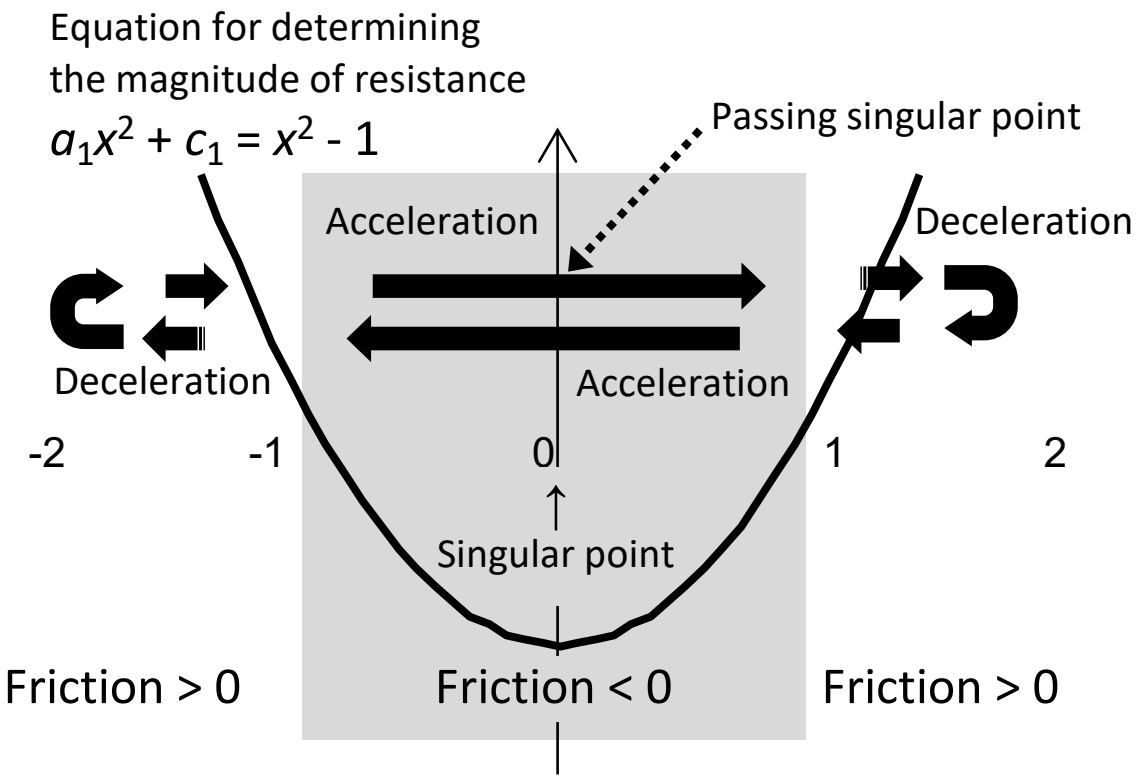


Figure 8. The van der Pol oscillator has a region of “negative” friction near the singular point.

III. Mutual Entrainment between Nonlinear Oscillators

As I explained in the previous section, nonlinear oscillators synchronize under appropriate interaction conditions. This phenomenon is called mutual entrainment between nonlinear oscillators.

The first person to discover this phenomenon was a 17th century Dutch scientist named Huygens. When two pendulum clocks were hung on a single board, they mysteriously always synchronized. It cannot be explained considering the accuracy of the clocks at that time. Now it is known that it happens depending on the slight interaction through the board, but it was completely incomprehensible to the people at that time.

Mutual entrainment also occurs between a large number of oscillators. A well-known example is the collective light emission of fireflies. There are huge groups in the tropics, and their synchronous luminescence is spectacular. The beating of the heart is also a collective synchronization between cardiomyocytes. When cells are separated, they beat individually. However, when they come together to form “tissue,” the individual beats synchronize. The entrainment between nonlinear oscillators is a very complex phenomenon. Here we focus on the phase oscillator (Kuramoto oscillator), which was devised to simplify and treat the phenomenon. The phase oscillator focuses only on the phase of the oscillation. This allows us to have a theoretical perspective on the entrainment phenomenon of nonlinear oscillators.

Kuramoto Oscillator -- Toward Understanding the Phenomenon of Mutual Entrainment

Nonlinear oscillators have the property of returning to their original amplitude in the presence of a disturbance. We can then simplify by ignoring the amplitude and focusing only on the phase, which is the angle of the rotating oscillator at any given time. This simplification is called phase reduction. To simplify even further, we can also assume that the phase velocity is constant. These ideas led to the creation of a phase oscillator (Kuramoto oscillator) by Professor Yuki Kuramoto.

The interaction between two oscillators is assumed to depend on the phase difference between the two oscillators. In other words, if one oscillator's phase is ahead of the other's, the phase velocity will decrease, and if it is the other way around, the phase velocity will increase. Using the sine function as the simplest expression of this relationship, the equation for the time development of the phase of the oscillator *i* interacting with another oscillator *j*, i.e. the differential equation of the phase, can be written as follows:

$$\frac{d\theta_i}{dt} = \omega_i + K\sin(\theta_j - \theta_i) \quad .$$

θ is the phase. ω is the phase velocity. K is the strength of the interaction between the oscillators (coupling constant). The differential equation for the time development of the phase difference between oscillators *i* and *j* is

$$\frac{d(\theta_i - \theta_j)}{dt} = \frac{d\Delta\theta}{dt} = \omega_i + K\sin(\theta_j - \theta_i) - \omega_j - K\sin(\theta_i - \theta_j) = \Delta\omega - 2K\sin\ \Delta\theta \quad .$$

The phase difference between the oscillators is $\Delta\theta$, the difference in phase velocity $\Delta\omega$, and the strength of the interaction (coupling strength) is K.

This equation shows the trade-off between the difference in the intrinsic frequency of the two oscillators and the coupling strength. The fact that the two oscillators are entrained, or phase-locked, means that the phase difference $\Delta\theta$ between them does not change with time, i.e., the left side of the above equation is zero. Given that the sine function here only takes a range of ± 1, the condition for the existence of a state in which the phase difference $\Delta\theta$ does not change with time is,

$$-1 \le \sin\ \Delta\theta = \frac{\Delta\omega}{2K} \le 1 \quad .$$

In other words, if the absolute value of the coupling s trength K is small, the oscillators will not be entrained even if the eigen-frequencies are not very different, and if it is large, they will be entrained even if the eigenfrequencies are very different.

The entrainment phenomenon between Kuramoto oscillators can be easily extended to the case of n oscillators. Figure 9 shows how, as the coupling strength is gradually increased, the oscillators, which were oscillating independently, become phase-transitionally synchronized after a certain critical value.

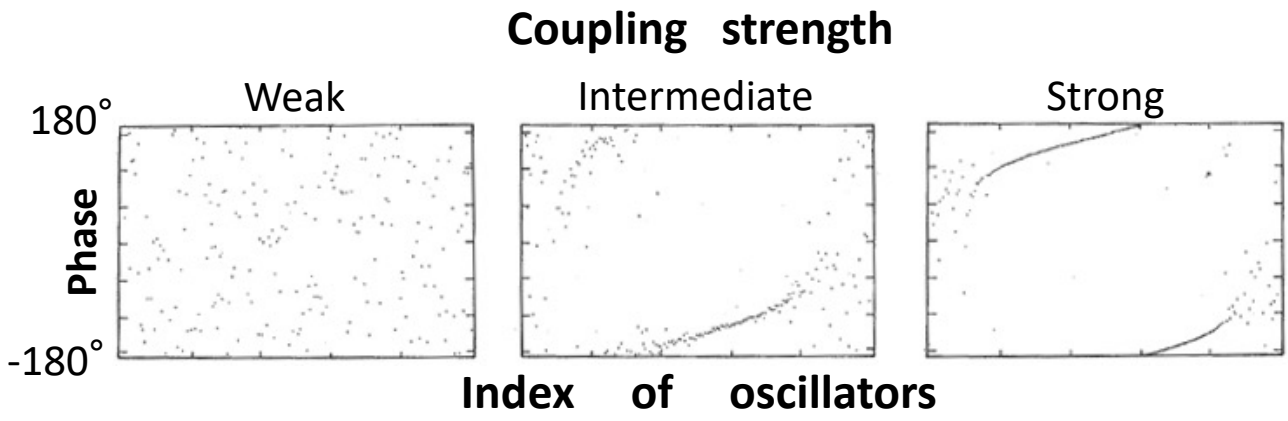


Figure 9. 200 oscillator entrainment simulation. The dots are the phase of each oscillator at a certain instant. They begin to synchronize as the coupling strength increases (from Ref. 7).

Why is Mutual Entrainment of Oscillators so Noteworthy?

As mentioned throughout this book, mutual entrainment of nonlinear oscillators is currently attracting attention as a mechanism for information processing and control by the brain. As for why it is worth paying attention to, researchers who have noticed it simply say that it is a phenomenon, while conversely, researchers who do not understand why it should be of interest often seem to resent the fact that the reasons are not explained. The author believes that there are the following general reasons.

In my opinion, entrainment is expected to be suitable for obtaining a globally consistent relationship between elements or factors. The reason for this is that oscillations or phases are a circular and neutral quantity. To visualize a consistent relationship between different factors or elements, imagine each of them as a ball rolling on a surface. And, when the balls come together, such a state is considered to be one in which a consistent relationship has been achieved between them.

Consider a case where balls roll in a space with infinity and infinity, such as the Cartesian coordinate plane system. If there is a "groove", the balls can come together, even though there is no interaction between them. However, there are usually many "grooves." Several balls are trapped in each "groove", and it is difficult for all the balls to gather in one place (Fig. 10A).

What if there is no “groove”? The balls would roll freely, but if they were rolling at different speeds the distance between them would get wider and wider, and they would move away from each other to infinity. Even if there was gravitational force between the balls, once they failed to attract each other there would be no chance (Figure 10B). Even if there is a gravitational force or something between the balls, there is no chance if they fail to attract each other once (Fig. 10B).

On the other hand, what about balls rolling on a sphere without "grooves"? The rolling speeds are so different that they may not come together (they may not synchronize). But since they are rolling on a sphere, even if they roll far away from each other, they will come back again. They will have another chance to interact, and they may come together depending on the conditions (Figure 10C).

I also find it fascinating that which balls gather together has the potential to flexibly change depending on the situation. In some situations, it is OK for some balls (oscillators) to rotate around without being entrained. These balls also remain on the sphere and do not disappear into the distance of infinity, so there remains the possibility that they will synchronize with other balls when the conditions or situation change.

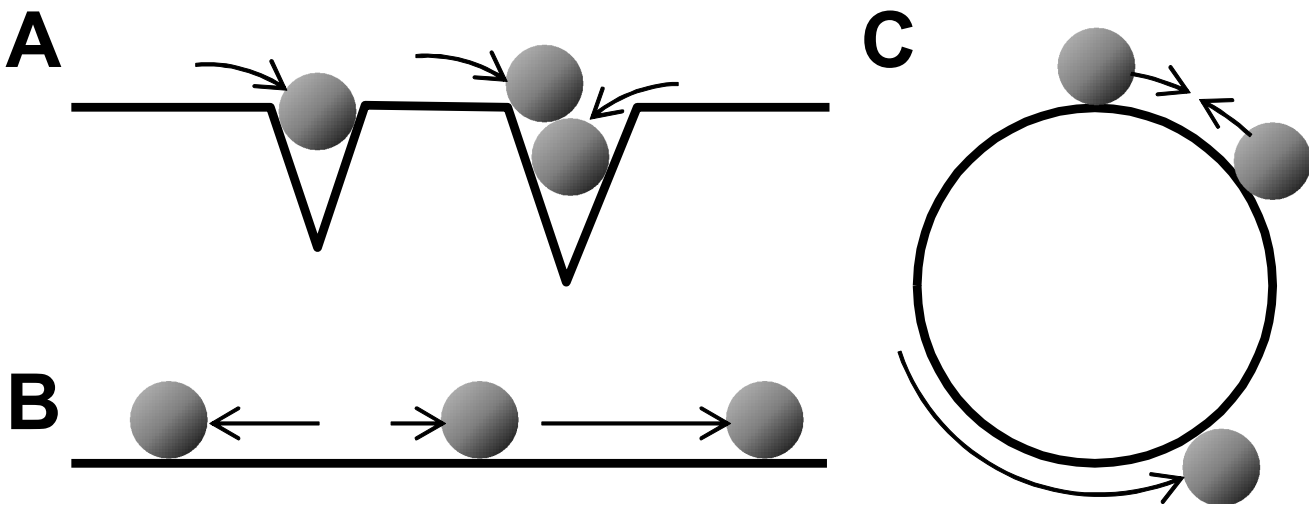


Figure 10. Comparative images of oscillating and non-oscillating interactions.

BOX *A Very Short Introduction to Chaos*

Unfortunately, there is little mention of chaos in this book. However, it is necessary to know the outline of it when talking about complex systems.

The first discovery of chaos was made by the American meteorologist Lorenz. The equation that led to his discovery is the following third-order differential equation:

$$\dot{x} = -\Pr x + \Pr y$$
$$\dot{y} = rx - y - xz$$
$$\dot{z} = xy - bz$$

For example, with Pr=10, b=8/3, r=28, a waveform shown in Figure 1 appears.

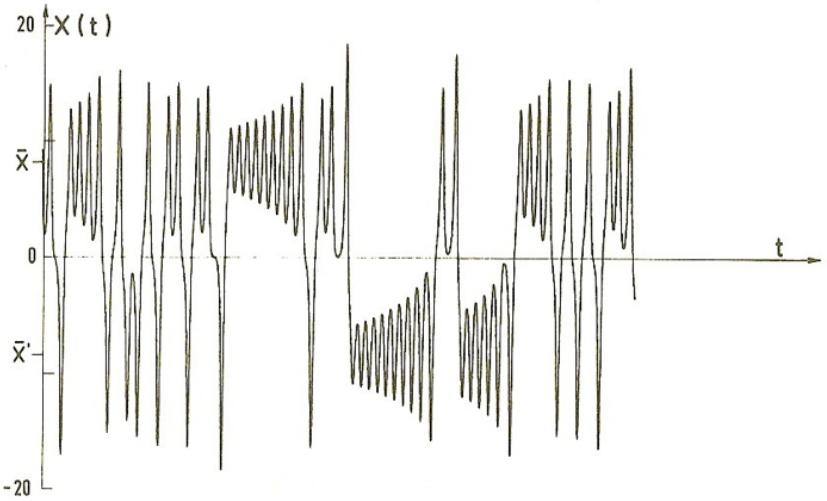


Figure 1. Example of Lorentz chaos waveform. Time development of x. Horizontal axis is time (from Ref. 8).

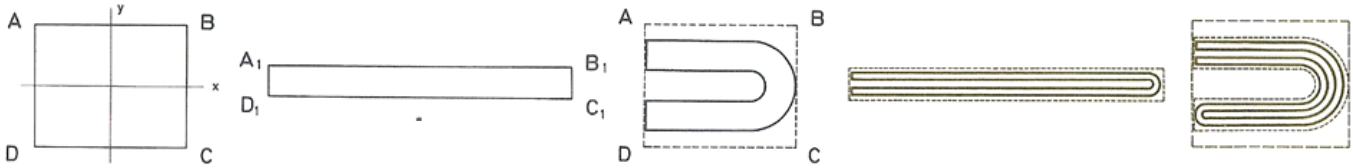


Figure 2. "kneading pie while snacking" proceeds from left to right (from Ref. 8)

The following is not a mathematically rigorous discussion, but chaos has the following characteristics: 1) It is somewhat periodic, but also random. Therefore, for chaos to emerge, two dimensions are required for the periodic behavior and at least one dimension for the randomness, for a total of three or more dimensions. 2) Even slight differences in the initial values can result in significantly different behavior. However, 3) it appears to converge to a limited region. Because of these properties, it is also called a strange attractor.

How should this behavior be understood?

I will not go into details, but examining the Lorenz chaos equation, we can see that it diverges in one direction in the three-dimensional space of variables x, y, and z. On the other hand, we can also derive that the phase volume (the range of variation in initial values) shrinks.

To understand the behavior, it is best to imagine "kneading pie while snacking" (Fig. 2). Let us consider the pie before kneading to be the phase volume. Since the direction of extension is the direction of divergence, we can now understand that a small difference in the initial value will cause a large difference in the subsequent behavior. But we can also understand that because it pinches off, the phase volume itself gets smaller and smaller, i.e., it appears to converge to a certain limited area. It is also easy to imagine that because it is folded, it does not diverge but becomes somewhat periodic.

Because of this “pie-kneading” in chaos, the trajectory of the strange attractor in phase space has a self-similar nested structure (i.e., a fractal structure) (Figure 3).

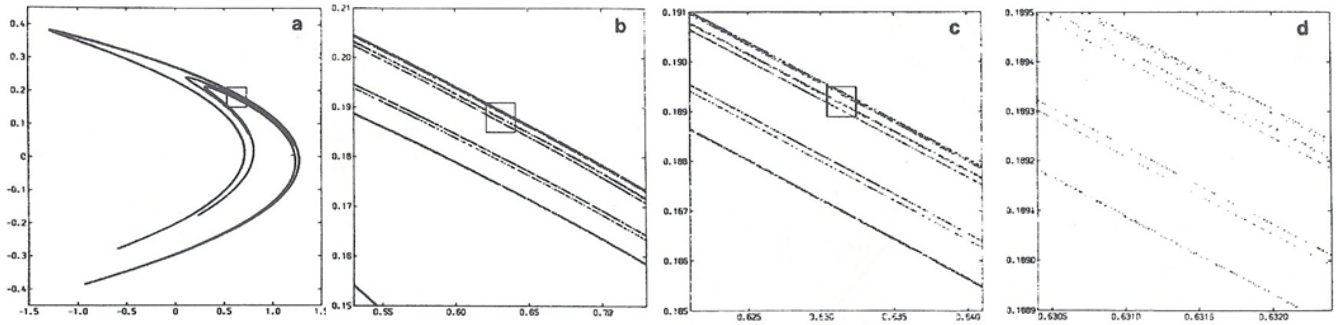


Figure 3. Self-similarity structure of trajectory in Hénon chaos. The squares in a, b, and c are enlarged to b, c, and d, respectively. Similar structures are repeated (from Ref. 8)

BOX *The End of the Deterministic Worldview*

There are "fatalists" in this world. They are people who believe that everything in this world is predetermined. This worldview, or deterministic worldview, is said to have reached its zenith at the end of the 19th century. The mathematician Laplace believed that if there existed an ultimate intelligence that could know the dynamic state of all matter at a given moment and the differential equations for its changes, such an intelligence would be able to perfectly predict the future (and the past as well). This intelligence is called Laplace's Demon.

The 20th century saw three discoveries that radically overturned this worldview: quantum mechanics, Gödel's Incompleteness Theorem, and chaos.

I won't go into the first two, but it is true that the Lorenz chaos we saw in the previous section arises from a completely causal description. Therefore, some people may think that as long as an actual object that can cause chaos can be described by differential equations, the behavior of the object can be predicted by simulations using current high-speed and high-precision computers, even if an analytical solution cannot be obtained. Unfortunately, this expectation is not met. This is because no matter how high-precision a computer is, its precision (significant digits) is always finite. As mentioned in the previous section, chaos magnifies even the slightest error. No matter how high-precision the computer is, as long as chaos appears, the simulation results and the actual behavior of the object will always be far apart.

Currently, attempts are being made all over the world to obtain previously unobtainable information through large-scale simulations. In brain research, efforts are being made to simulate the behavior of the brain as a huge neural circuit using highly realistic neuronal models. However, we should know that we can make a big mistake if we do not keep in mind what chaos tells us.

BOX *Synchronization Phenomena between People*

We often see heartwarming scenes of loving couples walking in unison. Walking is also a rhythmic phenomenon, so we can think about it from the perspective of synchronous phenomena.

Not only this example, but the phenomenon of nonlinear oscillator entrainment seems to give various suggestions about the state of human groups. Personally, I feel that I have gained various suggestions about more complex behavior than simple oscillations.

For example, the trade-off between the natural frequency difference and the coupling strength in the Kuramoto equation is also suggestive. When coupling strength is strong, a small amount of natural frequency difference is crushed and synchronized. I often feel that Japanese society is prone to falling into a "follow the crowd!" state. This is because of the strong pressure of conformity and assimilation in society, and not because of the lack of diversity of the true thoughts of individuals. On the other hand, even in a society like Germany, where individualism seems to be strong, it may be possible to take collective actions that seem extreme under certain circumstances, such as during World War II, from the perspective of "entrainment."

When I meet a close friend for the first time in in a long time and talk with him, I am sometimes surprised to find that we been reading the same books or thinking about the same things. This kind of coincidence, which cannot be dismissed simply by saying that “we have similar interests,” may be an “entrainment” in the broadest sense of the word. Some of you may think that I am not being scientific when I say this. That is true, but what I personally and the human race know and can know is only a limited part of the world. Even if the interaction is very faint and indirect, synchronization can occur when the conditions are right. It also seems premature to dismiss coincidences and synchronizations that cannot be considered coincidental as “unscientific.”

References

1) Schrödinger E. *What is life? The physical aspect of the living cell*. Cambridge University Press, Cambridge (1944).
2) *Rikagaku Jiten (5th)*. Iwanami, Tokyo (1998) in Japanese
3) http://www.ux.uis.no/~ruoff/BZ_Phenomenology.html
4) Nicolis G, Prigogine I. *Self-organization in nonequilibrium systems. From dissipative structures to order through fluctuation*. John Wiley & Sons, New York (1977)
5) Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Reports*, 2:342 (2012)
6) Haken H. *Synergetics – An introduction (2nd)*. Springer-Verlag, Berlin (1978)
7) Kuramoto Y. *Chemical oscillations, waves, and turbulence*. Springer-Verlag, Berlin (1984)
8) Berge P, Pomeau Y, Vidal Ch. *L'ordre dans le chaos*. Hermann, Paris (1984)
9) Tsuda I. *Nou no Naka ni Suugaku wo Miru*. Kyoritsu, Tokyo (2016) in Japanese

Appendix B: Parts and the Wholes – A Central Issue in Living Systems

Appendix A provides an overview of complex systems science. I noted that in systems that are out of equilibrium in the sense of thermal and statistical mechanics, that is, in systems in which there is a flow of energy and matter, temporal and spatial patterns can be generated autonomously, as if in defiance of the law of increasing entropy, which states that things move in a random direction.

Biological systems are also typical complex systems in that there is a flow of energy and matter in the form of eating and excretion, as well as fine spatial structures and periodic activities. However, there is one problem that requires special attention in living systems: the problem of parts and the whole.

Appendix B discusses the importance of the interaction between parts and the whole in biological systems as complex systems, focusing on the research on an artificial muscle called the stream cell[1)] and information processing in true slime molds (*Physarum polycephalum*)[2)] by my supervisors, Prof. Hiroshi Shimizu and Prof. Masafumi Yano.

In recent years, there seems to have been a great increase in the number of researchers studying living systems and the brain from a complex systems perspective. However, there seems to be a general lack of awareness of this part-whole, micro-macro issue. Through this discussion, which has been repeated in the text, I hope that I will be able to somehow convey the importance of this issue to the readers.

BOX *Muscle Structure and Contraction Mechanism*

Muscles are the main organs that enable animals to move, and generate force by contracting. They are broadly classified into cardiac muscle (heart muscle), smooth muscle (visceral muscle), and skeletal muscle. Skeletal muscles are connected to bones via tendons and are composed of muscle fibers (muscle cells) (Figure A). One myocyte contains several hundred nuclei. Within the myofibers are numerous columnar structures called myofibrils. Myofibrils contract when calcium ions (Ca2+) are released into the muscle cell from an intracellular organelle called the sarcoplasmic reticulum, triggered by an electrical impulse (action potential) that spreads across the muscle cell surface membrane.

Myofibrils are separated by disks called Z- discs. The area between two Z-discs is called a sarcomere. The Z-disc has fixed fibrous structure called an actin filament (a bundle of actin molecules). Between actin filaments is another fibrous structure called myosin filament (meaning bundle of myosin molecules). The actin filaments slide along the myosin filaments, bringing the Z-bands closer together. More details of the molecular mechanism are as follows.

When muscles are at rest, a protein called tropomyosin covers the actin molecule. When calcium ions, triggered by a command from the brain, are released and bind to a protein called troponin, tropomyosin shifts, exposing the binding site of the actin molecule to myosin. Once exposed, the protrusion of the myosin molecule can bind to the actin molecule (Figure B). This protrusion repeatedly pulls the actin molecule toward it using the energy of ATP (adenosine triphosphate). Specifically, steps (1) to (3) are repeated as long as calcium ions and ATP are present (Fig. C). (1) When inorganic phosphoric acid or ADP (adenosine diphosphate) is released from the protrusion of the myosin molecule bound to the actin molecule, the three-dimensional structure of the myosin molecule protrusion changes significantly, and the actin molecule is moved about 10 nm (nanometer) (This change is well known as the “swinging lever arm” theory, but there is also the “sliding” theory of myosin molecule, which is still controversial.) (2) When another ATP binds to the projection of the myosin molecule, the binding between the projection and the actin molecule is released. (3) When ATP is hydrolyzed into inorganic phosphate and ADP, the structure of the protrusion returns to its original state, and the protrusion binds to a different actin molecule position than (1).

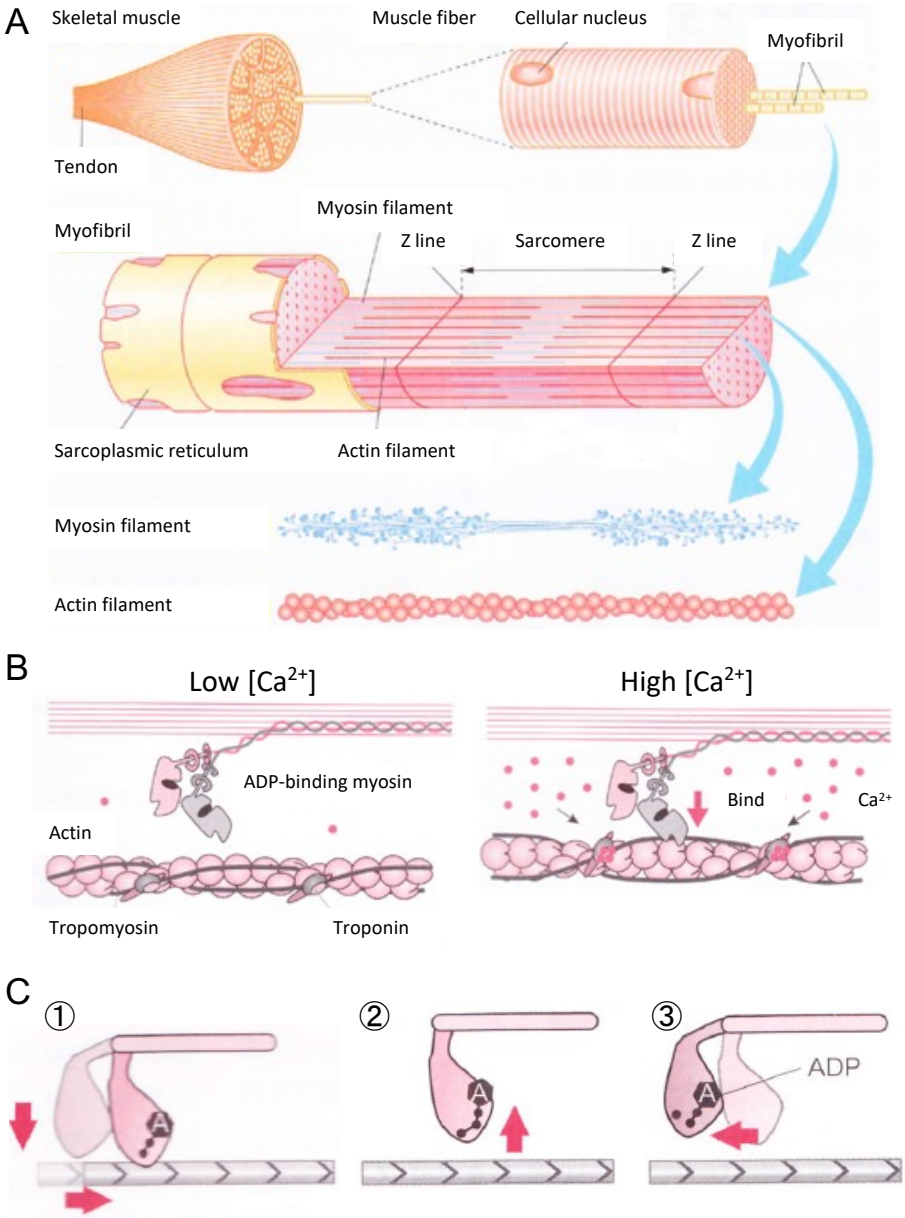


Figure. Overview of skeletal muscle structure and molecular mechanism of muscle contraction. A. Hierarchical structure of skeletal muscle (from Reference 3). B. Control of muscle contraction by calcium. C. Interaction between actin and myosin molecules (from Ref. 4).

I. "Muscles" - Interactions between the Part and the Whole

Professor Masafumi Yano, with whom I have worked for many years, created an artificial muscle motor called “the stream cell” by decomposing animal muscles into actin and myosin molecules and reconstructing them as an artificial system. Based on the results of a series of experiments using the stream cell, Professor Yano and Professor Hiroshi Shimizu established the theory that the rotation of a muscle motor is an self-organizing phenomenon. Although there are currently various approaches to artificial muscle research, the stream cell was extremely pioneering.

Muscle does not “waste” fuel, which is a mass of chemical energy, as a steam engine does by converting it into thermal energy with high entropy and then converting it into piston motion. In other words, muscles are extremely energy-efficient actuators because they convert the chemical energy of ATP (adenosine triphosphate), which is equivalent to fuel, directly into mechanical energy aligned in a certain direction, rather than converting it into randomly oriented molecular motion.

However, to elucidate the fundamental principles of muscle movement, it is necessary to clarify the relationship between a series of biological reactions involving microscopic actin and myosin molecules and macroscopic movement. Raw muscles contract, but if you try to use them as they are in experiments, it is difficult to eliminate the involvement of various molecules and structures in the muscles, which is not necessarily favorable for extracting basic principles. On the other hand, when studying biochemical reactions of actin, myosin, etc. in a beaker, the orientation of these molecules is random, and the movement of each molecule is not linked to the movement of the macroscopic solution.

The stream cell is a breakthrough that satisfies these conflicting requirements: to deal with as few types of molecular reactions as possible while generating macroscopic motion. The structure is quite simple (Figure 1). Between the inner and outer cylinders is an annular slit of 1 mm. The walls on both sides of the slit are covered with specially treated sheets called Millipore filters. Actin filaments (hereafter referred to as F-actin) obtained by processing rabbit skeletal muscle are attached to the slit in the same direction. To align the orientation, the solution containing F-actin is rotated in a certain direction. The filamentous F-actin molecules have an orientation and always attaches to the Millipore filter in one direction. If you continue to rotate the solution for a while, F-actin molecules stick to both sides of the slit. You can imagine that the leaves of water plants in a river are fluttering in the same direction. The key is the anisotropy of pasting in the same direction.

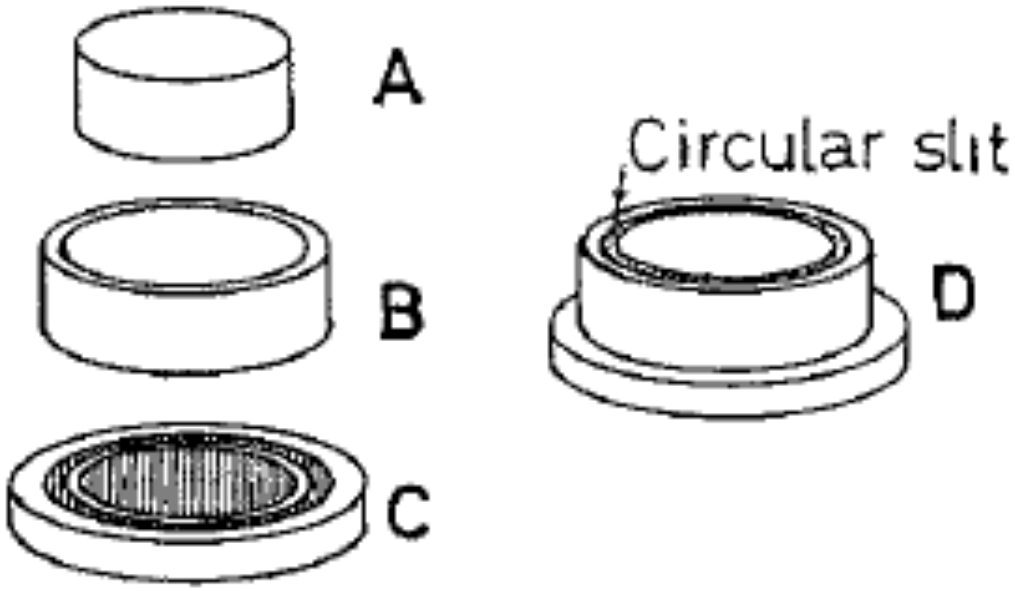


Figure 1. The stream cell. A. inner cylinder. B. outer cylinder. C. entire system combining A, B, and C (from Ref. 5).

Correlation between Flow and Reaction Rate

How are the chemical reactions between the molecules that make up muscle converted into movement? The stream cell, a device designed to conduct experiments on this, has an annular slit. F-actin molecules are attached to both walls of the slit in an aligned orientation. Another important molecule, myosin, is dissolved in water when cleaved with proteolytic enzymes to form heavy meromyosin (HMM). When this HMM and the energy source ATP are added to the slit, the circular solution begins to flow.

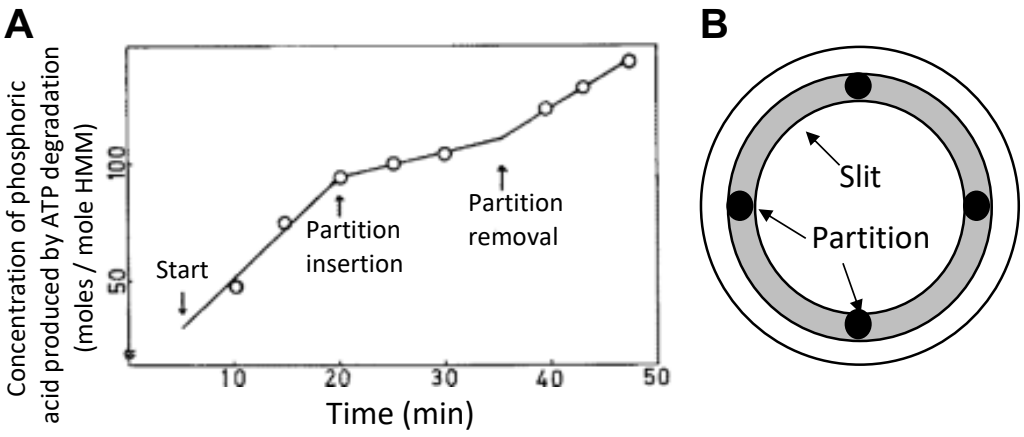


Figure 2. interaction between flow rate and reaction rate in the stream cell. A. Time course of reaction. B. Partitioning (from Ref. 6).

The direction of flow depends on the orientation of the fixed F-actin. Even if the initial flow direction is opposite to that of the fixed F-actin, it will rotate in the correct direction within 10 minutes. If the conditions of the chemical reaction, such as calcium ion concentration, are changed so that HMM and F-actin cannot break down ATP, the flow stops. If you revert the condition, the flow occurs again.

The flow lasts for about 90 minutes. During this time, the chemical reaction rate between actin and myosin molecules using ATP energy and the speed of muscle contraction (flow rate) can be measured, and the relationship between the two can be investigated in detail. The reaction rate can be quantified as the increase in phosphate concentration that occurs when ATP is degraded. The reaction rate when the solution in the slit is flowing normally at 20 micrometers per second can be seen as the slope of the graph between the start and partition insertion in Figure 2A.

Interestingly, if you put a partition in the groove to stop the flow (Fig. 2B), the rate of the chemical reaction decreases. This can be seen in the smaller slope between partition insertion and partition removal in Figure 2A. After that, when the partition is removed, the flow begins again, and the rate of ATP decomposition increases again. In other words, the faster the flow rate, the faster the reaction rate.

The solution in the stream cell begins to flow abruptly, i.e., in a phase transition manner, at temperatures above about 10°C (Fig. 3A). The reaction rate also increases at that temperature (Fig. 3B). When F-actin is fixed in an unaligned orientation, no flow occurs and the reaction rate increases linearly with increasing temperature (dashed line in Fig. 3B).

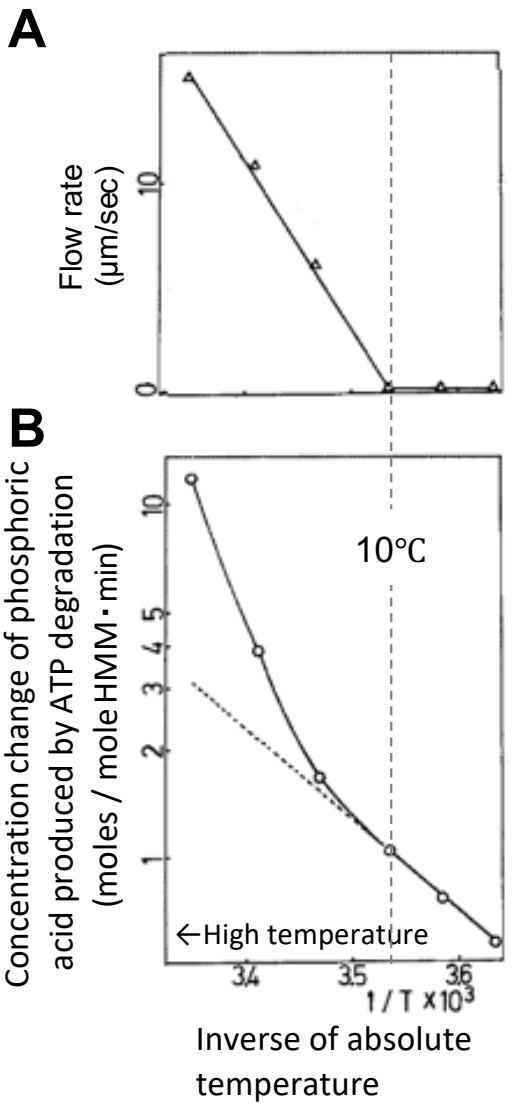


Figure 3. Temperature dependence of flow rate and chemical reaction rate. A. A. Flow suddenly occurs above about 10°C. B. Time-dependent change in phosphoric acid concentration as an index of reaction rate (from Ref. 6).

Micro-Macro Interaction

In the stream cell, the solution begins to flow when it exceeds the critical temperature. Then, the actin-myosin reaction using ATP is also enhanced. The faster the flow rate, the faster the reaction rate. The mechanism behind this is as follows. In general, a solution becomes more viscous at lower temperatures and less viscous at higher temperatures. Even at low temperatures, the actin-myosin reaction occurs. As a result, the solution around the molecule where the reaction has occurred is moved locally. However, due to the high viscosity of the solution, the solution only moves for a short distance (Figure 4A). On the other hand, as the temperature of the solution increases and the viscosity decreases, the movement of the solution caused by the movement of the myosin molecules reaches the neighboring myosin molecules and accelerates the reaction (Fig. 4B). F-actin is fixed in the same direction, so when a certain temperature is exceeded, this movement is facilitated like an avalanche across the solution. After that, the faster the flow, the faster the reaction.

Flow does not occur in a mere actin-myosin solution, i.e., in a situation where the F-actin is not fixed in the same direction. If the direction of the actin-myosin reaction is random, the direction of the induced local movement of the solution is also random (i.e., thermal), and they cancel each other out. Thus, the reactions of other actin-myosin molecules are not facilitated. The macroscopic order of flow is generated largely on the basis of the existence of a microscopic order of uniform orientation. In actual skeletal muscles, the individual actin-myosin molecules are also aligned in the same direction (so they do not generate much heat), and individual molecules have a mechanical effect on the surrounding molecules through Z membranes, etc. Thus, the importance of structure in the emergence of dynamic order that the stream cell tells us about seems to apply to actual living systems as well.

The stream cell gives us further insight into the nature of living systems. In particular, it is significant that while anisotropic reactions of actin and myosin at the microscopic level cause macroscopic solution flow, the macroscopic flow enhances the microscopic reactions, that is, the coordination, cooperation, and synchrony between individual actin-myosin reactions (Figure 4C). Micro-macro interactions are important in living systems. This is an important theme that comes up repeatedly in this book.

The fact that the microscopic and macroscopic orders are generated through cyclical interactions, like a chicken or egg situation, also means that the flow of the stream cell is self-organized synchronously. The significance of synchronization in this life system has been repeatedly discussed in the main text.

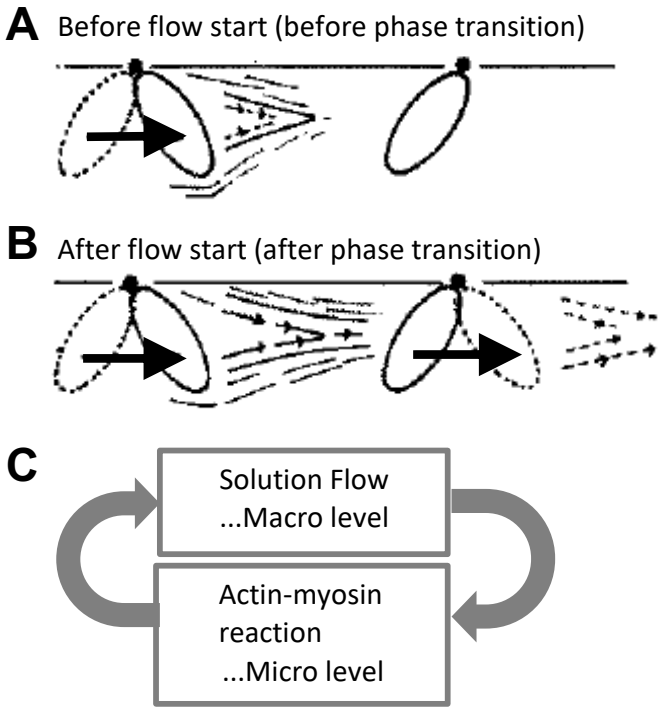


Figure 4. Mechanism of the stream cell. A. Below the critical temperature, individual actin-myosin react independently. B. Above the critical temperature, flow synchronizes the timing of the actin-myosin reaction. c. Macro and micro interact in the stream cell (from Ref. 7).

The Coherence of Parts and the Whole in Living Systems

In the artificial muscle, the stream cell, micro-level actin-myosin reactions caused the solution to flow, while macro-level flow promoted the cooperation and synchronization of individual actin-myosin reactions. The interactions between micro and macro, or between parts and the whole, were important for the stream cell to generate stable flow.

Some readers may be thinking, “Is that really such a big deal? Isn't there a lot of stuff out there?” Certainly, when it comes to young women's fashion, for example, individual tastes shape an overall trend in society, but each individual is influenced by the trend. There are many other similar examples. So, at what scale does such a micro-macro interaction begin? Is it at the society or group level? At the individual level? Or at the even more microscopic molecular level?

Currently, life science researches at the molecular level are very active. It is no exaggeration to say that most of them are. Of course, they are very important and there is still much to be elucidated. However, when I talk with researchers who are mainly engaged in such research, I truly feel that their vision of life is extremely element-reductionist. In other words, it feels like various molecules are working (reacting, etc.) independently and precisely. If we compare it to an orchestra, it would be like an individual player playing without listening to the other players, but the whole orchestra miraculously sounds like a piece of music because each player's sense of rhythm is so completely accurate and identical. However, this is unlikely. In contrast to the stream cell, the actin-myosin reaction in a test tube only reacts individually; the solution does not move macroscopically.

So what about the BZ reaction, the chemical reaction described in Appendix A? In the BZ reaction, macroscopic oscillations and patterns emerge. This means that individual chemical reactions are synchronized with nearby reactions through diffusion, etc. It is indeed a micro-level mutual entrainment between nonlinear oscillators. If you compare it to an orchestra, it is as if each performer only hears the performance around him or her. This may be fine for a chamber orchestra with a few musicians, but it is doubtful that it would work well for a symphony with many musicians. The reason why a conductor is needed is because it does not really work. A biological system consists of far more members than the performers needed for a symphony. Unlike the stream cell, the BZ reaction has an isotropic local reaction, so the reaction and the flow do not interact (this is called the Curie-Prigogine principle), i.e., there is no interaction between the micro and the macro.

It is important to note that in the stream cell, the microscopic molecular level reactions reflect the effects of the macroscopic flow. In other words, by generating macroscopic motion as a whole like a living organism, the stream cell has demonstrated the existence of microscopic and macroscopic interactions. The macroscopic pattern of the BZ reaction is merely the result of local interactions, but in a living system that must survive as a whole, it is important that the parts and the whole actively maintain consistency. We will explore this point in the next section on slime molds.

II. Slime Mold in Which Parts "Internally Observe" the Whole

A Simple Biological Model: Slime Mold

True slime molds (hereafter referred to as slime molds) are also known as myxomycetes. They live in the shade of forests, on cool, moist dead leaves and decaying wood, etc. Slime molds have both animal and plant properties. When they are in an amoeba-like form called a trophozoite, they move through the forest and feed on microorganisms. In that sense, they are "animals". However, when they are in the form of immobile mushroom-like fruiting bodies, they also have the "plant" aspects of forming spores and reproducing.

Figure 5. True slime mold, *Physarum polycephalum*.

Mathematicians, physicists, and mathematical biologists love to use *Physarum polycephalum* for research because it is easy to raise and cultivate (Figure 5). It is a huge yellow multinucleate unicellular organism. It is still alive even if it is cut to an appropriate size. The fact that it can be cut into various shapes according to the purpose of the experiment is also convenient for the experimenter[8), 9)].

When I first came into contact with slime molds, I was struck by how different they were from humans. The human body is functionally differentiated: the circulatory system, the brain and nervous system, the digestive system, etc. In slime mold, they are all integrated into one. However, the fact that there is no need to discuss the particularities of each organ is one reason why physicists and mathematicians prefer it as a model organism.

Slime molds have a mesh-like structure, and the threads of the mesh are called plasmodium. An enlarged schematic diagram of a plasmodial strand is shown in Figure 6. The outer part of the plasmodial strand is called the exoplasm or gel, and the inner part is called the endoplasm or sol.

In the experiment, light is illuminated from under a single strand, and the local thickness is observed as the intensity of the transmitted light. If the area of the strand to be observed is covered with a cellulose membrane and left for a while at room temperature and in saturated water vapor, the thickness of the sol and gel can be observed independently (Figure 7).

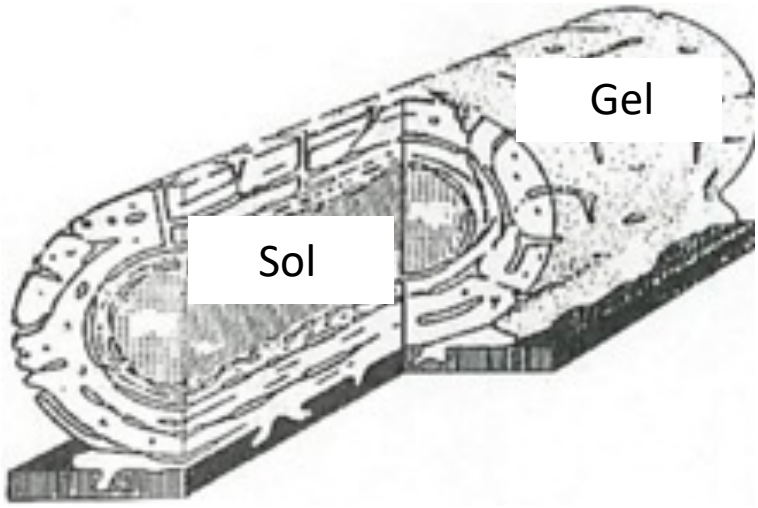


Figure 6. magnified and cross-sectional view of slime mold (from Ref. 10).

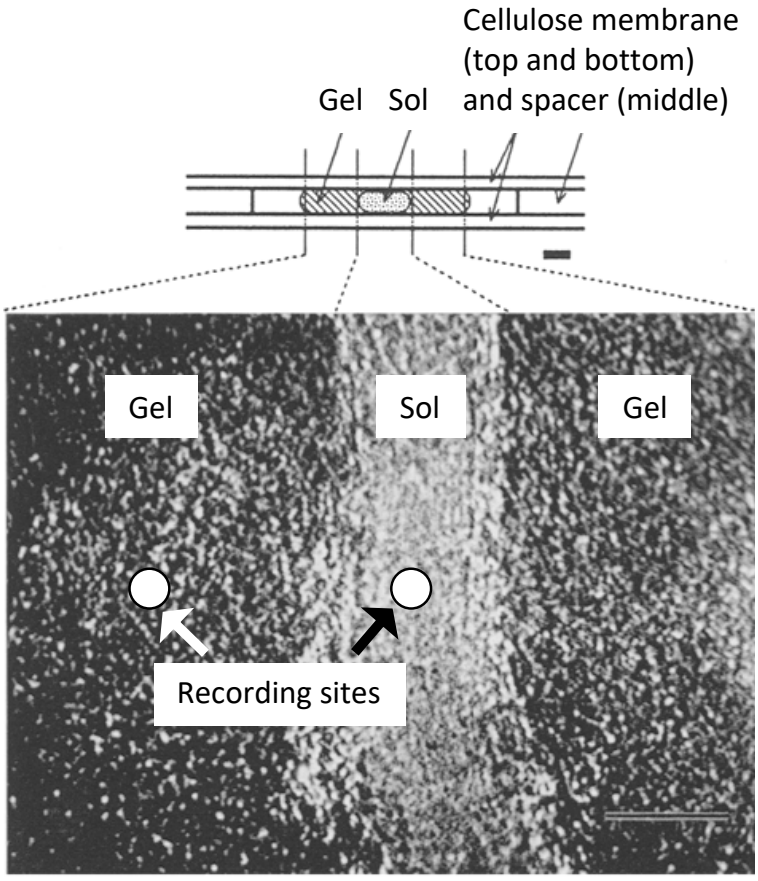


Figure 7. treatment of recording points (from Ref. 11).

Integration of the Whole through Oscillations

True slime molds are amoeba-like, multinucleate, unicellular organisms, and even when cut into small pieces, each piece continues to live and move normally. Thus, large amoeboid individuals can be regarded as a collection of elements that can live autonomously (although we cannot distinguish the boundaries between them). Moreover, they are not just a collection of elements, but are somehow integrated so that they can move through the forest as a whole, even as they transform. How are the elements or parts integrated into a whole? This is the main focus of slime mold research, and oscillation is the key.

In the body of a slime mold, everything oscillates (nonlinear oscillations, of course): ATP concentration, calcium concentration, etc. are often studied, but local thickness, etc. also oscillates. As mentioned in the previous section, by shining a light from below a plasmodium strand and observing the local thickness as the intensity of the transmitted light, it is possible to observe the oscillations of both the endoplasm and the exoplasm.

Although the endoplasm and exoplasm oscillate at approximately the same frequency, their phase relationship is not necessarily constant, and as shown in Figure 8, the phases are sometimes reversed. This suggests that the exoplasm and endoplasm can be considered as separate oscillator systems. In addition, looking at the phase difference between the exoplasm and the endoplasm between two points 200 micrometers apart, there is almost no phase difference in the endoplasm, while the phases vary relatively widely in the exoplasm. It can be assumed that the coupling between the oscillators is stronger in the endoplasm than in the exoplasm.

In the undergraduate training that I experienced at the university, we were given the task of measuring the oscillations in the thickness of the slime mold and considering its significance in terms of information processing. At that time, I was very reluctant to do this because I thought that even though there was no nervous system, it would be meaningless without measuring something that is responsible for information processing. But I was wrong. The information processing in the slime mold is not carried out by a specific substance or a specific tissue, but by the phenomenon of oscillation itself.

The fact that oscillations are used for information processing and motor control in such primitive creatures as slime mold, whose functional differentiation is unclear, is still very suggestive to the author, who has been engaged in neurophysiological research mainly on mammals. That is, it says, “oscillations come first.” Phylogenetically, animals have evolved brains and nervous systems specialized for information processing and motor control, but oscillations are more fundamental. In recent years, brain oscillation phenomena in the mammalian cerebrum such as electroencephalograms that can be recorded from the surface of the skull and their intracerebral counterparts, local field potentials (LFPs), have received renewed attention, but I think we need to take this message from the slime mold seriously. Information transmission and information processing by neuronal firing are based on LFP oscillations, and the general outline of information transmission and information processing may be preceded by LFP oscillations and synchronization between them.

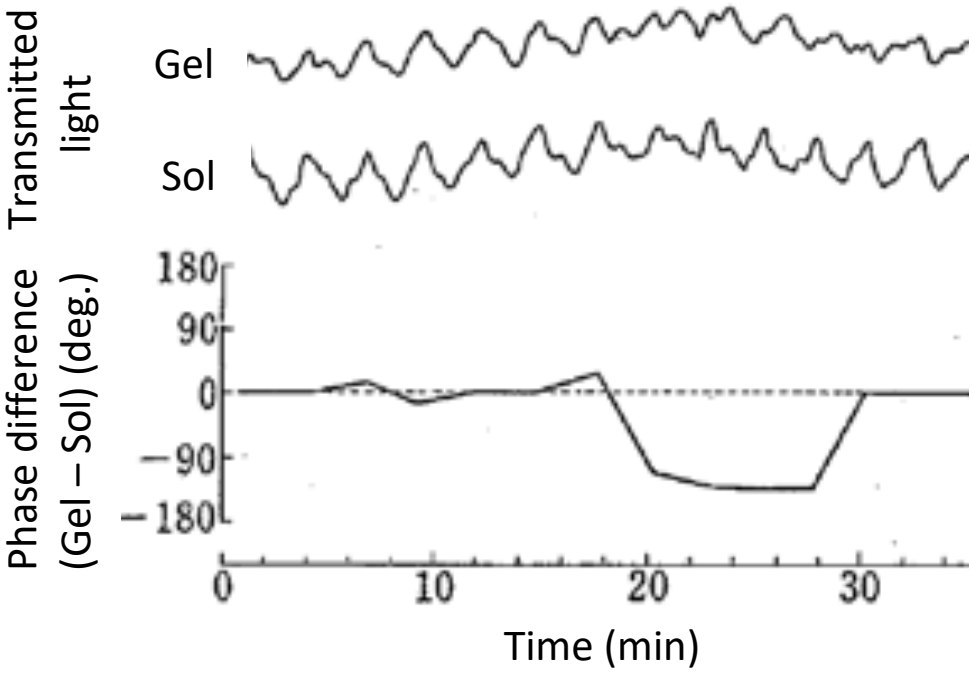


Figure 8. Example of oscillation of a plasmodial strand at rest (from Ref. 12).

Increase in Oscillation Frequency Due to Favorite Foods

Cutting out a plasmodial strand of slime mold and observing it in one dimension simplifies its basic behavior and makes it easier to model (Fig. 9A). Various innate likes (attractants) and dislikes (repellents) of slime molds are applied to one or both ends of the strand and observed. Below is a summary of the responses to the attractant stimuli (Figure 9BC).

When an attractant stimulus is applied to one end of the slime mold (Fig. 9B), the oscillation frequency of the stimulated site increases, but it also propagates to the opposite unstimulated site, where the frequency also increases to the same extent. However, the increase in frequency of the ectoplasm of the unstimulated area is delayed compared to that of the endoplasm. When air is injected into the endoplasm in the middle part of the plasmodial strand, blocking the endoplasmic interaction, the frequency of the unstimulated area does not increase even when an attractive stimulus is applied. These results suggest that the endoplasm is important for the propagation of the frequency increase.

Instead of having the same frequency at both ends, a phase gradient is created. The stimulated side is further ahead in phase. It's like two people running the same lap time on a track, but the stimulated site is in the lead. In addition, a calcium concentration gradient is created along the phase gradient. The gradient is important because it appears to be directly related to the control of the direction of movement of the slime mold. It is natural to assume that even in an amoeboid organism such as a slime mold, motor protein molecules such as actin-myosin are involved in movement, and calcium ions are related to the regulation of these reactions. Indeed, slime molds move in the direction of high calcium concentrations, i.e., in the direction of an attractive stimulus.

Surprisingly, even when an attractive stimulus is presented at both ends (Figure 9C), the body as a whole moves toward the stronger stimulus. The side presented with the weaker stimulus withdraws as if it "knows" that the stronger stimulus is present on the opposite side, even though the weaker stimulus is right in front of it. Again, the frequency of the whole body is pulled toward the higher frequency of the stronger stimulus, creating a phase gradient with that side at the top.

The information necessary for a part to behave according to its relative position in the whole, i.e. as if it "knows" its position in the whole, is called "positional information". The phase gradient seems to be related to this positional information. Before presenting a theoretical model of slime mold, I will briefly discuss the issue of positional information.

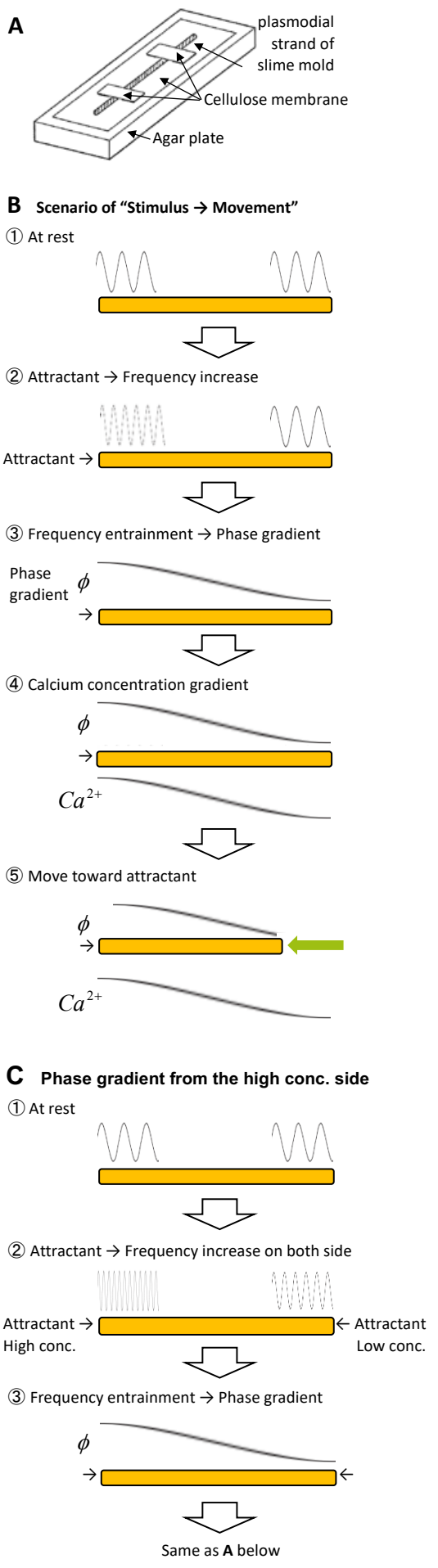


Figure 9. Response of slime mold to attractive stimuli. A. Experimental outline. B. Stimulus applied to one end. C. Both ends.

"Knowing" Where You Are

Slime molds are amoeba-like, multinucleate, unicellular organisms that lack functional differentiation of nervous, muscular, or skeletal systems. If one part is cut out, that part can still survive. However, when such "parts" come together to form a larger individual, they behave as an integrated "whole. Moreover, not only are the parts and the whole consistent, but the state of the parts changes according to their position in the whole. The information necessary for a part to behave as if it knows its position in the whole is called positional information. Positional information is a problem not only in slime molds, but also in many other living organisms.

The problem of positional information is particularly highlighted during the process of an individual developing from an egg. Figure 10 outlines an experiment conducted on a chicken at one stage of its body formation in the egg. Inside the egg, the legs and wings begin to form as protruding structures at about the same time. The cells of these two structures are very similar, and at this stage there is no obvious indication of the pattern of skeleton that they will later develop into, such as a thigh or a fingertip. At this stage, a mass of cells is cut out of the thigh in the area that will become the leg in the future, and transplanted to the position that will be the tip of the wing in the future. Strangely enough, when observing subsequent development, the site where the cells were transplanted becomes neither a wing nor a thigh. Instead, a toe is formed.

This result shows that the transplanted cells change depending on where they are transplanted. It seems that the transplanted cells were already preparing to become a leg. Therefore, they could not become a wing. However, the transplanted site was where they would have to become the "tip" in the future, rather than the "base" like the thigh. It is thought that the transplanted cells received some information about where they were being transplanted and changed accordingly. That is why they differentiated into the 'tip' of the toe.

Currently, developmental biology is very active in relation to regenerative medicine research. Genes involved in different developmental phenomena are being identified one by one. However, such research alone does not answer the questions: Why is this gene expressed at this time? Why is this gene expressed in this location? Why does it take the form that it does? Of course, some kind of mathematical research should be done in parallel, and it seems that such research is already underway. There will be many mathematical difficulties, but I am sure that the problem of positional information is an important and difficult problem that lies at the heart of the related problems.

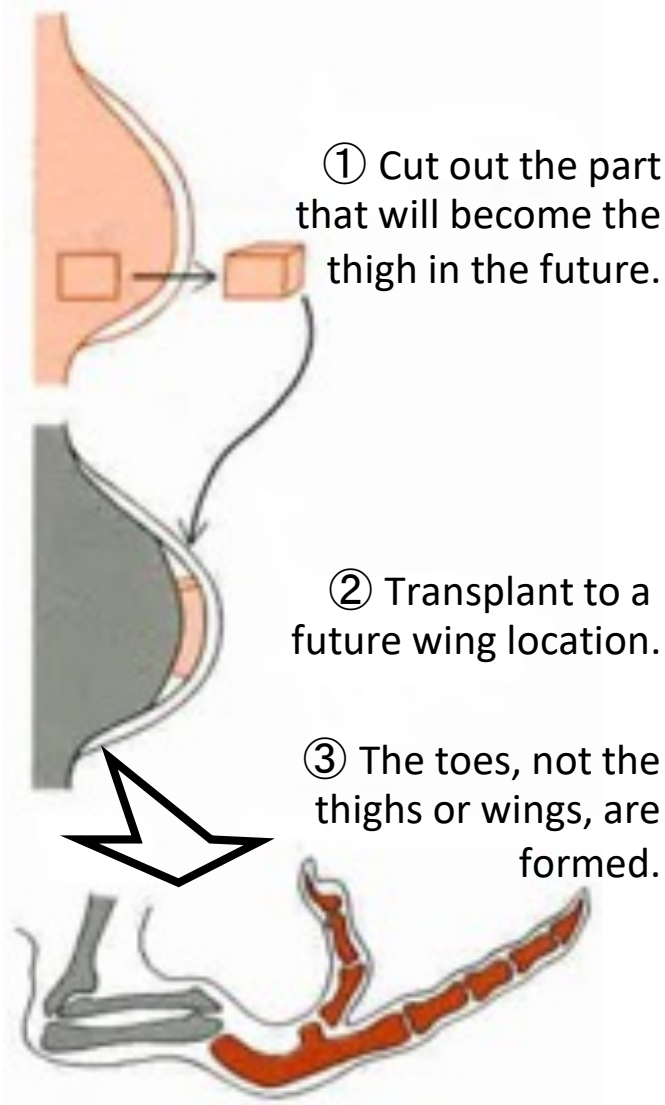


Figure 10. When a part of a chicken embryo is transplanted to a different position, it is affected by the position (from Ref. 14).

“Internal Observation” of Phase Gradient

Slime mold, an amoeba-like organism, is a collection of parts that can live independently but behave as a unified whole. For this to happen, each part must behave based on its position within the whole. The information necessary for this behavior is called positional information. Since amoebas vary in size, positional information should be relative to the overall size.

Let's review the behavior of the slime mold, simplified by cutting it into a string. When an attractant stimulus is applied to both ends of a slime mold, the oscillation frequency increases at both ends. However, the higher concentration of the stimulus causes a phase gradient from the higher concentration to the lower concentration as the frequency is entrained across the whole. This determines the polarity of the overall movement. Specifically, a calcium concentration gradient is generated that controls the motor system.

The central issue is the phase gradient. The calcium concentration gradient directly drives the movement. If the calcium concentration gradient changed frequently, the whole slime mold would wander around, making overall behavior inefficient. The calcium concentration, which reflects the overall decision-making process, should not change too much. On the other hand, the interaction of local oscillations can achieve a globally consistent state relatively quickly. The oscillations exhibited by slime molds are regarded as nonlinear oscillations. The phase gradients of the oscillations observed in the experiment, reflecting the direction of movement, are likely to be the result of the interaction of the nonlinear oscillations between the subparts.

The mechanism by which a local frequency change turns to a phase gradient is known in systems consisting of multiple coupled nonlinear oscillators (coupled oscillator systems). An oscillator with a high frequency entrains the surrounding oscillators to its own frequency, and forms a phase gradient with this oscillator as the leader (i.e., the source of the phase). This coupled oscillator system also has the property of competition. That is, even if several different oscillators increase their frequencies, the one with the highest frequency will eventually make the whole system to entrain to its own frequency, generating a phase gradient. These properties are quite similar to the properties of oscillation exhibited by slime molds. In a sense, the process of creating this phase gradient could be understood as a decision-making process for the slime mold.

The problem is that the phase gradient is only visible when the slime mold is viewed as a whole. The experimenter can observe the slime mold from the outside and know if there is a phase gradient in the oscillation. However, slime molds have no nervous system or other means of directly transmitting information about the state of distant body parts. Therefore, each part of the slime mold needs a mechanism to "know" its own position in the phase gradient from within. When modeling slime mold, it is crucial to consider such a mechanism for internal observation.

Phase Gradient is Reflected in Amplitude Difference

At each position of a string-cut slime mold, everything oscillates, including thickness and ATP concentration. When the preferred food is given to one end of the slime mold, a phase gradient in the oscillation is generated, i.e., the phase on the side of the preferred food is advanced. Then, according to this phase gradient, a calcium concentration gradient is generated, and the slime mold actually moves in the direction of the food. To understand this process, we need a mechanism that converts the phase gradient of the oscillation into a calcium concentration gradient. However, the phase gradient is only observable to the experimenter who can view the entire slime mold, and the mechanism by which each part of the slime mold internally observes its own positional information, that is, whether it is in an advanced or a lagging phase position, remains a mystery.

My former colleagues, Dr. Haruki Miura and Professor Masafumi Yano, proposed an interesting theoretical model based on a series of experimental results. Although the final version of the model was simpler, the following explanation is based on the initial model, which corresponds to the above findings on slime molds and is intuitively easy to understand.

Dr. Miura placed nonlinear oscillators in the endoplasm and exoplasm of the slime mold, respectively, so that the oscillators of the endoplasm and exoplasm at the same position are always synchronized. He also simplified the experimental results by assuming that only the endoplasm oscillators were diffusively coupled between adjacent oscillators (Fig. 11A). During computer experiments on this coupled oscillator system, Dr. Miura noticed something interesting. The amplitude of the exoplasmic oscillation at the leading edge of the phase gradient was larger than the amplitude of the endoplasmic oscillation at the same position, while the opposite occurred at the trailing edge of the phase gradient (Fig. 11B). He wondered whether this phenomenon could be used for internal observation of the positional information of the phase, and showed that there is a monotonic relationship between the phase source/sink (the leading edge of the phase gradient is the source and the trailing edge is the sink) of the phase and the difference between the endoplasmic and exoplasmic oscillations (Fig. 11C). Based on this, we formulated a mechanism that generates a calcium concentration gradient from the difference in amplitude between the endoplasm and exoplasm, instead of a phase gradient that cannot be directly observed by the parts, and qualitatively reproduced the series of behaviors of slime molds described above (Fig. 11D).

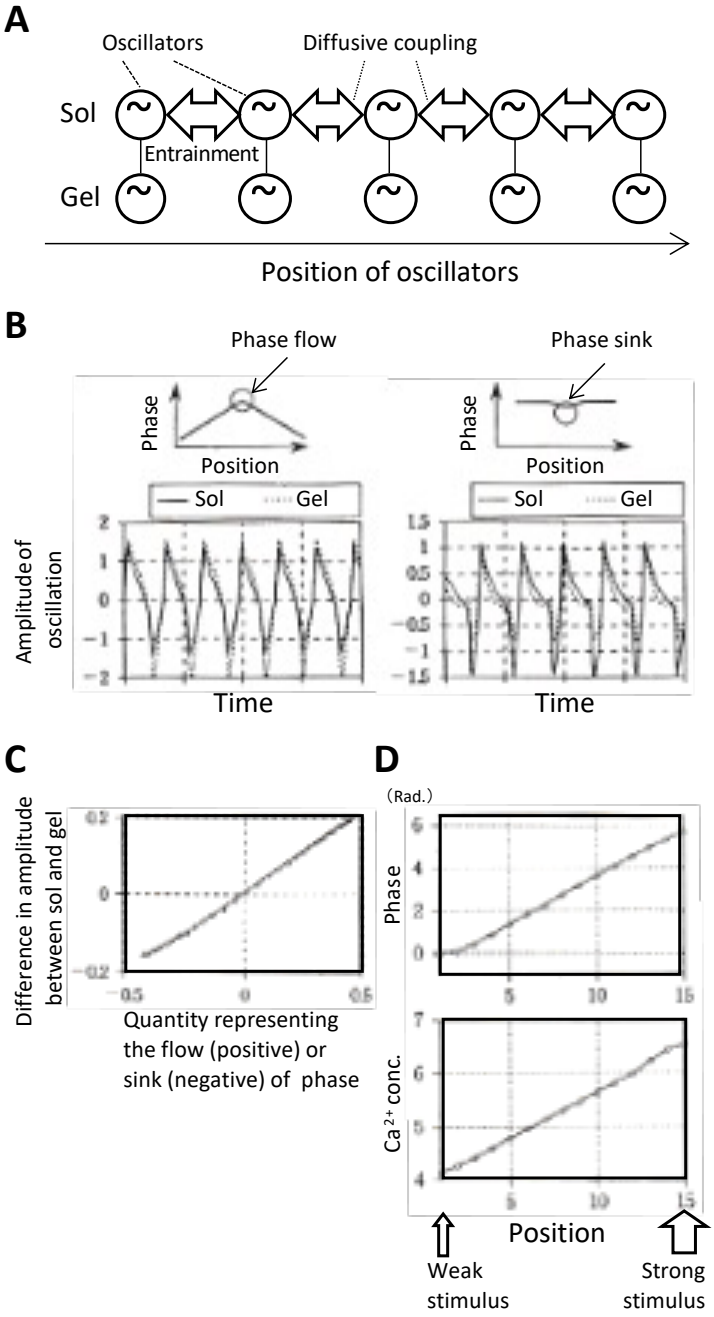


Figure 11. Miura-Yano slime mold model. A. A nonlinear oscillator is placed in the endoplasm and exoplasm at each position. Adjacent endoplasmic oscillators are diffusively coupled. B. Example of endoplasmic and exoplasmic oscillations at phase source (starting point of phase gradient) (left) and sink (end point) (right). C. Correspondence between phase source/ sink and amplitude difference. The quantity corresponding to the amplitude difference is the difference between the covariance of the endoplasmic/exoplasmic amplitudes and the variance of the endoplasmic amplitude. D. Example of phase difference → calcium concentration gradient reproduction (modified from Ref. 15).

Oscillators Coupling and Amplitude Modulation

Slime molds, which do not have information processing organs like the nervous system, behave in an integrated manner as a whole through the interactions between the oscillations of each part of their body, particularly through the generation of a phase gradient. However, the phase gradient is only visible when observed from the outside. Each part of the slime mold needs to observe the internal equivalent of a phase gradient and behave based on its position in the whole organism. In constructing a theoretical model for obtaining this positional information through internal observation, my former colleagues, Dr. Haruki Miura and Professor Masafumi Yano, took advantage of the fact that the start and end of the phase gradient correlate to the amplitude difference between the local endoplasm and exoplasm.

When an attractant is applied to a location, the exoplasm at that location increases its frequency and amplitude. The endoplasm at the same location initially oscillates at the same frequency as the exoplasm. However, this part of the endoplasm interacts diffusely with the neighboring parts and entrains them to the same frequency. A phase gradient is then generated and the stimulated site becomes the phase leader. On the other hand, the amplitude diffuses to the surroundings and eventually becomes equal. At this moment, the amplitude of the exoplasm is greater than that of the endoplasm.

In recent years, phase oscillators (Kuramoto oscillators) are often used in oscillator research because of their simplicity, ignoring changes in amplitude. However, it should be noted that the Miura-Yano model could not have been developed by focusing on this approach alone.

Importance of Size-Invariant Positional Information

The Miura-Yano slime mold model introduced so far also has a mechanism for obtaining size-invariant positional information. Size-invariant positional information means that a part behaves in a way that reflects its relative position in the whole, e.g., its state and behavior change depending on whether the part is located at one-third or two-thirds of the whole body.

In the Miura-Yano model, the oscillation phase gradient is converted into a calcium concentration gradient. The model includes a mechanism that ensures that the calcium concentration does not exceed an upper limit or fall below a lower limit, and that the gradient between these limits is smooth. The calcium concentration allows each part to "know" its relative position within the whole.

Unfortunately, the issue of positional information is not explicitly addressed in the main body of the book. However, I think that the importance of this issue should be properly recognized, especially now that research in regenerative medicine, triggered by the discovery of iPS cells, is accelerating. Without the help of theory, it would be impossible to understand when and where certain genes are expressed and what patterns are formed in the process of morphogenesis. The theory should be size invariant. For example, mice and rats are similar but quite different in size. However, it is unlikely that the position of ears and tails are determined by completely different mechanisms. There must be some mechanism that determines relative position independent of size.

References

1) Shimizu H. *Seimei wo Torae-Naosu (2nd)*. Chuko Shinsho, Tokyo (1990) in Japanese
2) Yano M. Hi-bunri no kagaku IV, Iichiko 111:104-114 (2011) in Japanese
3) Akasaka, K. *Yoku Wakaru Seibutsu Kiso + Seibutsu*, Gakken-Plus, Tokyo (2014) in Japanese
4) Editorial Board of Life Science Textbooks, The University of Tokyo. *Seimei Kagaku*, Yodosha, Tokyo (2013) in Japanese
5) Yano M, Yamada T, Shimizu H. Studies of the chemo-mechanical conversion in artificially produced streamings I. Reconstruction of a chemo-mechanical system from acto-HMM of rabbit skeletal muscle. *J. Biochem.*, 84:277-283 (1978)
6) Yano Y, Shimizu H. Studies of the chemo-mechanical conversion in artificially produced streamings II. An order-disorder phase transition in the chemo-mechanical conversion., *J. Biochem.*, 84:1087-1092 (1978)
7) Shimizu H, Yano M. Studies of the chemo-mechanical conversion in artificially produced streamings III. Dynamic cooperativity. A new cooperativity in actomyosin systems with a polarized arrangement of *F*-actin., *J. Biochem.*, 84:1093-1102 (1978)
8) Ueda T, Nakagaki T. Nenkin koudou no jikososhiki, in *Jikososhiki* (Toko K, Matsumoto G, eds.), Asakura Shoten, Tokyo (1996) in Japanese
9) Yano M. Shinsei nenkin ni okeru jouhou tougou to undo no jikososhiki, *Keisoku to Seigyo* 29:887-892 (1990) in Japanese
10) Fleishcer M, Wohlfarth-Bottermann KE. Correlations between tension force generation, fibrillogenesis and ultrastructure of cytoplasmic actomyosin during isometric and isotonic contractions of protoplasmic strands. *Cytobiologie*, 10:339-365 (1975)
11) Tanaka H, Yoshimura H, Miyake Y, Imaizumi J, Nagayama K, Shimizu H. Processing for the organization of chemotactic behavior of *Physarum* polycephalum studied by microthermography. *Protoplasma*, 183:98-104 (1987)
12) Miyake Y, Yano M, Shimizu H. Relationship between endoplasmic and ectoplasmic oscillations during chemotaxis of *Physarum* polycephalum. *Protoplasma*, 162:175-181 (1991)
13) Natsume K, Miyake Y, Yano M, Shimizu H. Development of spatio-temporal pattern of Ca^{2+} on the chemotactic behaviour of *Physarum* plasmodium. *Protoplasma*, 166:55-60 (1992)
14) Saunders et al. The differentiation of prospective thigh mesoderm grafted beneath the apical ectodermal ridge of the wing bud in the chick embryo. *Dev. Biol.*, 1:281-301, (1959)
15) Yano M, Miura H. Ketsugo shindoushi ni yoru shinsei nenkin no jouhou shori. *Suuri Kagaku*, 408:15-22 (1997) in Japanese
16) Miura H, Yano M. A model of organization of size invariant positional information in taxis of *Physarum* plasmodium. *Prog. Theor. Phys.*, 100:235-251 (1998)

Postscript

In this book, from the perspective of complex systems neuroscience, I have described the mechanisms that make creativity possible in the following order. First, I gave an overview of complex systems science, and then we looked at the problems faced by living organisms as complex systems, as well as problems in brain and neuroscience, mainly through the research of the former Hiroshi Shimizu Laboratory. As an extension of this, I introduced research on the action planning process in the prefrontal cortex that the author worked on in the former Tanji Jun and Hajime Mushiake Laboratory. After that, I talked mainly about research from the former Yano Masafumi Laboratory on specific ill-posed problems (problems with no unique answer) that the brain solves. In the final chapter, I presented my personal views on the "love" given to "thee" as the highest constraint for facing the most fundamental ill-posed problem of living, and on actions that do not seek something in return as a responsibility that comes from relationships.

This book introduced efforts of myself and my colleagues, but the order of the contents of this book does not reflect the author's own journey. What is discussed in the final chapter came first.

When I was young, I had a deep intuitive feeling about love and thought a lot about it. For example, I still remember the episode during forced labor in " Man's Search For Meaning" (1946) by Victor Frankl, who wrote about his experiences in a Jewish concentration camp.

Occasionally, I looked at the sky, where the stars were fading and the pink light of the morning was beginning to spread behind a dark bank of clouds. But my mind clung to my wife’s image, imagining it with an uncanny acuteness. I heard her answering me, saw her smile, her frank and encouraging look. Real or not, her look was then more luminous than the sun which was beginning to rise. A thought transfixed me: for the first time in my life I saw the truth as it is set into song by so many poets, proclaimed as the final wisdom by so many thinkers. The truth – that love is the ultimate and highest goal to which man can aspire. Then I grasped the meaning of the greatest secret that human poetry and human thought and belief have to impart: *The salvation of man is through love and in love*. I understood how a man who has nothing left in this world still may know bliss, be it only for a brief moment, in the complementation of his beloved.

I felt that the root of creativity lies here. At the same time, I came to the idea that what remains unchanged no matter what happens is "love" as something "given" to "thy". Even if I were to encounter injustice and be faced with death, I came to think that this is what is certain in this world. If you ask for something in return, there will always be room for betrayal. Furthermore, if you are connected to the world based only on attributes, that is, if you are connected only as "I-it", it is easy to have an exchangeable relationship with others. I thought that continuing to "give" to "thee" without asking for anything in return is something that springs (or must spring) from "I" and is therefore something that cannot be betrayed, cannot be replaced, and cannot be changed. As a result, "I" will reach an unspecified and irreplaceable relationship with undefinable "thy." I believe that what Frankl means by “pure bliss" is being able to position oneself as an irreplaceable and certain existence in an indefinite environment, even in a situation of miserable death. When Jesus Christ was crucified, he recited from Psalm 22 of the Old Testament, "Lord, Lord, why have you forsaken me?", to Psalm 31, "Lord, into your hands I commend my soul." By dedicating himself to the cross and entrusting his soul to the Lord's hands, Jesus may have been trying to prove to many people the existence of “bliss."

Then I wondered how I could get many people to realize the importance of "love" as something to give to "thee," and I thought that this idea could be achieved if it could be understood as a science. This is because science occupies an extremely large place in modern society. When something is called science, it is often assumed to be somehow correct. There is a reason for this. Science is not about rockets, computers, or DNA. It is a set of rules or etiquette for sharing the minimum of what is correct or important, even among people with different ideologies, beliefs, and claims. In today's world, where people travel frequently and there are many opportunities for people of different ideologies and beliefs to interact, it makes sense that

scientific thinking and judgment are important.

After that, I tried to discuss with many people whether the science of "love" and the science of "I-Thou" are possible (it sounds so suspicious that it makes me sick just to write it), but no one understood. Not only would they not understand, they would not take me seriously. Finally, one night in my sophomore year, I knocked on the door of Professor Hiroshi Shimizu's office. He must have been very busy, but he listened carefully to what I had to say. And at the end, he said briefly, "This is not easy. I was very touched because I felt that Dr. Shimizu recognized the existence of this problem, its importance, and the difficulty of dealing with it in science at that time. I thought, "I see. It's not easy even for Professor Shimizu. Then I should fully understand what he has been working on and what he thinks. As an extension of this, it may be possible to approach the problems of life and "love" in a unified way with the study of molecules and atoms. That's how I made up my mind, and I've continued to thrive to this day, barely surviving as a scientist.

This book is also my journey. I apologize for the incomplete content, but I have tried to give a rough overview of the topics of atoms and molecules, life, brain, and love as one connected story. I would be more than happy if the ruthless approach of this book could be a light for those who think, "There is no such thing as love in this world.

My friends Katsuhide Oishi, Kenichi Machida, and Osamu Wada read two pages of the manuscript each week from the perspective of a general reader, which was a great encouragement to me in writing this book. Mr. Machida, in particular, gave me sharp but thoughtful comments each week. If there is any aspect that is easy for the general reader to understand, it is largely due to him. Professor Hiroshi Shimizu, Professor Masafumi Yano, and Dr. Naoyuki Sato read parts of the drafts of the preface, Appendix B, and Chapter 5, Part 2 of II, respectively, and provided valuable comments that helped me gain a deeper understanding. Discussions with researchers from the Grant-in-Aid for Scientific Research on Innovative Areas "The study on the neural dynamics for understanding communication in terms of complex systems," "Elucidation of neural computation for prediction and decision making: toward better human understanding and applications," and "Non-linear neuro-oscillology: toward integrative understanding of human nature" were also of great help in writing this book.

Needless to say, the research I have conducted and mentioned in this book was made possible thanks to the guidance, cooperation, and understanding of my co-researchers and colleagues, including Professor Hajime Mushiake. I also received support from the NPO NeuroCreative Lab and the HAYAO NAKAYAMA Foundation for Science & Technology and Culture. I am grateful to Dr. Ichiro Tsuda and Dr. Shigetoshi Nara for their advice on the publication. Mr. Junsei Kishi of the University of Tokyo Press appreciated the planning of this book and gave me specific advice on how to make it more readable. I would like to take this opportunity to express my sincere gratitude to all these people. However, the author takes full responsibility for the content of this publication.

Thank you for reading to the end.