**A new theory of investment**

 Jing Chen

 School of Business

 University of Northern British Columbia

 Prince George, BC

 Canada V2N 4Z9

 Phone: 1-250-960-6480

 Email: chenj@unbc.ca

 Web: <http://web.unbc.ca/~chenj/>

 **Contents**

1. Major Factors in Biological and Social Systems

* 1. A common measure of performance
	2. The necessity of fixed costs
	3. Lifespans of organisms and investment projects
	4. Uncertainty
	5. Discount rate
	6. The volume of output, market size and abundance of resources
	7. On Demand and supply
	8. Concluding remarks

2. Resource and Technology

2.1. The importance of natural resources

2.2. Natural resources: Diverse forms and unifying principle

2.3. Living systems: Positive return technologies of utilizing resources

2.4. Human technologies and human societies

2.5. On inequality

2.6. Carbon and hydrogen as energy sources

2.7. Some patterns in energy economics

2.8. On the concept of renewable energy

2.9. Global warming and climate change: Some paradoxes

2.10. Concluding remarks

3. Production: A Mathematical Theory

3.1. A historic review of related ideas and mathematical techniques

3.2. A mathematical theory of production

3.3. Several numerical examples

3.4. Monetary policies and business cycles

3.5. Equilibrium and non-equilibrium theory

3.6. Physics, mathematics and predictability

3.7. Concluding remarks

4. Languages and Cultures: An Economic Analysis

4.1. Introduction

4.2. An application to cultural analysis

4.3. The evolutionary patterns of languages

## 4.4. Concluding remarks

**Chapter One: Major Factors in Biological and Social Systems**

* 1. **A common measure of performance**

For any biological or social system to survive and prosper, it has to provide a non-negative rate of return. In finance theory, the performance of a business is measured by its rate of return on monetary investment. In biological theory, the performance of an organism is measured by its rate of return on biological investment, or change of population size. In energy extraction, the rate of return is measured as energy return over energy invested.

However, in established economic theories, human beings and social systems are assumed to maximize utility instead of generating positive return. From our daily experience, when we are hungry, we try to obtain food; when we are thirsty, we try to obtain water; when we are sexually mature, we long for a mate. Our preferences and activities are indeed directed by our short term needs, or utilities. But short term needs and “utilities” are ways to achieve long term goals of positive rate of return. Over the long term, whether a system survives and prospers depends on its rate of return. A business will fail if it loses money over the long term, whether or not it maximizes its utility. An investor will have to exit the stock market if he loses all his money, whether or not he maximizes his utility. A species will go extinct if its numbers continue to decline, whether or not it maximizes its utility. This is not to suggest any short term activities we perform will promote positive returns. We are all constrained by the structures of ourselves and the environment. We all have limited capacity to forecast and respond. After all, most species that appeared on the earth went extinct; most businesses that once existed closed down. But every organism survives to this day is an unbroken chain of successful reproduction for several billion years. Every business or organization that survives to this day operates successfully since its inception. For example, Christianity is a successful social system that has been in operation for about two thousand years. As researchers of biological and social systems, we hope to understand why a system does well or fails.

Establishing an objective measure of performance does not mean the human mind and emotion are not important in understanding human societies. On the contrary, the human mind and emotion have evolved to generate positive biological returns. It is our love of children that enables us to spend tremendous amounts of effort to feed and protect helpless babies. It is because of the irresistible attraction to the opposite sex that two very different individuals live together to raise the next generation. This does not mean that our emotions always help us gain a positive return. The human mind is an evolutionary product of the past. The environment we live in today is very different from the past. What worked in the past may not work well in the future. It is comforting to know that we always maximize our “utility”, even if we don’t know what our utilities are. In contrast, a return based theory will allow us to analyze the long term impacts of our own behaviors, the strategies of our businesses, the structure of our societies, and the policies of our governments.

If a return based theory provides much clarity about our society, why the dominance of a utility based theory in economics? We can go back to the very first paper that introduced the concept of utility in economics. It was written by Daniel Bernoulli, originally published in 1738. In that paper, he showed that the arithmetic rate of return does not provide a good measure of return and proposed to use the concept of utility to replace arithmetic rate of return. Today, it is recognized by many people that the geometric rate of return provides a better measure of return than the arithmetic rate of return. Daniel Bernoulli’s utility function was equivalent to the geometric rate of return. If we adopt the geometric rate of return as the measure of return, we can resolve the problem raised by Bernoulli without resorting to the utility function.

By adopting a common measure of performance for both biological systems and social systems, we understand biology and social sciences as an integrated theory. Social systems, like other biological systems, are enabled and constrained by physical resources and physical principles. Social systems, like other biological systems, require non-negative return for their long term viability. There have been persistent suggestions that social theories should be built on the foundation of biology. Yet it is often assumed that there is a fundamental difference between the two: genetic mutations are generally considered random while human activities are considered purposeful. However, genetic mutations and human activities are problems at different levels. Human beings evolve through genetic mutations as well and many animal activities are purposeful. Furthermore, more precise observation shows that genetic mutations are not completely random. When, where and how fast genes mutate are influenced by many environmental factors. The regulation in genetic and epigenetic changes in organisms is highly directed to enhance their survival under different kinds of environments. Since directed and informed changes often provide a higher rate of return than complete random ones, purposeful changes evolve both in social and biological systems. But sometimes random changes can help bring new and better results to systems stuck in local but not global optimums. For example, some optimization software adopt random changes periodically to test whether the obtained results are global optimums. Because both purposeful and random activities are present in biological and social systems, there is no reason to segregate the study of social systems from other biological systems.

An integrated theory of biology and human society enables us to better understand the long term trends of human societies. After the rise of oil prices in the 1970s, many countries regained economic growth after brief recessions. Established economic theory shows us that markets can overcome scarcity of resources. But from a biological perspective, fertility in most wealthy countries dropped below the replacement rate after the 1970s, rendering their rate of return on biological investment negative. The initial drop in the fertility rate reduced the number of dependent children. This reduced the cost of raising children. Many more adults, especially women, became available as workers. Countries in demographic transition often enjoy a high rate of growth in economic output for several decades. But eventually, as the majority of the working population becomes old with far fewer next generation workers to replace them, the ratio of working population declines. Monetary activities decline eventually. Biological theory foretold the social problems of the former Soviet blocks, Japan, Europe and many other countries several decades earlier, when their biological return turned very negative. We could have avoided these social problems if our policies are guided by biological theories.

Some people contend that human beings are social animals and biological studies focus on individuals. However, many species of animals are social animals and a great amount of research has been done on social animals. Furthermore, each multicellular organism, such as a human being, is a community of different cells with different characteristics and different locations. Different cells communicate and coordinate with each other to accomplish goals that are impossible to achieve by individual cells. We can gain deep insight about communication, coordination and regulation of different cells from studies in animal physiology.

Our goal is to understand what factors impact the long term return of biological and social systems. In the next several sections, we will discuss some of the most important factors.

* 1. **The necessity of fixed costs**

Thermodynamics is often called the economic theory of nature. From thermodynamic theory, useful works can be obtained only when there exists a differential, or gradient between two parts of a system. Plants can utilize sunlight to generate chemical energy because the sun is much hotter than the earth. We can utilize water flow to generate electricity when water flows from higher place to lower place. However, fixed investment is required before positive return can be generated. For example, plants need to make chlorophyll before they can transform solar energy into chemical energy. We need to build hydro dam before electricity can be generated. In general, all organisms require a fixed set of genes before they can reproduce themselves.

In the language of economics, it requires fixed cost to transform resources to be used by all living organisms profitably. Specifically, fixed costs reduce variable costs. In general, a lower variable cost system requires higher fixed cost, although the reverse is not necessarily true. We will list several examples from engineering, biology and economics to illustrate the tradeoff between fixed and variable costs.

In electricity transmission, higher voltage will lower heat loss. But higher voltage transmission systems are more expensive to build and maintain. The differential of water levels above and below a hydro dam generates electricity. The higher the hydro dam, the more electricity can be generated. But a higher dam is more expensive to build and needs to withstand higher water pressure. In an internal combustion engine, the higher the temperature differential between the combustion chamber and the environment, the higher the efficiency in transforming heat into work. But it is more expensive to build a combustion chamber that can withstand higher temperature and pressure. Diesel burns at higher temperature than gasoline. This is why the energy efficiency of diesel engine is higher than that of gasoline engine and the cost of building a diesel engine is higher as well. In general, the higher the gradient, the more efficient energy can be transformed into useful work. This is the famed Carnot’s Principle, the foundation of thermodynamics. At the same time, creating and maintaining high differential itself is physically difficult and hence financially expensive. The economic principle is consistent with, indeed, a consequence of the physical principle.

Warm blooded animals can generate high energy output longer than cold blooded animals because their bodies are maintained at high energy levels to ensure fast biochemical reactions. So warm blooded animals can control and consume more resources than cold blooded animals. But the basic metabolism rates of warm blooded animals are much higher than the cold blooded animals. Warm blooded animals have to eat much more than the cold blooded animals of the same weight to keep alive.

Shops located near high traffic flows generate high sales volume. But the rental costs in such locations are also higher. Well trained employees work more effectively. But employee training is costly. People with higher education levels on average command higher income. But education takes time, effort and money.

The level of fixed cost of a system is often its defining characteristics. It is often used as a classification criterion in many research areas, although the term “fixed cost” is not necessarily used. In cultural study, cultures are often classified as high context cultures and low context cultures. In ecological study, species are classified as K species (high fixed cost) and r species (low fixed cost). In the study of social systems, societies are often classified as complex (high fixed cost) or simple (low fixed cost). Many debates in our societies are about fixed cost. When we say education is a right, we really mean education should be part of the fixed cost of the society.

For any living organism to be viable, its cost to obtain resources cannot exceed the value of the resources over its life cycle. Similarly, for a business to be viable, its average cost of operation cannot exceed its revenue. Costs include fixed cost and variable cost. The main decision for an organism or a business is to determine the structure of its fixed and variable cost so its total cost can be lower than its revenue in specific environments. Historically, lower fixed cost systems often appear earlier than higher fixed cost systems, which often evolve from lower fixed cost systems. Single celled organisms appear earlier than multicellular organisms. Cold blooded animals appear earlier than warm blooded animals. Small family owned shops appear earlier than large global companies. Fixed costs in wealthy societies are also higher than fixed costs in poor societies. Wealthy societies often have compulsory secondary education and easy to access university education while people in poor societies often start working at an early age. Wealthy societies often have well maintained roads, free public libraries, relatively open and fair legal systems, which require high maintenance costs, while poor societies usually don’t. Many people feel that the world will continue to evolve toward more complex, higher fixed cost systems, with only occasional setbacks. However, if we look at longer time spans and broader scales, the evolutionary patterns are subtler.

The ‘Doctrine of the Unspecialized’ ... describes the fact that the highly developed, or specialized types of one geological period have not been the parents of the types of succeeding periods, but that the descent has been derived from the less specialized of preceding ages. ... The validity of this law is due to the fact that the specialized types of all periods have been generally incapable of adaptation to the changed conditions which characterized the advent of new periods. ... Such changes have been often especially severe in their effects on species of large size, which required food in great quantities. ... Animals of omnivorous food-habits would survive where those which required special foods would die. Species of small size would survive a scarcity of food, while large ones would perish. … An extreme specialization…has been, like an overperfection of structure, unfavorable to survival. (Cope, 1896, p. 173-174)

We often have the impression that higher fixed cost systems are better than lower fixed cost systems. For example, simple organisms or societies are generally called primitive while complex organisms or societies are generally called advanced. However, biologists now recognize that systems of different fixed costs are adapted to different kinds of environments. Indeed, all life forms, simple or complex, are successful outcomes of unbroken chains of almost four billion years of survival and reproduction. The proper amount of fixed cost a system needs to invest mainly depends on the amount of resources available. In general, in an environment of abundant resources, large, high fixed cost systems dominate; in an environment of scarce resources, small, low fixed cost systems breakeven easier. However, the amount of resources a system can access depends partly on its ability to control and utilize particular resources. In general, high fixed cost systems are more capable of controlling and utilizing resources than are low fixed cost systems. The net impact can only be measured by the return of a system.

In the last several hundred years, with the large scale use of fossil fuels, higher fixed cost social systems did well most of the time, expanding their systems globally. But in the past several decades, with continuous depletion of high quality resources and continuously rising living standard, fertility rates in most wealthy countries, as high fixed cost systems, have dropped below the replacement rate. This means that the biological returns in these places have already turned negative. To mainstream economists, demographic changes are not the fundamental determinant of long term economic activities. So they could not foresee serious problems in the future. When economic recessions occur, they tend to view them as temporary breaks from the continuous growth. During the recent Great Recession, policymakers injected large amount of resources to pump up financial and auto sectors, two of the highest fixed cost industries. Policymakers expected these sectors generate high profit and large amounts of employment once economy start to grow again. However, from the perspective of resource scarcity, the bail out of these high fixed cost industries only delayed the necessary adjustments and made the whole society more vulnerable to future uncertainties.

We will look further into how the increase of fixed costs affects the rate of return. The increase of fixed cost is often associated with the division of labor. Instead of a single cell handling all biological functions, a multicellular organism often consists of many different organs, such as heart and liver, each responsible for specific functions. Unspecialized stem blood cells are capable to reproduce easily. But once these unspecialized stem cells become red blood cells and various white blood cells, they cannot reproduce themselves anymore. In general, specialized cells, such as red blood cells and neuron cells, often are infertile or less fertile. Similarly, in human societies, highly trained professionals, especially female professionals, are often less fertile. Being a parent requires a lot of general skills, such as changing diapers, feeding the babies and cooking for the family, that are possessed by so called unskilled workers but are lost or degenerated among many highly trained specialists. If the fixed cost of a system becomes too high and many people in the system become over specialized, the average fertility of the social system will drop below the replacement rate, rendering the system unsustainable.

The tradeoff between fixed cost and variable cost is universal in economic activities. However, this tradeoff is often not explicitly discussed in the same literature and often not considered in policy issues. For example, electricity generated from solar panel is considered clean energy because solar panel does not need fuels that will cause environmental problems. But the manufacturing of solar panels is highly resource intensive and highly pollutive. However, the pollution from manufacturing solar panels, the fixed cost part of the solar electricity, is rarely mentioned in policy discussion. While it is in the interest of the promoters of “clean” energy and “renewable” energy to avoid discussing such issues, a good economic theory should provide guidance to understand the big pictures.

The necessity of fixed cost in biological and social systems has broad implications. It takes time to recoup the initial investment. Therefore organisms and investment projects have lifespans or durations. How lifespans of different organisms and projects are determined? Are long life systems necessarily better? Uncertainty is an unavoidable part of nature. But the same level of uncertainty will have different impacts on systems with different fixed costs and durations. How they are related? The fixed costs are spent or committed at early stages of organisms or projects’ life. But the expected payoffs, which occur in the future, are only estimates. Therefore expected payoffs from investments are subject to discounting. How should discount rates be determined? What are the consequences of different level of discount rates? When market size is large, high fixed cost systems, with low variable cost, will generate high rate of return. Are larger market sizes always better? We will explore these and other questions in the next several sections.

* 1. **Lifespans of organisms and investment projects**

How do lifespans or durations of organisms and projects affect the rate of return on our investments? If the duration of a project is too short, we may not be able to recoup the fixed cost invested in the project. If the duration of a project is too long, the variable cost may become too high and the rate of return will turn negative. This is why individual life does not go on forever. Instead, it is of higher rate of return for animals to have finite life spans and produce offspring. In social sciences and policy discussion, longer life span is often used as an indicator of higher quality of a social system. However, societies that enjoy a long life span, such as Japan, often struggle with below replacement fertility. We will analyze the relation among fixed cost, lifespan or project duration, and rate of return in greater detail.

When the level of fixed cost increases, it often takes longer time for a project to breakeven. Large animals and large projects, which have higher fixed cost, often have longer lifespan. There is an empirical regularity that animals of larger sizes generally live longer (Whitfield, 2006). The relation between fixed cost and duration can be also applied to human relation. In child bearing, women spend much more effort than men. On average women value long term relation while men often seek short term relation (Pinker, 1997).

When the duration of a project keeps increasing, variable cost will keep increasing and the return of a project will eventually turn negative. Hence, duration of a project or an organism cannot become infinite. For life to continue, there has to be a systematic ways to generate new organisms from old organisms. From earlier discussion, for a system to have a positive return, fixed assets have to be invested first. Thus old generations have to transfer part of their resources to younger generations as the seed capital before younger generations can generate positive return. Therefore, there is a universal necessity of resource transfer from one generation to the next generation in biological and social systems. “Higher” animals, such as mammals, generally provide more investment to each child than “lower” animals, such as fish. In human societies, parents provide their children for some years before they become financial independent. In general, wealthy societies provide more investment to children before they start to compete in the market than poor societies. In businesses, new projects are heavily subsidized at their beginning stages by cash flows from profitable mature projects.

Empirical evidences exhibit the inverse relationship between lifespan and fertility. Lane (2002) provided a detailed discussion about the tradeoff between longevity and fecundity in the biological systems.

Notwithstanding difficulties in specifying the maximum lifespan and reproductive potential of animals in the wild, or even in zoos, the answer is an unequivocal yes. With a few exceptions, usually explicable by particular circumstances, there is indeed a strong inverse relationship between fecundity and maximum lifespan. Mice, for example, start breeding at about six weeks old, produce many litters a year, and live for about three years. Domestic cats start breeding at about one year, produce two or three litters annually, and live for about 15 to 20 years. Herbivores usually have one offspring a year and live for 30 to 40 years. The implication is that high fecundity has a cost in terms of survival, and conversely, that investing in long-term survival reduces fecundity.

Do factors that increase lifespan decrease fecundity? There are number of indications that they do. Calorie restriction, for example, in which animals are fed a balanced low-calorie diet, usually increase maximum life span by 30 to 50 per cent, and lower fecundity during the period of dietary restriction. … The rationale in the wild seems clear enough: if food is scarce, unrestrained breeding would threaten the lives of parents as well as offspring. Calorie restriction simulates mild starvation and increase stress-resistance in general. Animals that survive the famine are restored to normal fecundity in times of plenty. But then, if the evolved response to famine is to put life on hold until times of plenty, we would expect to find an inverse relationship between fecundity and survival. (Lane, 2002, p. 229)

Lane went on to provide many more examples on the inverse relation between longevity and fecundity.

The necessity of the fixed cost investment and the finiteness of lifespan determine that resource transfer from old generation to new generation is essential for the long term viability of a system. However the process of resource distribution is often the source of many conflicts between generations and within the members of the same generation. Each child wants more resources from parents. But parents prefer children to become independent early so resources can be distributed to younger or unborn children. Old mature industries, which need little R&D expense, prefer low tax systems. But young high tech industries, which rely heavily on universities to provide new technologies, employees and users, strongly advocate government support in new technologies. Businesses prefer lower tax rates. But educational institutions, which mainly train the younger generation and receive much of their incomes from governments, prefer higher government revenues.

The conflict between generations often starts with pregnancy. “The greater the amount taken by the fetus, the greater its birth weight, but the less its mother would have for other purposes. Lighter babies would have had a reduced probability of survival, but costly pregnancies would have increased the mother’s vulnerability to disease, reduced her ability to care for existing children, and decreased her chances of reproduction again. (Haig, 1993, p. 496).”

In our daily life, we all experience conflicts between generations. When contraception technology is available, many people decide to delay raising children to delay the transfer of resources to the next generation. With aging parents, more and more children are born with genetic defects. In many societies, the fertility rates also drop below the replacement rate because of the delaying or avoidance of resource transfer to the next generation.

The increase of retirement age is often advocated as the solution for potential labor shortage in societies with low fertility rate. But many countries with low fertility rates have very high youth unemployment rates. Younger people are more efficient in most works. Economic performance will improve if more young people can have jobs. In general, systems with below replacement rate fertility need to shorten the duration of the work period to restore the viability of the systems. If older people can retire earlier, more young people can move out of unemployment. With financial security, young people in their prime reproductive age can have more children. However, in a society with low fertility and long lifespan, the proportion of seniors will be very high. Furthermore, young people will grow old but old people will not turn young. Senior issues generally gain a lot of attention. Since unborn and small children have little political clouts, their interests are less often represented in public discussion. This poses a great challenge to societies already in a state of low fertility rate and long lifespan.

**1.4. Uncertainty**

Uncertainty is an integral part to all living systems. But different systems respond to uncertainty in different ways. In general, simple low fixed cost, short duration systems are quicker to adapt to environmental changes than complex high fixed cost, long duration systems. However, high fixed cost systems often possess more resources to work with. They often spend more resources to maintain stability in internal and external environment to reduce uncertainty. They also evolve mechanisms to adjust themselves periodically. We will use some examples to illustrate the strategies adopted by systems with different fixed costs.

Lower fixed cost systems in general have shorter life span than higher fixed cost system. The mutation rates of lower fixed cost systems are higher. This gives lower fixed cost systems advantages in initiating and adapting changes. For example, AIDS virus is much smaller than human beings and can mutate much faster. This makes it difficult for humans to develop natural immune response or develop drugs to fight against AIDS virus. However, higher animals develop a general strategy in immune systems that has been very effective most of the time. Instead of developing one kind of antibody, our immune systems produce millions of different types of antibodies. It is highly likely that for any kind of bacteria or viruses, there is a suitable type of antibody to destroy them. This strategy is very effective but very expensive, because our body needs to produce many different kinds of antibodies that are useless most of the time. When we are too young, too old, too weak, or too stressed, our bodies don’t have enough energy to produce large amount of antibodies. That is when we become vulnerable to infection.

Other than adapting to the ever changing environment, organisms also regulate internal and sometimes external environment. In general, higher fixed cost systems spend more resources to regulate environment to reduce the level of uncertainty. Human beings, as warm blooded animals, regulate our body temperature around 37 degrees, regardless of external temperature. The constancy of body temperature greatly reduces the uncertainty for our body’s many chemical and physical processes. However, the maintenance of constant temperature is extremely energy intensive. Warm blooded animals require a lot more food intake than cold blooded animals of the same size. When they cannot find any food for several days, their body functions weaken seriously. This causes great increase in uncertainty.

Living systems not only regulate internal environment, they also regulate external environment. Among all animals, human beings have the greatest capacity to modify our external environment. We clear land for agriculture, which provides much more certain supply of foods than hunting and gathering. We build houses to shelter us from the uncertainty brought by wind, rain and snow. In wealthy societies, we take many measures to reduce uncertainty in many aspects of life (Galbraith, 1958). When we lose jobs, we get unemployment payment. When we are disabled, we get disability assistance. Our buildings are cooled in summered and heated in winter. Water is chlorinated to eliminate bacteria. Populations are vaccinated to reduce infectious diseases. Many measures are taken to make life as predictable and as pleasant as possible. In general, uncertainty in wealthy countries is very low compared with poor countries. In environment with low uncertainty, investment with high fixed cost and long duration often generate high rate of return. But measures to reduce uncertainty require resources. For example, to keep buildings warm in winter, we need natural gas to heat the building.

It is more difficult for high fixed cost systems to change. But high fixed cost organisms do develop mechanisms to change periodically to adapt to changing environment. Most complex animals reproduce sexually. Sexual reproduction is very expensive compared with asexual reproduction. Just imagine how much effort we spend on relationship and how difficult relationship can become. However, by reproduce sexually, the genes of our offspring are reshuffled substantially. The genetic diversity of our offspring makes some of them more adaptable to the changing environment. The genetic changes also make parasites in our bodies less able to get established on our offspring. Democratic countries also have elections periodically to change the leaders of the systems.

In general, mature industries with low levels of uncertainty, such as household supplies, are dominated by large established companies such as P&G. Fast changing industries, such as IT, are pioneered by small and new firms. Microsoft, Apple, Yahoo, Google, Facebook and many other successful businesses are started by one or two individuals and not by established firms with capital and expertise. Similarly, in scientific research, mature areas are generally dominated by top researchers from elite schools, while fundamental ideas are often initiated by newcomers or outsiders. Most economists are trained in economics. But some of the most innovative thinkers in economics are outsiders. Neoclassical economics, the dominant economic theory today, was founded around 1870 by Jevons, Walras and others. Both Jevons and Walras, the greatest economists in the nineteenth century, were trained in science.

We will compare the properties of RNA and DNA molecules to discuss the relation between fixed cost, uncertainty and duration. DNA is made from RNA. So DNA is more costly to make than RNA. The extra step in chemical reaction makes DNA more stable than RNA. So a DNA molecule lasts longer than a RNA molecule. Because DNA is more stable than RNA, DNA is the preferred molecule to carry genetic information. The only organisms to use RNA to carry genetic information are some viruses, which are very small and mutate very fast. For example, AIDS viruses use RNA to carry genetic codes. Any larger organisms, from bacteria on, use DNA to carry genetic codes. But RNA molecules are used in our bodies for many functions that do not require high level of precision and do not last very long, for RNA molecules are cheaper to make than DNA. The chemical properties and biological functions of DNA and RNA show that fixed cost, uncertainty and duration are intimately correlated.

* 1. **Discount rate**

From monetary policies to the climate change problem, from the burden of private credit card debts to the evaluation of public projects, discount rate is the central issue. Much effort has been made to understand the tradeoffs among investment and consumption behaviors at different time intervals, yet there is little clear understanding about the nature of discounting. Martin Weitzman, the world’s leading expert on the social discount rate, commented:

The concept of a “discount rate” is central to economic analysis … Because of this centrality, the choice of an appropriate discount rate is one of the most critical problems in all of economics. And yet, to be perfectly honest, a great deal of uncertainty beclouds this very issue. … The most critical single problem with discounting future benefits and cost is that no consensus now exists, or for that matter has ever existed, about what actual rate of interest to use. (Weitzman, 2001, p. 260)

The main problem in the theory of discounting, as pointed out by Weitzman, is that “an economist who knows the literature well” is “able to justify *any* reasonable social discount rate by some internally consistent story”. We will discuss how the discount rate is related to other factors in biological and social production systems, such as fixed cost in production, duration of production or life span, and uncertainty. The relations among different factors in a production system will put constraints on the ranges of discount rate that are viable in particular environments. These constraints will help us understand how discounting should be applied in different situations. But before our analysis, we would like to list some puzzles related to discount rates.

The borrowing rates for banks are very low. But the credit card interest rates that banks can charge their customers are very high. How can the large interest rate differential be maintained over a long time? From another perspective, individuals can obtain a line of credit at a much lower interest rate than the credit card interest rate. Why do so many people still maintain a large amount of credit card debt, without replacing it with a line of credit?

Most economists are staunch proponents of the efficiency of markets, especially when the market is very liquid and transparent. The short term money market is among the most liquid and transparent markets in the world. Yet, most economists support that the short term discount rate, possibly the most important factor affecting economic performance, should be determined by a small group of “independent” professionals from central banks. Does that mean the market is only capable of being efficient on minor issues and not on major issues?

In general, yield curves slope upward. Loans with longer maturity pay higher interest rates than loans with shorter maturity. At the same time, empirical evidence suggests that humans discount the long term future at lower rates than the short term future (Ainslie, 1992; Berns, Laibson and Loewenstein, 2007). Many policy papers also advocate discounting long term projects at lower rates than short term projects (Weitzman, 2001; Newell and Pizer, 2003**).** Why do market discount rates and psychological and policy discount rates move in different directions in relation to the increase in project duration?

In the following, we will discuss how discount rate is related to other main factors in production. This will help resolve the puzzles. First, discount rates are closely related to fixed cost in production. When we have invested a large sum on something, we will take extra care that the value of investment depreciates slowly. From another perspective, in a low interest rate environment, the cost of borrowing is low. Investments with higher fixed cost will benefit. Investments with lower fixed costs are less sensitive to the level of discount rates. Human beings instinctly understand the relation between fixed cost and discount rate. Many psychological experiments show that the “rate of temporal discounting decreases with the amount of reward” (Thaler, 1981; Green, Myerson, McFadden, 1997) In the field of human psychology, this empirical regularity is called the “magnitude effect” (small outcomes are discounted more than large ones)

An earlier work by Ainslie and Herrnstein (1981) provides similar understanding:

The biological value of a low discount rate is limited by its requiring the organism to detect which one of all the events occurring over a preceding period of hours or days led to a particular reinforcer. As the discounting rate falls, the informational load increases. Without substantial discounting, a reinforcer would act with nearly full force not only on the behaviors that immediately preceded it, but also on those that had been emitted in past hours or days. The task of factoring out which behaviors had actually led to reward could exceed the information processing capacity of a species.

Differences in fixed costs in child bearing between women and men would affect the differences in discount rates between them. Women spend much more effort in child bearing. The high fixed investment women put in child bearing would make women’s discount rate lower than men’s. An informal survey conducted in a classroom survey showed that discount rates of the female students are lower than that of the male students.

In poor countries, lending rates are very high; in wealthy countries, lending rates charged by regular financial institutions, other than unsecured personal loans, such as credit card debts, are generally very low. To maintain a low level of lending rates, many credit and legal agencies are needed to inform and enforce, which is very costly. As wealthy countries are of high fixed cost, they are willing to put up the high cost of credit and legal agencies because the efficiency gain from lower lending rate is higher in high fixed cost systems. In the last several hundred years, there has been in general an upward trend in living standard worldwide. There has also been a downward trend in interest rates (Newell and Pizer, 2003).

Second, discount rates are closely related to the duration of production or lifespan of organisms. In steady states, the rate of reproduction is equal to the rate of death. Therefore, in biological literature, the discount rate is often set to be equal to the rate of reproduction (Stearns, 1992). Bacteria can reproduce themselves in just thirty minutes. Humans can reproduce only after ten or more years old. Hence, discount rates for bacteria are measured in hours while discount rates for humans are measured in years. In general, organisms with longer life spans have lower fertility rates and hence they have lower discount rates (Lane, 2002). Many empirical studies have documented that humans, as well as other animals, often discount long duration events at lower rates than short duration events (Frederick, Loewenstein and O’Donoghue, 2004). This pattern is called hyperbolic discounting.

Third, discount rates are closely related to uncertainty. In general, organisms facing a high level of environmental uncertainty, such as predation, have higher discount rates (Stearns, 1992). In economic theory, the standard quantitative model on the relation between discounting and risk is the capital asset pricing model (CAPM), which states that systems with higher risk is discounted at higher rate. The same idea about the relationship between discount rate and uncertainty was reached in the study of human psychology. “The same discount curve that is optimally steep for an organism’s intelligence in a poorly predictable environment will make him unnecessarily shortsighted in a more predictable environment (Ainslie, 1992, p. 86)”.

The above discussion shows that discount rates, fixed costs, duration of production and uncertainty are highly related. Lower discount rates are positively correlated with high fixed costs, long duration, and low uncertainty. High discount rates are positively correlated with low fixed cost, short duration, and high uncertainty. These results are consistent with the classification of organisms in ecological studies. Organisms are often classified as K strategists and *r* strategists. K strategists often have high fixed costs, long duration and low discount rates. They are highly competitive in stable environment. *r* strategists often have low fixed costs, short durations and high discount rates. They thrive in volatile environment (MacArthur and Wilson, 1967). Our results about the relation between discount rate and other factors are very similar to ones that are obtained in the study of mind of humans and other animals (Ainslie, 1992). This shows that the human and animal mind, a product of evolution, understands the important relations in life very well.

In most of our evolutionary past, uncertainty was high, lifespan was short, fixed investment was low. So our discount rate is high. Many of us are willing to borrow at high interest rate. They feel that high interest rate credit card debt is normal. So they accumulate credit card debts casually and don’t put much effort in looking for lower rate alternatives, such as line of credit. The accumulation of debts often has tragic consequences to individuals, their families and the society as a whole. This is probably why in many cultures and religions, money lending businesses are discouraged or banned. However, modern societies have developed technologies to extract abundant fossil fuels and other resources. With abundant resources and hence high economic output, high fixed cost investments generate high rate of returns. Low interest rate environment will stimulate high fixed investment. But interest rates will be high in a free financial market because the discount rate in our mind is high. In many countries, governments take over the task of setting short term interest rate to keep interest rate low. By maintaining low interest rates, governments help stimulate high fixed cost investment and overall economic growth. This is why short term interest rate is determined by governments and not markets. For mainstream economists, free market provides the optimal result. Any government action requires explanation from externality or imperfection, which are often arbitrary. But once we recognize the relations among major factors in biological and social systems, it is easy to understand why governments, not markets, are setting short term interest rates in many countries.

There is another puzzle related to interest rate. Why consumers are allowed to accumulate large amount of credit card debts, which charge very high interest rate? It is often argued that free market provides optimal results. Hence the issue of credit card borrowing should be left to individuals. If this is so, why legislations force people to save for retirement funds, which usually generate very low rate of return? If people could use pension money to pay off credit card debt, they will be much better off. At a theoretical level, why different logics are used to explain different things? But from the perspective of returns for financial institutions, it is easy to understand. Financial institutions generate huge revenues from both credit card businesses and pension fund management. Naturally, they will support both types of businesses. Logical inconsistency is unimportant compared with profit. This is another reason why a return based theory provides more explanatory power than other approaches.

* 1. **The volume of output, market size and abundance of resources**

The increase of market size and volume of output often increases profit or reduces average cost. However, average costs do not always decline with increasing outputs. Both economy of scale and increasing marginal costs play a role in overall costs. In educational system, primary schools are usually smaller than secondary schools, which in turn are smaller than universities. It would be difficult for primary school students to walk for a long distance to go to school. So a primary school caters to a very small market. Secondary school students can walk for longer distance but usually live with their parents. So a secondary school caters to mostly local students. University students can live independently, often away from home. So a university caters to local and distant students. As a result, the market size of a university can be larger. Market size of an organization is closely related to its fixed cost. A primary school only needs one teacher for each grade. A secondary school needs specialized teachers for English, mathematics, science, physical educations and other subjects. A university usually has many specialized programs. Each program often has several specialized fields that need different faculty members. For example, a business school generally has department of accounting, finance, marketing and other programs. With increasing market sizes from primary school to university, their corresponding fixed costs increase as well.

Many factors determine the market size of a product or the volume of output of a business. An important factor is the level of resource abundance. When resource is abundant, more resources can be transformed into products people need and transportation cost is low. People will buy a lot of different things in large quantity. When many large oil fields were discovered in late 20s and early 30s in the last century, oil prices dropped, many new cars were built and many new roads were paved. Many shopping malls and supermarkets with ample free parking spaces were built. They provide large variety of merchandises and serve customers within driving distances. Gradually they replace general stores, which provide small quantities of essential goods and serve customers within walking distance, as the main destinations for shopping. Low transportation cost also greatly increases the size of international trade, making the global economic activities more integrated. Most economists believe economic activities will become more integrated in the future. However, the productions of many crucial resources, such as oil, are near their peak or already peaked. So it is highly likely that trend of globalization and increasing market size may reverse soon. Historically, free trade zones have been growing and declining alternatively. Volume of economic output has been growing and declining over different time as well. Some people have already prepared and practiced various forms of localized economy, such as growing and eating local foods.

The market size for commercial goods has been expanding continuously for the last several centuries. The same is true for the market sizes of languages. Because of the decreasing cost of transportation and communication, we tend to interact with more and different people. The benefit of using major languages, especially English, increases. Many languages with small number of speakers have disappeared or are disappearing. However, the market size for the states, the most important social units, has been in decline for some time. At least since the end of the World War II, the number of countries has been growing steadily. This means the average market size of the states, measured by area, has been declining since then. Soon after the end of the World War II, many Asian and African regions became independent countries. In the nineties of the last century, Soviet Union and Yugoslavia split into several different countries. From time to time, some countries, such as Czechoslovakia, Sudan and Ethiopia, split into two countries. The expansion of free market and free trade often intensify the social division, making societies less stable (Chua, 2003). It is likely that the number of countries will continue to increase in the future.

Market size is closely linked to fixed cost of a system. When fixed costs are low, an area may contain many independent units of production. But with higher fixed cost and increasing division of labor, less number of integrated systems can live in the same area. For example, a single grizzly bear, which is composed of trillion cells, require a large piece of land to survive. On the same piece of land, trillions of single celled organisms live there. The grizzly bear is very powerful, much more powerful than any single celled bacteria. But when any organ of the grizzly bear becomes weak, most of the trillion cells suffer as well. When the kidney cannot filter out waste effectively, all the trillion cells on the body of the grizzly bear have to live in a toxic environment. When the heart fails, all the trillion cells on the body of the grizzly bear die together. On the other hand, even many single celled bacteria die constantly, other bacteria can survive independently and the community of bacteria can prosper.

There is a parallel in human societies. In a simple society, families or tribes are often the independent units of production. While each family or tribe struggle mightily to stay afloat, the lives of different families or tribes often are not highly integrated. But in a high fixed cost system with extensive division of labor, most members of the society are integrated into the system. Soviet Union was once a splendid highly integrated system. It did so well in its early days that many people were fearful of its dominance in the near future. But a high fixed cost system is difficult to adapt. In a high fixed cost system, the deficiency of a small part will crumple the whole system. Over time, such deficiency will surely develop and spread, ultimately destroying the whole system. While the capitalist system is supposed to operate on private ownership, everyone and every business is obliged to pay tax. It means that everyone and every business is partly owned by the government. When the tax rate is high, the percentage of government ownership is high and our social system becomes highly integrated. In such systems, the freedom and right of each member is much less than what we would like to believe. For example, during the recent financial crisis, wealthy financial institutions are bailed out with taxpayers’ resources, while ordinary people have no say in the decision making process. When the collective resources are forced to support the financial industry whenever the collapse of financial system seems imminent, the whole society becomes a single large market size, high fixed cost system, whether we like it or not. If everyone is forced to live in a single large building, the collapse of the building will bury us all.

Some biological processes become easy to understand from the sizes of output. There are twenty types of amino acids. Glutamine is often the first amino acid to be synthesized. But glutamine is sometimes converted to arginine before amino acids are transported to somewhere else (Willey, Sherwood, Woolverton, 2011, p. 703). It may be difficult to understand why organisms take extra step in performing a task. But each arginine molecule contains four nitrogen atoms, the most nitrogen atoms for any amino acids. So transporting one arginine molecule carries four times as many nitrogen atoms as glutamine. Similarly, a hemoglobin molecule can carry four oxygen molecules while a myoglobin molecule only carries one oxygen molecule. Hemoglobin molecules transport oxygen over long distance and myoglobin molecules deliver oxygen to nearby cells. By carrying four oxygen molecules, hemoglobin reduces the cost of oxygen transportation. By carrying one oxygen molecule, myoglobin supplies the oxygen need for nearby cells.

* 1. **On demand and supply**

Demand and supply are two basic concepts in economics. Biologists and ecologists recognize that conditions of an ecosystem are ultimately determined by the supply of resources, such as size of land, amount of water, sunlight, carbon dioxide and other nutrients. However, most economists often blame the lack of demand as the cause of poor economic conditions. They often prescribe policies to stimulate demand in times of economic downturn, while at the same time lamenting the high debts accumulated by the general public. Why people from different disciplines have different emphasis on demand and supply? We will apply the concepts discussed in earlier sections to analyze their arguments.

An economic system requires fixed cost investment. A system with higher fixed cost generally needs higher output to breakeven. At the same time, a higher fixed cost system, with lower variable cost, has the possibility of generating higher rate of return when the output is high. In a mostly growing economy, anticipation of larger market size encourages higher fixed cost investment. However, the initial market size may not be able to support the high fixed cost investment. If the expected market growth doesn’t materialize, the investment will not be able to breakeven and face the risk of collapse. So it is essential to maintain high level of demand to keep high fixed cost investment profitable. Very often, the demand is maintained through borrowing. Since the economy is on the upward trend, borrowing can generally be repaid in the next cycle of economic boom. In a rising economy, stimulating demand will encourage high fixed cost investment, which provides high rate of return in a large market. Since modern economic theory is an adaptive product in a growing economy, most economists emphasize the demand side of the equation.

Biologist and ecologists usually consider more general patterns over long terms. From time to time, some biological systems find new ways to obtain resources to expand their supply. When resources are abundant, organisms multiply rapidly to consume more resources. Over long term, organism will occupy all ecological niches. In the end, the growth of any biological system is constrained by resource supply. Because of technology developments to utilize fossil fuels and other resources in great scale, economic growth and population growth have been the norm for the last several centuries. However, in most wealthy countries, where per capita resource consumption is high, fertility rates have already dropped below the replacement rate. This suggests that supply of resources has already become the main constraint of the economic activities, at least in high resource consumption societies.

Politically, it is more attractive to promote the increase of demand. However, many households and governments are already heavily indebted. The amount of debt has been increasing over time. The problem with current economic system is not the lack of demand. Rather the problem lies at the high fixed cost of our economic system in wealthy countries, which requires high level of demand to breakeven. If the fixed cost of our social system becomes lower, we can generate positive return and make the system viable with lower level of demand. However, with the continuous economic growth of several centuries, we are getting used to the steady increase of spending, or demand. It is very difficult for us to accept the reality that continuous growth is never the normal state in any biological system, including human societies.

* 1. **Concluding remarks**

According to established economic theories, human beings are assumed to maximize utility. Since utility is subjective, maximizing utility can be many things to many people. The established economic theories also assume that market is the most efficient way to organize economic activities. But due to “imperfection” or “externality”, regulations are needed from time to time “to make the market work”. Education should be promoted by the governments, because education generates positive “externality”. Short term interest rates are set by the central banks, because markets can be “myopic”. These are rationalizations, not explanations. They don’t tell us how much government should spend on education or what levels interest rates should be at.

In this chapter, we propose that the rate of return as the common measure of all biological and social systems and discuss the major factors that affect the rate of returns. Currently, most high resource consumption societies have negative biological rate of return, rendering these systems unsustainable. For these systems, the reduction of resource consumption will make these systems viable again. Specifically, for these systems, fixed costs shall be lower, duration of investment shall be shorter, less resource shall be spent to reduce uncertainty, discount rate shall be higher and market size shall be smaller. In practice, reducing the tax rate, which will reduce government revenue, will achieve most of these goals. With lower government revenues, the sizes of governments, as the highest fixed cost systems, will be reduced; number of years of mandatory education will be reduced; average retirement age will be earlier; entitlement programs will be scaled back; governments and central banks will lower ability to influence interest rate and financial markets; supranational organizations, such as the Eurozone, will be abolished or scaled down, making each individual state more flexible in dealing with their own problems.

Among the major factors in social systems, the change of fixed cost often leads to corresponding changes in other factors. However it is often difficult to reduce fixed cost in an economic or social system, for several reasons. First, higher levels of fixed cost are widely associated with progress. Established economic theories generally emphasize the advantages of division of labor and of specialization, which tend to increase fixed costs in economic systems. The advantages of higher fixed cost systems are certainly real in an expanding environment while resources are cheap and abundant. However, over a longer time horizon, resource constraint is often more pronounced. Jesus once said, “Blessed are the meek, for they shall inherit the earth.” This means low fixed cost systems often have advantages over long term.

Second, changing the structure of an economic system is very costly and disruptive. It is something not likely to be undertaken while there is doubt about the long-range outlook. If the decline or difficulties are thought to be short term, it is often better off to reduce output while maintaining the essential production system intact; only when the decline is known to be long term does incurring the cost of change make sense. For this reason, strategies of denial about the long-term character of change in the resource environment argue powerfully against making changes in the structure of fixed costs.

Third, reducing fixed cost often involves cutting the number and the incomes of people at high income levels. For example, the trial of Robert Pickton, a serial killer from the Vancouver area, costs the government over one hundred million dollars. Most of the money goes to highly paid professionals, such as lawyers. If the tax rate is lowered, tax income will be reduced. Governments won’t be able to support as many highly paid professionals, who are politically influential. This is obviously difficult.

The reduction of fixed cost, with the corresponding increase of variable cost, will make our daily life less convenient. The reduction of resource consumption will reduce our living standard. Hence it is often difficult to adopt policies that will reduce the fixed cost of a society. So far, it is mainly through the natural selection instead of adaptation that low fixed cost communities spread out. In biological systems, low fixed cost systems have higher fertility than high fixed cost systems. The same is true in human societies (Rushton, 1996). Low fixed cost communities generally have higher fertility rates than high fixed cost communities. With currently low mortality rates among both high fixed cost and low fixed cost communities, low fixed cost communities spread out with respect to high fixed cost communities. Both selection and adaptation are at work in the evolution of human societies, like any other biological systems (Jablonka and Lamb, 2006; Cochran and Harpending, 2009).

In general, human beings, as the most successful evolutionary product on the earth, know what will benefit us. If a policy has both short term and long term benefits for most of us, it is unlikely that it has not been implemented. If a policy with significant long term benefit to the majority of people is not adopted, it is almost certain that the policy has negative short term impacts to many of us. Most of us understand the long term harm caused by drug use. But many people choose to use drugs because of the short term pleasure they bring. Short term benefits of many social policies are reaped by members of the current generation who can influence decision making, while long term costs are paid by youth, children and future generations, who have little political influence. This makes it difficult to abandon such policies. Inevitably most of the policy recommendations derived from this economic theory will be politically incorrect or politically unpopular.

The established economic theories often assume technology development will resolve the problem of resource scarcity. New technologies often enable us to utilize resources that previously cannot be tapped. At the same time, technology itself requires resource input. The net output of resources from new technologies may or may not be positive. In the next chapter, we will discuss in greater detail on the relation between resources and technologies.

**Chapter Two: Resource and Technology**

**2.1. The importance of natural resources**

The standard economic theory states that natural resources are only one factor in economic activities, which can be easily substituted by other factors, such as technology. The depletion of natural resources is of little concern for technology advance will substitute the need for resources. Since the standard economic theory is taught in almost all universities worldwide, it has a great influence on public opinion.

To understand better the relation between resource and technology, we will examine regions where resource base is very narrow so the impact of resource depletion can be measured more clearly. Sunlight is the most universal resource to the world. Many other resources, such as fresh water, fertile land, are derived from abundant sunlight. We will examine population change in a mining town in deep North, where solar energy, the most important and universal natural resource, is scarce. Suppose in one mining town, there are 10,000 residents, of which 2,000 are miners or mining related service providers. The rest 80% population are policemen, teachers, doctors, pastors, bakers, grocery store cashiers, other service providers and anti-mining environmental activists. Can we conclude, from this figure, that only 20% of the town depends on the mining activities? Suppose, after some years, the mine is exhausted. Can new technologies provide additional job opportunities or at least support the remaining 80% non-mining population? Some historical examples will offer a hint. Barkerville, in northern British Columbia, Canada, which at its heyday of gold rush had over 10,000 residents, was once the largest town along the west coast of North America. Its population dwindled to zero when the gold mines were exhausted. This is not an isolated example. In almost all mining towns in deep North, once mines are exhausted, the towns become ghost towns.

When resource bases are narrow, it is easy to recognize natural resource as the ultimate source of all economic activities and technologies are means to utilize resources. But often there are many different kinds of natural resources in the same place. People will move on to other natural resources after the depletion of one kind of natural resource. For example, with the depletion of gold mines in California after the gold rush, people move to agriculture and other activities that require different kinds of resources, such as fertile soil, fresh water, sunlight, and petroleum.

With the increase of commodity prices, more and more people become aware of the importance of natural resources. However, the total costs on gasoline and other commodities are still small for most people. Output from resource industry is still only a small part of overall economic activities, even for major commodity producers, such as Canada. This gives an impression that resources constitute only a small part of the overall economic activities. But this is really a matter of definition. Transportation industry is not defined as a resource industry. However the manufacturing and operating of vehicles, ships and planes are totally dependent on the availability of energy resources.

To better understand the extent of our dependence on natural resources, we will discuss the diverse forms of natural resources and a unified understanding of resources from the thermodynamic theory.

**2.2. Natural resources: Diverse forms and unifying principle**

Most of the natural resources on the earth can be attributed to the temperature differential between the sun and the earth. The surface temperature of the sun is 6000 K while the surface temperature of the earth is around 300K. The high temperature solar surface emits sunlight, which carries high quality energy. The earth receives high quality solar energy and emits low quality waste as infrared wave. This temperature differential is what drives most thing, including living organisms, on the earth. Intuitively, this temperature differential is like differential in water levels at a hydro dam, which drives turbine to produce electricity.

Part of solar energy is captured by plants through photosynthesis into chemical energy, which can be stored for a longer period of time than photons. The chemical energy stored in plants can be released to work for plants when and where it is needed to maintain various life activities of the plants, including photosynthesis process. Animals, by eating plants, obtain some of the chemical energy stored in plants. Almost all of the energy sources in the food web on the earth ultimately come from the solar energy.

Fresh water is so common that we often take it for granted. In many places, its economic value is very low. However, fresh water is vital for our survival. Fresh water is also very scarce compared with salty water. It comprises of only 3% of all water. The rest 97% is salty water. Salty water is of lower free energy level than fresh water. Hence salty water is in the stable state of chemical equilibrium, while fresh water is in the unstable non-equilibrium state. This non-equilibrium state is maintained by the solar energy. Solar energy distills mostly salty water into vapor, which returns fresh water to the earth in the form of rain or snow. Since fresh water is of higher free energy, the intake of fresh water instead of salty water saves living organisms a lot of energy. That is why areas with abundant fresh water have higher level of biomass density than areas with little fresh water, such as desert or ocean, where water is salty. Most cities are by rivers or lakes, where freshwater is abundant. Fresh water is another gift from the sun.

Because of gravity, water will flow from high to low places. Without solar energy to enable water vapor to escape the pull of gravity, all water will stay at low places. Clouds will not form and rain will not drop. The whole earth will be a desert. No river will flow on the earth. Rivers are vital to human life. Most early civilizations were originated by the riverside. Rivers formed the major channels of transportation before the age of cheap fossil fuels. With rivers comes the hydro power, which is the only renewable natural resource that generates significant amount of electricity for human society.

Today, most of the energy needs of human society come from fossil fuels. We often call our civilization the fossil fuel civilization. Coal, oil and most of the natural gas are from ancient biological deposits. These biological deposits, which were formed over millions of years, are transformed for use by human societies over last several hundred years. The abundant use of the fossil fuels is the foundation of the economic prosperity many the world over have enjoyed in recent times.

More detailed and systematic discussion about natural resources can be found from the standard references, such as Hall et al (1986) and **Ricklefs (2001).** While the forms of natural resources are diverse, most natural resources can be classified into two classes. The first class is low entropy sources, or more popularly understood as energy sources. The second class is raw materials that can be used as building blocks to harness energy sources for our use. The second class includes most metals. Many resources belong to both resources. Wood can be used as cooking fuels or building material. Petroleum can be processed into vehicle fuels or vehicle parts. Proteins, fats and carbohydrates can be used as energy sources or as building blocks for organisms.

**2.3. Living systems: Positive return technologies of utilizing resources**

Resources are low entropy sources, or materials having gradient against the environment. But to utilize resources, it requires structures to harness entropy flows. We will look at the periodic table to see what kinds of chemical elements are good raw materials to build up structures to harness entropy flows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| H |   |   |   |   |   |   |
|  Li | Be | B | C | N | O | F |
| Na | Mg | Al | Si | P | S | Cl |

**Table 2.1**: The first three row of periodic table without the column of noble gases, which are chemically inactive.

Table 2.1 shows the first three row of periodic table without the column of noble gases, which are chemically inactive. Carbon is the first element at the fourth column, the center column of the periodic table. Hence carbon is the lightest of a group of elements that are largely chemically neutral and have four valence bonds. The chemical neutrality and large number of valence bond makes it easy for carbon to link many different atoms to form large molecules. Large molecules are essential for the preservation of life to withstand random dissipation (Schrodinger, 1944). Large molecules are also essential to perform complex and various tasks. This is why carbon, and carbon alone, is the backbone of life (Atkins, 1997). The chemical neutrality of carbon makes it easy to combine with and detach from other atoms. The stable and weak bond of carbon is ideal for storing and releasing energy. This is why carbon is an essential part for any type of organic energy, including natural gas and petroleum. Hydrogen is the smallest and the lightest element. That is why hydrogen is an efficient carrier of energy and building block of a good technology.

For a technology to last, it must be able to utilize energy resources to make another copy of itself before it wears out. Living systems are technologies that last. Carbon and hydrogen atoms are natural raw materials that build living systems. Because the buildup of living systems embodies energy resources, living systems themselves become the resources that can be used by other living systems. Animals eat plants. Animals eat other animals. Bacteria and fungi eat plants and animals. Viruses eat bacteria, plants and animals. Human beings eat plants, animals and fungi. We also use fossil fuels, which are transformed from dead bodies of living systems, in great scales.

The greatest amount of resource can be accessed from the earth is the solar energy. The earliest organisms that developed the structure for photosynthesis were probably bacteria. These bacteria utilize solar energy to build up structures with carbon, hydrogen and other kinds of atoms. Bacteria obtain carbon atoms from carbon dioxide. Initially, bacteria get their hydrogen from hydrogen sulfide (H2S) in photosynthesis. Compared with water (H2O), the chemical bond in hydrogen sulfide is weak. From Table 2.1, both oxygen and sulfur are at the sixth column of the periodic table. Their chemical properties are similar. But sulfur is one row below oxygen. So it takes less effort, or lower fixed cost, to get hydrogen from hydrogen sulfide. “However, when hydrogen is removed from hydrogen sulfide in the interior of a bacterium, the excrement is sulfur. Sulfur, being a solid, does not waft away, so the colony of organisms has to develop a mode of survival based on a gradually accumulating mound of its own sewage.” (Atkins, 1995, p. 22) Eventually, organisms developed techniques to break the strong bond in water to get hydrogen. Oxygen is a byproduct of this technology. Since oxygen is a gas, the pollution does not accumulate locally. It spread out globally. Another advantage of obtaining hydrogen from water is the abundance of water. With water as a raw material in photosynthesis, living systems spread all over the earth.

Oxygen molecules, the waste product from photosynthesis, are highly energetic, or have great chemical gradient against the environment. This energy destroys many early living systems and rusts many materials. But eventually, some organisms evolved the structure to harness the energy from oxygen and managed to reduce the destructive power of the energy from oxygen to tolerable levels with antioxidant technology. Animals, whose active lifestyle require large amount of energy consumption, evolved to take advantage of the abundant atmosphere oxygen. All materials with gradient against the environment have the capacity to destroy. Only after we develop technologies to harness the gradient of these materials and contain their destructive power to manageable levels, these materials become resources.

Since the amount of solar energy to early photosynthesis bacteria was almost infinite, those bacteria multiply quickly. Over time, photosynthesis organisms cover most parts of the world. Photosynthesis is probably the most important technology organisms have ever developed. It transforms the vast amount of light energy, which is difficult to store, into chemical energy, which is easy to store. Since the invention of photosynthesis, almost all living organisms derive their energy source from solar energy, directly or indirectly.

If a technology helps an organism to earn positive return on resource utilization, the technology will be duplicated and spread out. However, not all technologies developed by organisms are able to provide non-negative return consistently. Most genetic mutations or innovations are harmful. Even among those mutations where new species are established, over 99.9% of the species eventually went extinct. Technology and innovation are not a guarantee to the prosperity and safety of any biological species or human societies.

A technology is not only expensive to develop but also expensive to maintain. Some fish live in dark underground caves become blind because functional eyes are expensive to maintain. When eyes are no more useful, their structures degenerated. Similarly, the sense of smell of human beings is highly degenerated compared with our ancestors. Walking upright increases our dependence on vision and reduce our dependence on smelling. As a result, our sense of smell is under less selection pressure and degenerates. A technology will be developed and maintained if its return on investment is positive during that period. Otherwise, the technology or its host will degenerate.

With the multiplication of organisms, they have to compete for the limited resources or to defend themselves to be consumed by other organisms. There are two main strategies, either to form social groups to increase the group power or to make each individual more powerful. Social sciences are usually defined as research areas that are concerned about human beings exclusively. But almost all organisms have their social aspects. Bacteria can live alone. But they often grow much better in groups. Most organisms have similar problems when communicating or coordinating with other organisms. Social biology was developed to study these problems. Very often studying the social behaviors of other species provides valuable insights to understand human behaviors. Some single cell organisms evolved into multi-celled organisms. Multi-celled organisms are larger than single cell organism. An individual multi-cell organism can grow into a collection of many trillion cells. Large animals can often overpower small animals and control more resources. But multi-cell organisms need to resolve the same problems of communication and coordination among different cells inside their own body, very much like different individuals in a society need to communicate and coordinate with each other. There is really no chasm between social and natural sciences.

**2.4 Human technologies and human societies**

Technologies in human societies are generally understood as tools we make. These tools are outside our body. The making of tools is not constrained by the physical and chemical environment of our body. Iron smelting require a temperature over 1000 degrees, which is much higher than our body temperature. Aluminum smelting require strong electric currents that would kill us instantly. However, all biological and human technologies are bound by the same economic principle. If a technology generates positive return in biological system, it will be developed and maintained. If not, it will decline. Since different technologies require different social structures and bring in different amount of resources, technical structures and social structures are closely intertwined with each other.

We study past events to estimate the possible patterns in the future and to guide our actions today. The events that greatly affected the lives of people in the past, such as the Industrial Revolution, the Great Depression in 1920s, oil crisis in 1970s and the recent Great Recession, have especially strong grip on the minds of the public. The interpretations of these events greatly influence the public’s vision about the future and the policies adopted by governments. We will put more efforts to examine these more recent events.

**The use of fire and cooking**

 What is the most important activity that separates human beings from other animals? Different people have different answers.  From the energy perspective, we might say the use of fire and cooking. Cooking kills most of the pathogens in the foods. So our body doesn't have to spend as much energy on our immune systems to destroy these pathogens. Cooking, by breaking down large organic molecules, predigest foods before we eat them. So our body doesn't have to spend as much energy on our digestive systems. The energy saving can be used to nourish other important systems, such as brain. So human beings can supply more energy to brains for more complex thinking. Cooking, by killing most of the pathogens in food, reduces the risk of infectious disease. It enables human beings to live in higher density, which stimulates the need for better communication. Languages, cultures and religions flourished to bind people together. Cooking makes us familiar with a chemical process of high energy intensity. This will be crucial in the development of future technologies. It seems no other activity has the same level of impact as cooking and the use of fire to influence the evolution of human beings and human societies.

**High resource density, sedentary lifestyle and agriculture**

Most researchers suggest that sedentary lifestyle followed the advent of agriculture. But sedentary life could precede agriculture. Some places have very high concentration of resource densities. Several times we travelled along Highway 16 toward Prince Rupert, British Columbia. We would pass by a canyon at a place called Morristown. In salmon spawning season, large amount of salmon fish waited just below the canyon, preparing to jump over the rapids to reach their final spawning destinations. Anyone can use a net to scoop up a fish from the river. The salmon fish are so abundant that local people live a sedentary life around the canyon year around. They depend on salmon instead of agriculture. Some people could live sedentary lifestyles before the development of agriculture. But in general, sedentary lifestyle provided the incentive to practice agriculture, which requires intensive work in the fields for prolonged periods. So it could be sedentary lifestyle leading to agriculture, instead of agriculture leading to sedentary lifestyle.

While sedentary lifestyle may precede agriculture, large scale sedentary lifestyle followed the invention of agriculture. Agriculture is the growing of a few selected crops with active exclusion of most competing plants.  Because crops are mainly selected from the nutrition value to humans instead of their competitiveness in the fields, they are usually not very competitive. The growth of crops requires intensive human intervention to remove weeds growing in the same fields. Because of the high resource density of the crops, they become the favored food source for microbes, other animals and other people. So farmers have to remain vigilant against other animals and other people. The increase of resource density always increases the potential of wars against microbes, other animals and other people. Since agriculture produces higher energy yield than hunting and gathering, it can support higher human density. Gradually, farming communities replaced hunting and gathering communities in many places of the world.  High human density comes with complex society with many hierarchies.

**Bronze Age, Iron Age and the Dark Age**

Making bronze tools require special smelting technology. The advent of Bronze Age is a significant step in the mastery of using energy resources. The melting point of bronze is 232°C while the melting point of iron is 1535°C. It is much more difficult to smelt iron than to smelt bronze. So the technology of making iron is developed much later than the technology of making bronze. But iron is much harder than bronze. Iron made weapons are more powerful than bronze made weapons. Iron made tools can be used to cut down trees, from which charcoal is made. Charcoal is used to smelt iron. More charcoal led to more iron and more iron led to more charcoal. This positive feedback greatly increased the output of energy and iron made equipment, such as sword and plow. Swords enabled iron making people to expand their territories. Plows enabled iron making people to get more nutrients from deep soil, thus enhancing crop yields and increasing population density. The arrival of Iron Age generated burst of military, economic and cultural activities in human history. However, the amount of iron production was ultimately constrained by the amount of available trees around iron mines. Sharp increase in iron production in a place quickly deforested the surrounding area, which limited the scale of iron output in most of the Iron Age. Indeed, deforestation and soil erosion often turn once prosperous civilizations into desolate areas. Dark Age, which consumes less resource, often followed Iron Age.

**Coal, iron and the Industrial Revolution**

Industrial revolution has been interpreted in many ways. Here we offer another interpretation, not necessarily inconsistent with other interpretations. Shortly before 1750, a technology of iron making with coal was invented in England. Mining and transportation of coal require many iron made equipments. Since the beginning of the Iron Age, iron smelting was by charcoal, which is made from wood. Before 1750, the need for charcoal in iron making largely deforested whole England (Jevons, 1865). The limited supply of charcoal limited the supply of iron, which limited the supply of coal. After the invention of iron smelting by coal, coal replaced charcoal as the main fuel in iron making. Coal is much more abundant than wood. More coal led to more iron and more iron led to more coal. The positive feedback between the output of coal and iron, the most important energy source and the most important material, enable human beings to grow tremendously in number and in prosperity. This is the essence of the Industrial Revolution. Jevons was very aware of the importance of this invention. In *The Coal Question*, he described it in great detail.

**The age of oil: the Great Depression, oil crises, and the Great Recession**

Since the beginning of the Industrial Revolution, global economy has been growing steadily most of the time. But there had been several slowdowns, such as the Great Depression, the oil crisis in 1970s and the recent Great Recession. The Great Depression was often attributed to structural weaknesses in various economic sectors and policy mistakes by various government agencies. From the energy and technology perspective, the transformation from a coal-based economy, centered on railways, to an oil-based economy, centered on cars, was partly responsible for the Great Depression. In the decade of 1920, the number of cars increased tremendously. But the supply of oil had been quite limited. So cars were regarded as a supplement but not replacement to the railroad economy. However around the end of 1920s and the early 1930s, many gigantic oilfields were discovered over a short period of time due to the development of better exploration methods (Deffeyes, 2001). It became very clear that petroleum and car economy will replace the coal and railroad economy. The railroad economy fell apart immediately because of this realization. But it took time for the car economy to grow enough to take the place of railroad economy. The structures of railway centered economy are very different from that of the highway centered economy. Areas around the train station are often the prime real estates and the center of most economic activities in a city in railway centered economy. In many cities, the street where the train station locates is called the first avenue. However, in a highway centered economy, shopping areas are relocated to malls and residential areas are moved to suburbs. The shift of economic gravity often devastated the downtown areas in most cities. Many once prosperous towns on railways became ghost towns when highways bypassed them.

Since the discovery of the giant oil fields around 1930 happened over a short period of time, the adjustment was very sudden and painful. The Great Depression was unavoidable, regardless of the government policies. Yet the abundance of oil, an energy source of higher quality than the coal, also set the stage of economic boom after the Second World War, when highways and gas stations were built in many parts of the world. The tremendous growth of petroleum and car based economy greatly increased the consumption of petroleum. In 1950s, Hubbert proposed oil output will eventually peak and decline. Around 1970, many people became concerned by the increasing consumption of resources. A representative work at that time was The Limit of Growth published in 1972. In 1970, US oil production peaked. In 1971, US government stopped converting US dollar to gold at the fixed rate of 35 dollar per ounce, thus delinking the value of dollar to gold, a major commodity. After that the value of US dollar depreciated sharply against gold. This put heavy pressure on the price of other commodities, such as petroleum (Galbraith, 2008). In 1973, the oil price increased substantially due to the collective action of major oil exporting countries. This increase of oil prices generated deep recession in many oil importing countries, including most wealthy countries.

By 1980s, most wealthy countries that experience economic recession regained economic growth. The standard explanation is that these countries were able to overcome high oil price with proper economic policies. Hence good economic policy can overcome resource scarcity. However, if we understand human society as a biological system, we will observe that in most wealthy countries, where resource consumption is high, fertility rates dropped below replacement rate after 1973, the year of high oil price. This shows that the biological rate of return has turned negative. However, the drop of fertility rate temporarily reduced the investment cost of raising the next generation. This generates temporary economic growth for several decades. But eventually, negative biological return will inevitably lead to negative monetary return in a society.

Another adjustment from wealthy countries was to move manufacturing activities to poor countries where ordinary people have little political powers. This greatly reduced the energy consumption in dealing with waste pollution. This also greatly reduced the salary paid to workers. Lower salary means less consumption of resources. By moving manufacturing to poor countries, the energy consumption is greatly reduced.

In 1970s, the global oil output was still rapidly growing and the average physical cost of oil production was low. But the control of price setting of oil gradually shifted from major oil consuming countries to oil producing countries. This caused major economic recessions and negative biological return in most wealthy countries. By the end of 1990s, with the rapid depletion of easy to extract oil, the average cost of oil output increase steadily. In a now classic paper titled The End of Cheap Oil, Campbell and Laherrere (1998), after carefully examining the oil exploration and production data, concluded “What our society does face, and soon, is the end of the abundant and cheap oil on which all industrial nations depend.”

When their paper was published in 1998, oil prices were in the low teens. Since then, the price of oil, as well as other major commodities, has increased substantially. If we recognize the fundamental importance of resources to the overall economy, we would have stopped the measures to stimulate economy after the burst of the internet bubble in 2000. However, the authorities were convinced that the economic recessions in 1973 was caused by improper policy response to high oil price instead of high oil price itself. They believed that they have mastered the proper policy response now and were not worried about the steady increase of oil prices. That is why the authorities were unprepared when the financial crisis broke out in 2007 and 2008, although the prices of major commodities had been advancing for some time.

From the established economic theories, recessions are short term interruptions from long term economic growth. After each recession, economic growth will eventually resume. But from a biological and resource perspective, long term economic growth is not assured. It has been for several decades that the biological rate of return has been negative in most wealthy countries. With the demographic structure of inverse pyramid, economic activity will eventually decline. High resource cost makes it more difficult to maintain both high living standard and non negative biological return.

The resource and technology based interpretation provides a simple and consistent interpretation of the major events inhuman history. There are many other interpretations. Financial crises are often blamed on human greed. This is certainly true. But humans are always greed, before and after financial crises. Financial crises are often blamed on bad monetary policies, keeping interest rate too low for too long. But in a system with abundant resources, low interest rate policy only has small impact to generate inflation. Financial crises are often blamed on wide spread fraud. This is certainly the case in the recent Great Recession. But why frauds are so systematically practiced this time? This is because in an environment with increasing resource cost, fraud becomes the only viable way to generate high rate of return systematically, still expected by the public who are accustomed to the good old days of abundant resources (Chen and Galbraith, 2012b).

**2.5. On inequality**

When the water levels inside and outside of a hydro dam are unequal, electricity can be generated. When temperatures inside and outside of an engine are unequal, work can be generated. In popular term, the inequality of gravitational potential, chemical potential, electric potential and other potentials are called energy. From the thermodynamic theory, inequality in potential, or energy is the driving force in the nature. It is also the destructive force. We all need energy provided by oxygen. At the same time, our body produces many antioxidants to prevent oxygen from reacting and destructing our tissues. Sugar is the vital energy our body needs. But too much sugar in our blood system and in cells will damage our health. When we cannot maintain a low level of sugar in our blood system, we get diabetes. Human societies depend very much on high energy input. But we carefully regulate the energy sources. “Playing with fire” is always considered dangerous. Whenever possible, systems with high gradient, or high inequality are carefully regulated in isolated places or remote places. Furnaces are usually located in basements. Electric generators are usually placed very far from residential areas.

In North America, electric voltage in residential areas is 110 volt while in most other parts of the world, the electric voltage is 220 volt. To carry the same amount of electric energy in a 110 volt system requires much thicker wire than in a 220 volt system. But when accidents occur, 110 volt causes less shock than 220 volt. In a system with abundant natural resources, such as North America, we often choose less resource efficient but safer options. In systems with scarce natural resources, we often choose more resource efficient but riskier options. There is a parallel in social systems. In a social system that controls more resource, its internal inequality is often low. But such system can utilize abundant resources as “energy slaves” (Nikiforuk, 2012) or impose inequality on other weaker social systems. In a social system that controls less resource, its internal inequality is often high. In such system, efficiency is very high for the elites, the designers of the system. But such system also has high probability to experience violent revolution. When factories are located in wealthy countries, much resource is used to control pollution. This lowers the ratio of output over resource input. But when factories are moved to poor countries, where local population has little political power, little resource is allocated to control pollution. This increases the ratio of output over resource input. Pollution is the reduction of chemical potential. Increasing pollution is the increase of inequality in chemical potential. By increasing inequality, the designers of the system gain higher efficiency and obtain cheaper products as a result.

It requires higher fixed cost to maintain a more unequal society. Dominant parties of a society do not necessarily hope to increase inequality all the time. When British Empire was expanding rapidly in the 19th century, it abolished slavery, a more extreme form of inequality. By adopting less unequal social systems, Britain was able to maintain and expand a huge empire with relatively little cost and huge profit. The inequality of a system also depends on how long the dominant parties expect the system to last. When we go fishing, we hope to have some inequality over fish. We use a fishing line to hook fish. But if the general public is allowed to use fishing nets in rivers, lakes and oceans, few fish will be left in a very short period of time. For an unequal system to last, the level of inequality cannot be too extreme, both in nature and in human societies. When the dominant parties expect the system to end soon, the inequality of the social system tends to increase so that dominant parties can extract more profits while the system lasts.

**2.6. Carbon and hydrogen as energy sources**

Carbon and hydrogen are the main components of organism. They are also the main component of the energy sources in organisms and in fossil fuels. Most energy sources we encounter, from foods we eat to gasoline we use to power our cars, are mainly combinations of carbon and hydrogen atoms. Hydrogen is lighter and has much higher energy density than carbon. Hence hydrogen has lower cost of transportation than carbon. At the same time, hydrogen energy is more costly to produce. For animals, because of their mobile lifestyle, the cost of transportation is high. So animals store a lot of energy as fat, which has high hydrogen content. Plants, because of their sedate lifestyle, are more economical to use carbon energy directly. They contain little fat in their bodies, with the exception of seeds, which have high energy demand and need to be more mobile than the plants themselves. This pattern applies to human society as well. Automobiles and airplanes, as transportation tools, are highly mobile. Their energy supply are petroleum products, such as gasoline, jet fuel and diesel, which contain high hydrogen content and are relatively light. Electricity generators are not mobile. The main energy input in electricity generation is coal, which is mainly carbon, heavy but cheap. The transportation of electric energy is to transport electrons, which are much lighter than atoms. So transporting electricity is a cheap way to transport energy. This is one reason electricity is universally used in most daily activities.

Among main energy resources, natural gas has the highest hydrogen content. That is why natural gas is preferred over coal as the energy source in home for cooking and heating. The last century can be thought as the transformation from carbon economy represented by coal to hydrogen economy represented by oil and natural gas. Therefore, people have moved toward a hydrogen economy long before official mandates from governments. This is also why natural gas and oil, which have much higher hydrogen content than coal, deplete much faster than coal. With the fast depletion of high quality (high hydrogen content) fossil fuel sources, can we create a man-made hydrogen economy?

Compared with coal, oil and natural gas burn more complete and emit less pollutant. In general, an energy source that reacts with the environment easier will leave less harmful residue. At the same time, since a cleaner energy react easier in the natural environment, it will be more difficult to be preserved. This is why coal is much more abundant than oil and natural gas. In the end, we have to rely on coal as our main energy supply after the depletion of oil and gas. The twenty-first century will become more a carbon economy instead of a hydrogen economy envisioned in some literature.

When the supply of high hydrogen content energy sources, such as oil and gas, are depleted, a man-made hydrogen energy can be produced from two possible sources. One is from renewable energy, such as solar, wind and biomass. We will discuss this option in a later section. The other is from low quality energy sources, such as coal. From the thermodynamic law, producing certain amount of high quality energy will require more low quality energy. Hence a hydrogen economy will produce more, not less pollution on the global level. The consequence of a hydrogen economy can be seen from an electricity economy. Electricity is a very clean form of energy. Its cleanness enables average households to utilize a huge amount of energy without feeling its negative impact. But power plants are the largest consumer of coal and other energy sources.

A parallel understanding can be made from the separation of residential area and industrial area. The separation makes residential areas cleaner. But the separation adds the extra pollution from transportation as people now need to commute between residential areas and work areas. The longer the distance between residential and industrial areas, the cleaner the residential areas are. But the total pollution will be higher because the extra pollution caused by transportation will be higher. At the country level, trade allows heavily polluted industries to be moved to poor countries where general population has little political power. While the rich countries enjoy cleaner environment, total pollution on the earth will increase because of the added transportation and communication costs. The concept of ecological footprint, which represents the consumption level of each country, provides a better measurement to the burden of human society to the environment (**Wackernagel and Rees, 1995).**

The prevailing wisdom on energy consumption is that hydrogen based energy should be promoted and the carbon based energy should be suppressed. However, hydrogen based energy is scarce and has already been depleted at fast pace because of their high quality. A further restriction on carbon as an energy source will accelerate the depletion of high quality energy sources and leave future generations in a worse shape. In a more sensible strategy of energy consumption, different energy sources should be utilized in different ways according to the differing physical and chemical properties of hydrogen and carbon. Natural gas, with the highest hydrogen content among carbohydrate fuels, is the cleanest. It can be used as fuels in densely populated residential areas. Gasoline, being a liquid and high energy density fuel because of its high hydrogen content, can be primarily used as transportation fuel, where energy supply has to be carried on vehicles. Coal, being largely carbon, is heavy, abundant and hence cheap. It can be economically used to generate electricity. Utility companies are large companies that can afford to make high fixed cost investment. Since power plants use large amount of fuel, they are in the best position to use expensive equipment to reduce the pollution from coal burning. Before the large price increase of oil during oil crisis in 1970s, many power plants used oil as fuels. But after that most power plants use coal as fuel (Dargay and Gately, 2010).

**2.7. Some patterns in energy economics**

Almost all the energy sources on the earth ultimately come from the sun. However, not all living organisms use solar energy directly. Several factors determine the pattern of energy use. The first important factor is energy density. Solar energy is vast. But the net return from transforming light energy into chemical energy, which organisms can store and transport easily for their further use, is low. Plants, whose sedate lifestyle requires low level of energy consumption, can effectively utilize solar energy directly through photosynthesis. Most animals, whose mobile lifestyle requires high level of energy input, could not support themselves through photosynthesis internally. Instead, some animals consume plants, which store high density chemical energy transformed from solar energy over a period of time. Other animals eat plant eating animals. Fossil fuels, which are further concentration of biomass in large scale, provide much higher energy density than biomass. The consumption of high energy density fossil fuels is the foundation of economic prosperity enjoyed by human societies in the last several hundred years.

The second factor of energy economics is the relation between the ease of storage and the ease of use. Electricity, which is very easy to use, is very difficult to store. The energy of biomass and fossil fuels are stored in the form of chemical energy, which is less easy to use compared with electricity, is easier to store. Nuclear energy, which requires very expensive system to harness, can be preserved for billions of years. This fundamental tradeoff is determined by the potential well of an energy source. The deeper the potential well, the easier to store the energy and the harder to use it. This is why attempts to reduce the storage cost of some easy to use energy sources, such as electricity, seem so elusive. On the other hand, it is often difficult for easy to store energy, such as fat, to be used easily. That is why it is so difficult for people to lose fat in their body. It is also difficult to achieve high energy density for easy to use energy sources, which react easily due to low potential well. Electricity is easier to use than gasoline. But the energy density in battery, a form of chemical energy, is generally low. For example, the energy density of lead battery is 0.16 \* 106 J/Kg while the energy density of gasoline is 44\* 106  J/Kg (Edgerton,1982, p. 74). Great progress has been made to increase energy density of batteries by utilizing smaller atoms, such as lithium. However, potential for further increase of energy density of batteries significantly is limited by the physical properties of electric energy. This is why progresses to develop electric cars that can drive similar distance to gasoline cars without recharging have been slow, although electric cars have a long history.

The third factor of energy economics is the efficiency of energy use and the total consumption of energy. Many people have advocated the increase of efficiency as a way of reduce energy consumption. Will the increase of efficiency reduce overall resource consumption? Jevons made the following observation more than one hundred years ago.

 It is credibly stated, too, that a manufacturer often spends no more in fuel where it is dear than where it is cheap. But persons will commit a great oversight here if they overlook the cost of improved and complicated engine, is higher than that of a simple one. The question is one of capital against current expenditure. … It is wholly a confusion of ideas to suppose that the economic use of fuel is equivalent to the diminished consumption. The very contrary is the truth. As a rule, new modes of economy will lead to an increase of consumption according to a principle recognized in many parallel instances. (Jevons, 1965 (1865), p. xxxv and p. 140)

Put it in another way, the improvement of technology is to achieve lower variable cost at the expense of higher fixed cost. Since it takes larger output for higher fixed cost systems to breakeven, to earn a positive return for higher fixed cost systems, the total use of energy has to be higher than before. That is, technology advancement in energy efficiency will increase the total energy consumption. Jevons’ statement has stood the test of time. Indeed, the total consumption of energy has kept growing, almost uninterrupted decades after decades, in the last several centuries, along with the continuous efficiency gain of the energy conversion (Inhaber, 1997; Smil, 2003; Hall, 2004).

 Figure 2.1: Global energy consumption (Million ton oil equivalent) from 1965 to 2012. Source: BP.

Figure 2.1 displays the total primary energy consumption worldwide from 1965 to 2012, a period of rapid technology progress. During this period, energy consumption grew steadily, with only two brief interruptions. From 1979 to 1982, a period of Iranian Revolution, oil price jumped from 13.60 US dollars per barrel in 1978 to 35.69 in 1980, causing serious recession in industrial world. The drop of energy consumption from 1979 to 1982 was due to sharp jump in oil price and the ensuing contraction of economic activities, not due to technology progress. Another brief period of interruption is year 2009. In 2008, oil price rose sharply, reaching 147 dollars per barrel at its peak. The ensuing Great Recession diminished the capacity for energy consumption. But the decline of energy consumption only last for one year.

We may also examine energy consumption of individual innovations such as hybrid cars. Hybrid cars need two engine systems, internal combustion engine and electronic motor. Hence their production requires more resource and energy input. If the owners of hybrid cars drive very little, the total energy cost of a hybrid car is actually higher than a conventional car due to the high resource consumption in manufacturing hybrid cars. Only when the owners of hybrid cars drive extensively, hybrid cars become more economical than conventional cars. Hence the use of more efficient cars will encourage, indeed enforce, the high consumption of energy. It should also be noted that hybrid cars, equipped with two engines, are heavier than conventional cars. Hence hybrid cars will be less fuel efficient in highway driving where frequent stop and acceleration is not required. In the end, hybrid cars, because of their high resource consumption in production and hence high prices, will become a symbol of status, just like SUV at present time.

**2.8. On the concept of renewable energy**

The deepening public concern about the long term availability of fossil fuels has promoted governments worldwide to adopt policies to subsidize research and production of renewable energy. To understand the effectiveness of these policies, we need to clarify the concept of renewable energy.

Almost all life forms depend on solar energy directly or indirectly for billions of years. So the use of renewable energy is not a new adventure. Instead, it has been practiced for billions of years by all life forms.

Almost all of the energy sources are renewable to some degree. Fossil fuels are generally considered to be non-renewable. But in fact they are been produced every day in various geological structures. However the rate of production of fossil fuels is much lower than the rate of consumption. A resource is renewable when its consumption rate is lower or equal to its regeneration rate. Otherwise it is not renewable. Hence, the concept of renewable resource is intrinsically linked to the level of consumption of that resource. For example, many supposed renewable resources, such as fishery, were collapsed when consumption levels became higher than regeneration. This will clarify a lot of confusions around renewable resources.

Producing ethanol from corn is an important part of renewable energy industry. Pimental and Patzek (2005) reviewed the past research works on the production of ethanol from corn and found the following results.

1. The total energy input to produce a liter of ethanol is higher than the energy value of a liter of ethanol. Thus there is a net energy loss in ethanol production from corn.
2. Producing ethanol from corn causes major air, water pollution and degraded environmental system.
3. There are over 3 billion people in the world are malnourished. Expanding ethanol production, which divert corn needed to feed people, raises serious ethical issues.

Despite the fact that ethanol production from corn is not a renewable energy, it pollutes environment and raise serious ethical issues, government subsidy on bio-energy has been expanding rapidly over time. Why is that? There could be several reasons.

1. Because of government subsidy, companies participating bio-energy production have been very profitable (Pimental and Patzek, 2005). Hence they will lobby hard for this type of subsidy.
2. The term “renewable resources” provides a psychological satisfaction for the general public. People can continue their grand lifestyle that requires high level of resource consumption while still feel morally superior because they invest on renewable resources, which are supposedly to be beneficial to future generations.
3. Governments are eager to be seen as doing “something” for the environment. Spending money on environment related issues projects a positive image. Government actions are generally channeled into directions with maximum political support and minimum political resistance. This means that government policies often benefit the current generation at the expense of future generations, who are not here to vote and lobby. This is true even for government policies that are supposed to benefit future generations.

Like any other viable investment in life, viable investment in energy should yield positive rate of return. Estimations of energy return on investment (EROI) of many renewable energy resources have been produced by experts in different areas. Very often the estimations are very high. For example, many estimations of energy return on investment of wind power is around 20 and some estimations of energy return on investment of solar panel is around 7. If the energy returns of these resources are really that high, the construction of these projects will spread like wild fire, even without any government subsidy. There will be no more problem of energy shortage because of the vast amount of solar energy. It is beyond my expertise to estimate the rate of return on each type of renewable energy resources. Instead, I will discuss some general economic principles that suggest the costs of most renewable energy programs have been seriously underestimated in some literature.

Living organisms, including human beings, generally utilize resources from easy to difficult, or in economic terms, from low fixed cost to high fixed cost. For example, humans use wood, coal, oil and natural gas, in that order, because the fixed cost of using them increases in each case. Promising renewable energy sources have not been brought into market place because of their high fixed costs. These high fixed costs are heavily subsidized through tax dollar funded university and industry research. The development and maintenance of most renewable energy technology requires high level of technology expertise, which is developed through the expensive education system. While education is funded by general tax revenue, it mainly benefits high tech industries.

All proponents of alternative renewable energy acknowledge the high fixed cost inherent in these new energy resources and argue that scale economy will eventually bring down the average cost. There are several types of economy of scales in the resource industry. We will discuss them separately.

The first type of scale economy is the scale of high tech research. The expensive high tech research only pays off when it market size is large. For example, people have been harnessing wind power for many centuries. But only with light and strong new materials such as fiberglass, generating electricity via wind power becomes feasible. Fiberglass, as a synthetic material made from petroleum, is a direct outgrowth from the fossil fuel industry. Indeed, most of metal based technologies are supported by the abundance of fossil fuels. When the fossil fuel becomes scarce, most modern industries, with their high fixed cost, may not have sufficient large scale to be viable. While the fixed costs in developing some renewable energy seem high today, they could be much higher in the future when fossil fuels become less abundant.

The second type of economy of scale is the use of large quantities of fossil fuel itself. For example, electrical transmission systems are very expensive to build and maintain. They only become economically viable when large amount of electricity, most of which is generated by fossil fuels, is being transmitted. It is the scale economy of fossil fuel generated electricity that supports the infrastructure that is also been used by alternative energy sources. If fossil fuels are excluded in electricity generation or depleted in the future, will the alternative energy sources be able to provide sufficient large market size to support the high cost electrical transmission system? This leads to the third type of scale economy: the scale of each type of renewable energy.

Hydro power, which consists of 7% of total electricity output worldwide, has already achieved substantial scale economy. At certain locations, hydro power has the advantage of high energy density, just like fossil fuels. This is also why hydro power has achieved such a large scale. For other renewable energy resources, such as biomass, solar and wind, the energy density may not be very high and steady, which poses a physical limit on the reduction of cost.

The current biosphere is the result of more than three billion years of evolution. In the competition for survival, many different ways of utilizing resources more economically have been explored. The most discussed forms of renewable energy, such as solar, wind, and bio energy, have been around for billions of years and have been extensively explored by many species, including human beings. While it is difficult to rule out the possibility that human beings can develop new technologies that has significantly higher overall efficiency in energy use than other living organisms and our own ancestors, the likelihood will be low. Human beings, like other dominant species, excel at controlling more resources, not at utilizing resources more efficiently (**Colinvaux, 1980).** Furthermore, research activities themselves are very resource intensive and accelerate the depletion of natural resources. Hence government policies about future should not rest on the assumption that technology progress will automatically substitute the demand for natural resources, as mainstream economic theory asserts (Samuelson and Nordhaus, 1998, p. 328).

**2.9. Global warming and climate change: Some paradoxes**

Global warming has become a hot topic in media for many years. Global warming is closely related to the increasing use of resources by human societies. Instead of providing a systematic review about this topic, we will simply point out some paradoxes in public discussion about this topic.

First, it is often argued that global warming will increase the intensity of natural disasters, such as hurricanes. Indeed, systems with higher energy level have higher power, including the power of destruction. For example, an airplane crash generates much more damage than a bicycle crash because an airplane embodies much higher energy level than a bicycle. Yet human beings embrace air travel. It is logically inconsistent to celebrate a high energy lifestyle and at the same time condemn the high energy level of the atmosphere, which is the byproduct of our high energy lifestyle.

Second, it is often pointed out that climate change brings high cost of adjustment. But so is technology change, which constantly renders existing technology and knowledge obsolete and worthless. While we deplore climate change, we applaud technology change, which is really the driving force behind the climate change. While we worry intensely about the high adjustment cost brought by the climate change, we congratulate the lifetime adjustment or lifetime learning brought by the technology change.

Third, it is often argued that global warming could bring some serious unpredictable consequences, such as global freezing. But so is innovation. While innovation is lavished rewarded with patent protections and government subsidy, global warming, which is brought by the innovative technologies developed in human societies, is viewed with suspicion. We talk about conservation. But we put our faith to further and accelerated innovation.

Fourth, we seem to worry a lot about endangered species or extinct species. But at the same time, we praise the efficiency gain of free competition in the market, which causes the extinction of many businesses each day. The dominance in a sector by some firms, such as Microsoft, often reduces the number and diversity of firms. However, competition in the business world is generally regarded as a good thing while competition in nature is judged by a different criterion.

Fifth, people seem genuinely concerned about global warming. However, given opportunities, people will move to truly warm places, such as Florida and Southern California and move away from truly cold places. Among the top ten towns of largest population decline in the last several years in Canada, five are from northern British Columbia, where land is covered by snow for more than half a year. Global warming becomes an issue because most people, if given options, prefer to live in warm places. As a result, more and more people live in warm places. For people living in warm places, hot summer months are less than ideal, hence the harm of global warming. For those who live in cold places who have to endure long and bitter winter months, a little bit warming is great news. However, since the percentage of people who live in warm places is high and growing, media and academic research, in catering to their clients, prefer to talk about the damage of global warming. It is true that even people living in true North have to complain about the damage caused by the global warming, such as the disappearance of glaciers, to get media coverage. However, countries around arctic region have intensified their effort to claim sovereignty to various areas in the arctic in recent years. Apparently, warming increases instead of decreases the value of cold places, which is of course hardly surprising.

Sixth, most programs that supposed to battle global warming, such as carbon offset program, actually accelerate the pace of global warming and resource depletion. Carbon offset programs, which promise, for a fee from clients, to reduce carbon dioxide from the atmosphere. So clients can continue to consume large amount of resources while continuing to enjoy a sense of moral superiority. One popular offset program is tree planting. However, if we consider economy of human society as part of the economy of nature, we will recognize that active planting trees by humans will increase, not decrease, carbon emission. Since the beginning of the agricultural civilization, most human interventions in land management have been to increase and maintain farm land from originally forest land. As a result, forests are gradually replaced with farm land or pasture land, often with great effort from human beings. In most cases, forests can be restored without any human effort. You simply leave the land alone and trees will grow back. Human intervention, such as carbon offset programs, will only increase energy consumption and carbon emission. Because most carbon offset programs increase instead of decrease carbon emissions, they inevitably become the hotbed for frauds (Schapiro, 2010). Despite the heated rhetoric against global warming, most of our social institutions continue to encourage activities that cause global warming and climate change. In particular, many prominent anti global warming crusaders are the ones who generate the largest amount of carbon footprint themselves and promote programs that actually produce more carbon dioxide.

The negative impacts of carbon dioxide in the atmosphere are well documented and publicized in the literature and media. However, higher concentration of carbon dioxide in the atmosphere also has great beneficial effects. Carbon forms the structural foundation of all organisms. Carbon dioxide in the earth’s atmosphere provides the raw material for all plants. The usefulness of carbon makes carbon dioxide an extremely scarce resource in nature. It is present in the Earth's atmosphere at a concentration of only 0.038%. The low concentration of carbon dioxide is the bottle neck in plant’s growth in many cases. The plants’ need for more carbon dioxide is so great that C4 plants, such as corn and sugarcane, were evolved, which can utilize carbon dioxide more efficiently, but at a higher energy cost (Atkins, 1991). The increase of concentration of carbon dioxide in the atmosphere improves the growth condition of plants and contributes greatly to the increase of food production in the last several decades, which is essential to feed the world’s growing population. It will be very helpful to provide a systematic study to weigh the negative and positive effects of the increase of concentration of carbon dioxide in the atmosphere. However, there has been a long term misconception about carbon dioxide, which makes the objective evaluation of it very difficult.

The human mind is inherently inclined to take moralistic view of nature. Prior to the modern scientific era ... nearly every problem was viewed as an alternative between good and evil, righteousness and sin, God and the Devil. This superstitious slant still distorts the conceptions of health and disease; indeed, it is mainly derived from the experience of physical suffering. Lavoisier contributed unintentionally to this conception when he defined the life supporting character of oxygen and the suffocating power of carbon dioxide. Accordingly, for more than a century after his death, and even now in the field of respiration and related functions, oxygen typifies the God and carbon dioxide is still regarded as a spirit of Evil. There could be scarcely a greater misconception of the true biological relation of these gases. (Henderson, 1940)

 His words are as relevant today as more than seventy years ago.

**2.10. Concluding remarks**

Because of the importance of energy in our life, people have pursued the dream of renewable energy since time immemorial. In the course of history, many people believe they have discovered an inexhaustible source of energy, such as battery, or invented one or another kind of perpetual motion machine. They think their discoveries or inventions can be put into practical use in large scale once necessary technical improvements can be made in the future. However, more rigorous investigation leads to the development of thermodynamic theory, which rules out the possibility of a perpetual, or renewable energy source without external input.

In this chapter, we further investigate how physical environment enables and constrains living organisms and economic systems by integrating the economy of human society into the economy of nature. We explore the relation between natural resources and technology in human society. It helps us envision the future of human society in an environment of increasingly scarce and costly natural resources. The main results can be summarized as follows. First, the survival and prosperity of human society depends entirely on the availability of natural resources. Second, while the forms of natural resources are diverse, they can be understood from the unifying principle as low entropy sources. Third, to utilize natural resources, fixed structures are required, which consume resources themselves. Fourth, when certain structures can generate positive returns on the use of natural resources over an indefinite period of time, these structures are called living organisms. Fifth, it is the unique chemical properties of carbon that enables it to become the backbone of life. The major non-renewable resources that our industrial civilization builds on, such as coal, petroleum and natural gases, are generated from the remains of the living organisms. They all contain carbon.

Some practical implications emerge from our theoretical discussion. First, we prefer high quality resources over low quality resources. This is why we move gradually from more carbon based fuels, such as coal, to less carbon based fuels, such as natural gas and petroleum. However, as high quality resources, such as conventional oil, are seriously depleted, human society will be forced to move back toward a carbon based economy from the current mixed carbon and hydrogen economy. This contradicts the often dreamed hydrogen economy in the future. Second, the programs that supposed to help environment, such as carbon offset program, are decorative frills in a society. Empirical evidences show that decorative frills in all animals, including human beings, increase instead of decrease the burden to the environment. Third, increasing energy efficiency, which requires the increase of fixed cost, will increase total resource consumption. This was pointed out by Jevons more than a hundred years ago. Fourth, due to the levels of potential well, energy sources that are easy to use, such as electricity, are difficult to store. This is why it is so difficult to develop electric cars that can drive long distance without recharging.

**Chapter Three: Production: A Mathematical Theory**

**3.1. A historic review of related ideas and mathematical techniques**

Because of the fundamental link between thermodynamics and life, many attempts have been made to develop analytical theories based on the principle of thermodynamics and apply them to living systems and human society. Two of the influential theories are Lorenz’ chaos theory and Prigogine’s far from equilibrium thermodynamic theory. Lorenz, a meteorologist, simplified weather equations, which are thermodynamic equations, into ordinary differential equations. He found chaos properties from these equations. Prigogine developed the theory from some chemical reactions. Ping Chen has written extensively to apply Prigogine’s theory to social sciences, offering great insights into many social problems (Chen, 2010). The theories of Lorenz and Prigogine greatly influenced the thinking in biology and social sciences. However, these theories, as well as other related mathematical theories, do not model life process or social activities directly.

Since uncertainty is an integral part of life processes, the advancement of stochastic calculus is essential for the development of an analytical thermodynamic theory of life and human society. In the past several decades, some fundamental works in the area of stochastic calculus were undertaken by people with very diverse backgrounds. Three works are particularly relevant to the development of our theory. The first is Ito’s Lemma, which provides a rule to find the differential of a function of stochastic variable. Ito’s Lemma was obtained in 1940s. But its importance was not recognized until its wide spread application in financial economics several decades later.

The second tool is Feynman-Kac formula, which maps a stochastic process into a deterministic thermodynamic equation. In natural science, there is a long tradition of studying stochastic processes with deterministic partial differential equations. For example, heat is a random movement of molecules. Yet the heat process is often studied by using heat equations, a type of partial differential equations. Richard Feynman (1948) attempted to simplify calculation in quantum mechanics by transforming problems in stochastic processes into problems in deterministic processes. The new mathematical technique enabled him to perform many computations in quantum mechanics which were very difficult in the past. With this he established the theory of quantum electrodynamics. The breakthrough in physics is often generated by the breakthrough in new mathematical methods, which enables us to describe the subtler parts of the nature. An important motivation in Feynman’s research was his seek for universality. “The question that then arose was what Dirac had meant by the phrase ‘analogous to,’ and Feynman determined to find out whether or not it would be possible to substitute the phrase ‘equal to.’” (Feynman and Hibbs, 1965, p. viii) Kac (1951) extended Feynman’s method into a mapping between stochastic process and partial differential equations, which was later known as the Feynman-Kac formula. Despite its highly technical nature, Feynman-Kac formula is a very general result and has proved to be extremely useful in many different fields (Kac, 1985).

Kolmogorov developed more systematic approach to stochastic calculus (Kolmogorov, 1931). He defined two different types of mappings from stochastic processes to deterministic differential equations as backward equations and forward equations. Kolmogorov backward equations are equivalent to Feynman-Kac formula. Kolmogorov forward equations are equivalent to Fokker-Planck equation in physics. Two types of equations have different physical meanings. It will be very important to distinguish them in specific applications, as we will see later.

The third is Black-Scholes (1973) option pricing theory, which provides an analytical formula of observable variables to price a financial instrument whose payoff depends on a stochastic process. This is a landmark contribution in social sciences. It shows that a complex economic problem can be effectively modeled by a simple analytical theory and much information about it can be obtained through such an analytical theory. Fischer Black, one of the co-developers of the Black-Scholes theory, was a legendary figure in finance. Jack Treynor, who introduced Fischer Black to the field of finance, had the following observation:

Fischer never took a course in either economics or finance, so he never learned the way you were supposed to do things. But that lack of training proved to be an advantage … since the traditional methods in those fields were better at producing academic careers than new knowledge. Fischer’s intellectual formation was instead in physics and mathematics, and his success in finance came from applying the methods of astrophysics. Lacking the ability to run controlled experiments on the stars, the astrophysist relies on careful observation and then imagination to find the simplicity underlying apparent complexity. In Fischer’s hands, the same habits of research turned out to be effective for producing new knowledge in finance. (Mehrling, 2005, p. 6)

Jack Treynor (1996) summarized:

Fischer’s research was about developing … insightful, elegant models that changed the way we look at the world. They have more in common with the models of physics --- Newton’s laws of motion, or Maxwell’s equations --- than with the econometric “models” --- lists of loosely plausible explanatory variables --- that now dominate the finance journals.

In Black-Scholes option theory, the price movement of financial assets is modeled with lognormal processes

where *r* is the rate of expected return and *σ* is the rate of uncertainty. Option prices, as functions of prices of its underlying assets, satisfy the following Black-Scholes equation.

I had been thinking about an analytical thermodynamic theory of life systems for many years when I learned about the Black-Scholes option theory. The most fundamental property of life is their ability to extract low entropy from the environment to compensate continuous dissipation. Soon I realized this property can be represented by lognormal processes, where *r* is the rate of extraction of low entropy and *σ* is the rate of diffusion. Every stochastic process can be mapped into a deterministic thermodynamic equation, which is often easier to handle and yields more results. So I hope Black-Scholes equation and option theory may offer some insight for an analytical thermodynamic theory of life systems. After several years, I first developed such a theory based an analogy between option theory and living systems. Later I was able to derive the theory directly without depending on its analogy with option theory. In the next section, we will provide an updated version of this theory.

**3.2. A mathematical theory of production**

The theory described in this section can be applied to both biological and economic systems. For simplicity of exposition, we will use the language of economics. However the extension to biological system is straight forward.

We start the investigation by asking: What are the most fundamental properties of organisms and organizations? How do we represent these fundamental properties in a mathematical theory? First, organisms and organizations need to obtain resources from the environment to compensate for the continuous diffusion of resources required to maintain various functions. This fundamental property can be represented mathematically by lognormal processes, which contain both a growth term and a dissipation term.

Suppose *S* represents the amount of resources accumulated by an organism or the unit price of a commodity, *r*, the rate of resource extraction or the expected rate of change of price and σ, the rate of diffusion of resources or the rate of volatility of price change. Then the process of *S* can be represented by the lognormal process

where

 

The process (1) is a stochastic process. Although a stochastic process will generate many different outcomes over time, we are mostly interested in the average outcomes from such processes. For example, although the movement of individual gas molecules is very volatile, air in a room, which consists of many gas molecules, generates a stable pressure and temperature. We usually study the average outcomes of stochastic processes by looking at the averages of the underlying stochastic variables and their functions. These investigations often transform stochastic processes into their corresponding deterministic equations. For example, heat is a random movement of molecules. Yet the heat process is often studied by using heat equations, a type of deterministic partial differential equation.

Feynman (1948) developed a method of averaging stochastic processes under very general conditions, which is usually called path integral. Kac (1951) extended Feynman’s method into a mapping between stochastic processes and partial differential equations, which was later known as the Feynman-Kac formula. According to the Feynman-Kac formula (**Øksendal**, 1998, p. 135), if

 

is the expected value of a function of *S*  at time *t* discounted at the rate *q*, then *C(t,S)* satisfies the following equation

with

 

It should be noted that many functions of *S* satisfy equation (3). The specific property of a particular function is determined by the initial condition (4). This is similar to Black-Scholes option theory. The Black-Scholes equation is satisfied by any derivative securities. It is the end condition at contract maturity that determines the specific property of a particular derivative security.

Second, for an organism or an organization to be viable, the total cost of extracting resources has to be less than the amount of resources extracted, or the total cost of operation has to be less than the total revenue. Costs include fixed cost and variable cost. In general, production factors that last for a long time, such as capital equipment, are considered fixed cost while production factors that last for a short time, such as raw materials, are considered variable costs. If employees are on long term contracts, they may be better understood as fixed costs, although in the economic literature, they are usually classified as variable costs. Typically, a lower variable cost system requires a larger investment in fixed costs, though the converse is not necessarily true. Organisms and organizations can adjust their level of fixed and variable costs to achieve high level of return on their investment. Intuitively, in a large and stable market, firms will invest heavily in fixed cost to reduce variable cost, thus achieving a higher level of economy of scale. In a small or volatile market, firms will invest less in fixed cost to maintain a high level of flexibility. In the following, we will examine the relation between fixed cost and variable cost in a very simple project.

Suppose there is a project with a duration that is infinitesimally small. It only has enough time to produce one unit of product. If the fixed cost is lower than the value of the product, in order to avoid arbitrage opportunity, the variable cost should be the difference between the value of the product and the fixed cost. If the fixed cost is higher than the value of this product, there should be no extra variable cost needed for the product. Mathematically, the relation between fixed cost, variable cost and the value of product in this case is the following:

where *S* is the value of the product, *C* is the variable cost and *K* is the fixed cost of the project. When the duration of a project is of a finite value *T*, relation (3.5) can be extended into

as the initial condition for equation (3.3). Equation (3.3) with initial condition (3.6) can be solved to obtain

where

The function *N(x)* is the cumulative probability distribution function for a standardized normal random variable. From (3.6), the solution of the equation (3.3) can be interpreted as the variable cost of the project. However, we will investigate shortly whether the function represented in formula (3.7) has common properties of variable costs. For a given investment problem, different parties may select different discount rates. To simplify our investigation, we will make the discount rate equal to the expected rate of growth. This is to set

 

This choice of discount rate can be understood from two perspectives. From a biological perspective, fast growing organisms also have a high probability of death. In a steady state, the growth rate has to be equal to the death rate. In the biological literature, the discount rate is usually set equal to the growth rate (Stearns, 1992). From the perspective of economics, in option theory, the discount rate is set equal to the risk free rate. The level of risk of an option contract is represented by implied volatility, which does not necessarily equate with past volatility or future expected volatility. Some people do not agree with the economic logic behind the mathematical derivation of Black-Scholes equation that made the risk related discount rate disappear (Treynor, 1996). However, the disappearance of the separate discount rate greatly simplified our understanding of how option values are related to market variables. From both a biological and economic perspective, this choice of discount rate provides a good starting point for further investigation.

With *q* equals to *r*, equation (3.3) becomes

and solution (3.7) becomes

This takes the same form as the Black-Scholes formula for European call options. But the meanings of the parameters in this theory differ from that in the option theory. Formula (3.10) provides an analytical formula of variable cost as a function of product value, fixed cost, uncertainty, duration of project and discount rate of a firm. Similar to understanding in physics, the calculated variable cost is the average expected cost.

We will briefly examine the properties of formula (3.10) as a representation of variable cost. First, the variable cost is always less the value of the product when the fixed cost is positive. No one will invest in a project if the expected variable cost is higher than the product value. Second, when the fixed cost is zero, the expected variable cost is equal to the value of the product. When the fixed cost approaches zero, the expected variable cost will approach the value of the product. This means that businesses need to make a fixed investment before they can expect a profit. Similarly, all organisms need to invest in a fixed structure before they can extract resources profitably. Some do not agree with this statement and provide examples of low fixed cost investment with high profits, such as J. K. Rowling writing Harry Potter books. Our results are about the statistical average. While a small percentage of authors earn high incomes from blockbusters, an average author does not earn a high income. Third, when fixed costs, *K*, are higher, variable costs, *C*, are lower. Fourth, for the same amount of the fixed cost, when the duration of a project, *T*, is longer, the variable cost is higher. This shows that investment value depreciates with time. Fifth, when risk, σ, increases, the variable cost increases. Sixth, when the discount rate becomes lower, the variable cost decreases. This is due to the lower cost of borrowing. All these properties are consistent with our intuitive understanding of and empirical patterns in production processes.

After obtaining the formula for the variable cost in production, we can calculate the expected profit and rate of return of an investment. Suppose the volume of output during the project life is *Q*, which is bound by production capacity or market size. During the project life, we assume the present value of the product to be *S* and the variable cost to be *C*. Then the total present value of the product and the total cost of production are

respectively. The net present value of the project is

 $QS-\left(QC+K\right)=Q\left(S-C\right)-K (3.12)$

The rate of return of this project can be represented by

It is often convenient to represent *S* as the value of output from a project over one unit of time. If the project lasts for *T* units of time, the net present value of the project is

 

The rate of return of this project can be represented by

Unlike a conceptual framework, this mathematical theory enables us to make quantitative calculations of returns of different projects under different kinds of environments. Jack Treynor’s comment about Black-Scholes theory provides a relevant background to understand our production theory (Treynor, 1996):

Time has always been a pesky problem for economists, who have dealt with it by

1. Restricting their model to perpetuities (Modigliani and Miller).
2. Focusing on one-period problems (Markowitz’s portfolio balancing model).
3. Reducing the dynamic flow of economic events to a static long run and a static short run (Alfred Marshall).

That these pioneers in quantifying the previous unquantifiable ducked the problems is a measure of what Black-Scholes accomplished.

Our production theory, as an extension of the Black-Scholes methodology, can be applied directly to refine the theories of these pioneers. By considering corporate finance problems in a finite time horizon, we are able to provide a more precise understanding of the problems related to capital structure considered by Modigliani and Miller (Chen, 2006a). By working in a continuous time framework instead of a one period framework, we are able to obtain a more refined understanding of relations among risk, discounting and duration of projects. This helps us understand patterns such as hyperbolic discounting (Ainslie, 1992). By identifying long run cost as fixed cost and short run cost as variable cost and establishing their relations, we provide an analytical theory of economic dynamics that was conceived by Alfred Marshall qualitatively.

Soon after Black-Scholes (1973), it became apparent to many researchers that similar approaches may be applied to capital investment. These approaches are generally called the real option theory. The book by Dixit and Pindyck (1994) is the acknowledged classic in real option theory. In that book, many partial differential equations were derived, but no analytical results about the key factors in capital investment were obtained. As a result, the real option theory “either use stylized numerical examples or adopt a purely conceptual approach to describing how option pricing can be used in capital budgeting” (Megginson, 1997, p. 292). In comparison, the production theory presented here provides simple analytical formulas for the key parameters in capital investment. A detailed literature review and comparison between this theory and the real option theory was provided in Chen (2006b).

From mathematical perspective, Dixit and Pindyck (1994) adopted Kolmogorov forward equations in their book. This is a little bit surprising, for in their book, they stated, “Feynman could be claimed as the father of financial economics” (Dixit and Pindyck, 1994, p. 123). Probably they were not aware that Kolmogorov forward equations and Feynman’s method were different methods. Currently, Kolmogorov forward equations, which are called Fokker–Planck equations in physics, are more widely used in economic theories (Aoki and Yoshikawa, 2006). Kolmogorov forward equations describe the evolution of probability distributions of a system while Kolmogorov backward equations describe the evolution of average values of a system. Since many decision makings are about the average values of gains or losses, Kolmogorov backward equations and Feynman-Kac formula provide natural representations in many important economic problems.

Ping Chen (2010) questioned the validity of using lognormal processes in modeling economic processes. He suggested that lognormal processes grow exponentially while no economic system will grow forever. He is certainly right that no economic system grows forever. But our theory is concerned about biological and economic processes over each generation or duration of a project, which lasts for a finite period of time. Over a finite period of time, biological and economical systems can and do grow exponentially. So our theory is not inconsistent with basic scientific principles.

Since our theory is very similar to Black-Scholes option theory, it is essential to compare the basic equations of Black-Scholes theory and our theory. The Black-Scholes equation is

The basic equation in our theory is

The Black-Scholes equation has a negative sign in front of the time derivative. From the physics perspective, our equation represents a thermodynamic process while the Black-Scholes equation represents an inverse thermodynamic process. From the economic perspective, Black-Scholes equation solves the current option price when the future payout is determined while our equation solves the expected variable cost in the future when the current fixed cost is determined. This is the main difference between two theories.

This theory provides quantitative relations about major factors in economic and biological systems. First we will examine how variable costs change with fixed cost at different levels of uncertainty. By calculating variable costs from (3.10), we find that, as fixed costs are increased, variable costs decrease rapidly in a low uncertainty environment and change very little in a high uncertainty environment. To put it in another way, high fixed cost systems are very sensitive to the change of uncertainty level while low fixed cost systems are not. This is illustrated in Figure 3.1.

**Figure 3.1. Fixed cost and uncertainty**: In a low uncertainty environment, variable cost drops sharply as fixed costs are increased. In a high uncertainty environment, variable costs change little with the level of fixed cost.

The above calculation indicates that systems with higher fixed investment are more effective in a low uncertainty environment and systems with lower fixed investment are more flexible in high uncertainty environment. Mature industries, such as household supplies, are dominated by established large companies such as P&G while innovative industries, such as IT, are pioneered by small and new firms. Microsoft, Apple, Yahoo, Google, Facebook and countless other innovative businesses are started by one or two individuals and not by established firms. Similarly, in scientific research, mature areas are generally dominated by top researchers from elite schools, while scientific revolutions are often initiated by newcomers or outsiders (Kuhn, 1996).

We now discuss the returns of investment on projects of different fixed costs with respect to the volume of output or market size. Figure 3.2 is the graphic representation of (3.13) for different levels of fixed costs. In general, higher fixed cost projects need higher output volume to breakeven. At the same time, higher fixed cost projects, which have lower variable costs in production, earn higher rates of return in large markets.

**Figure 3.2. Fixed cost and the volume of output**: For a high fixed cost investment, the breakeven market size is higher and the return curve is steeper. The opposite is true for a low fixed cost investment.

We can see from the above discussion that the proper level of fixed investment in a project depends on the expectation of the level of uncertainty and the size of the market. When the outlook is stable and the market size is large, projects with high fixed investment earn higher rates of return. When the outlook is uncertain or market size is small, projects with low fixed cost breakeven easier.

In the ecological system, the market size can be understood as the size of resource base. When resources are abundant, an ecological system can support large, complex organisms (Colinvaux, 1978). Physicists and biologists are often puzzled by the apparent tendency for biological systems to form complex structures, which seems to contradict the second law of thermodynamics (**Schneider and Sagan**, 2005; Rubí, 2008). However, once we realize that systems of higher fixed cost provide higher return in the resource rich and stable environments, this evolutionary pattern becomes easy to understand. An example from physiology will highlight the tradeoff between fixed and variable cost with different levels of output.

An increased oxygen capacity of the blood, caused by the presence of a respiratory pigment, reduces the volume of blood that must be pumped to supply oxygen to the tissues. …The higher the oxygen capacity of the blood, the less volume needs to be pumped. There is a trade-off here between the cost of providing the respiratory pigment and the cost of pumping, and the question is, Which strategy pays best? It seems that for highly active animals a high oxygen capacity is most important; for slow and sluggish animals it may be more economical to avoid a heavy investment in the synthesis of high concentrations of a respiratory pigment. (Schmidt-Nielson, 1997, p. 120)

For high output systems (highly active animals) investment is fixed cost (respiratory pigment) is favored while for low output systems (slow and sluggish animals) high variable cost (more pumping) is preferred. Pumping is variable cost compared with respiratory pigment because respiratory pigment lasts much longer.

In the next section, we will provide some examples to apply the mathematical theory to understand different economic phenomena. In Section 3.4, we will apply the theory to understand monetary policies and business cycles, which are among the most important problems in economic theories.

**3.3 Several numerical examples**

With a simple analytical theory, we can easily observe the impacts of particular parameters or relations among parameters from simple calculations. In the following, we will present some examples.

All parameters in our theory, except uncertainty, correspond to directly observable quantities. This is very similar to option pricing theory, in which all parameters, except volatility, correspond to directly observable quantities. In option pricing theory, volatility is often called implied volatility because volatility is implied from the option prices. Similarly in our theory, uncertainty is implied from the expected variable costs. Indeed, the value of uncertainty rate can involve many factors. The meaning of uncertainty rate can be very different in different applications.

**Economy of scale and the law of diminishing return**

All economic systems experience economy of scale and the law of diminishing return at the same time. This can be modelled by setting uncertainty as a function of output. This can be modeled with uncertainty, σ, as an increasing function of the volume of the output. Specifically, we can assume

 

Where σ0 is the base level of uncertainty, *Q* is the volume of output and *l* > 0 is a coefficient. Intuitively, when the size of a company increases and the business expands, the internal coordination and external marketing becomes more complex. With the new assumption, we can calculate the rate of return of production from formula (3.13). The result from the calculation is presented in Figure 3.3. From Figure 3.3, the rate of return initially increases with the production scale, which is well known as the economy of scale. When the size of the output increases further, the rate of return begin to decline. This is the law of diminishing return. In specific applications, we can analyze how each factor influence the shape of return curves and try to obtain high rate of returns for our investment.

**Figure 3.3 Volume of output and the rate of return:** The rate of return of a project with respect to volume of output, when uncertainty is an increasing function of volume of output

**Increasing fixed cost to reduce uncertainty**

When the fixed cost of a system increases, the increased fixed cost can often help reduce uncertainty. Air conditioning and heating systems can reduce the temperature uncertainty of a building. Insurance can reduce uncertainty of large losses. This pattern can be modeled with uncertainty, σ, as a decreasing function of the fixed cost. Specifically, we can assume

 

where σ0 is the base level of uncertainty, *K* is the fixed cost and *l* > 0 is a coefficient. Assume the unit product value is 1, discount rate is 5% per annum, duration of project is 10 years. Assume σ0 is 20% per annum and *l* is 0.2. Calculated rate of return from the project with different levels of fixed cost is shown in Figure 3.4. When the level of fixed cost is increased, the rate of return increases initially and then declines.

Many decision makings involve the spending of fixed cost to reduce uncertainty, such as unemployment insurance, old age insurance, medical insurance, and governments’ guarantee to financial institutions. There are often debates about how much we need to spend to reduce uncertainty in our life. A proper model will help us reach compromise among various parties.

**Figure 3.4. Increasing fixed cost to reduce uncertainty**

**Resource abundance and investment patterns**

One way to represent the resource abundance is by the level of uncertainty rate. In physics, the term representing uncertainty rate in lognormal process is often called diffusion rate. Higher diffusion rate means that less energy is applied to do useful work, indicating low quality of energy fuels. For example, when a dry cell gets discharged, its internal resistance gradually increases and more energy turns into unusable heat. So the abundance of resources can be represented by the (inverse of the) uncertainty rate.

Table 3.1 lists the fixed costs, durations, net present values at different levels of uncertainty rate when net present value from formula (3.13) is maximized with respect to fixed cost and duration, assuming a constant discount rate (4%). Low uncertainty rates, which represent abundant resources, correspond to the choice of high fixed costs and long durations, and to high net present values. Since fixed costs and durations are more visible than resource abundance, highly valuable investments are often seen to be the result of those investment choices. More generally, high fixed costs and long durations are often associated with progress.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Uncertainty rate | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 |
| Fixed cost | 10.3 | 8.1 | 6.4 | 5.1 | 4.1 | 3.3 |
| Duration | 42.8 | 35.7 | 29.7 | 24.9 | 20.9 | 17.8 |
| Net present value | 18.4 | 13.6 | 10.1 | 7.6 | 5.7 | 4.3 |

**Table 3.1**: Investment decisions and values of investments with different levels of resource abundance

From Table 3.1, when resources are very abundant (uncertainty rate equals 0.2), the investment that generates highest net present value has a fixed cost of 10.3 and duration of 42.8 years. The performance of such a level of investment with different levels of resource abundance can be calculated from (3.13). The results are presented in Table 3.2. When resources become increasingly scarce, net present values of the same investment decline and eventually drop below zero.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Uncertainty rate | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 |
| Fixed cost | 10.3 | 10.3 | 10.3 | 10.3 | 10.3 | 10.3 |
| Duration | 42.8 | 42.8 | 42.8 | 42.8 | 42.8 | 42.8 |
| Net present value | 18.4 | 13.0 | 8.3 | 4.1 | 0.6 | -2.2 |

**Table 3.2**: Values of high fixed cost, long duration investment with different levels of resource abundance

In the past several centuries, with continuous improvement in the extraction and use of fossil fuels, the amount of resources consumed has increased steadily. Social institutions make many adjustments, generally increasing the fixed cost and duration of investments to take advantage of this abundance. Because of the high correlation between fixed cost, life span and general prosperity, decision makers often reflexively adopt policies that increase fixed costs and life span, expecting to take advantage of the greater efficiency those systems yield. Secondary education becomes mandatory and tertiary education becomes easily available in most wealthy countries, greatly increasing the fixed cost of social life. Retirement age has been continuously extended. However, in the last several decades, some social problems, such as high youth unemployment, below replacement fertility and population aging, have become increasingly prominent. From our calculation, in an environment with high resource cost, high fixed cost investment and long duration working years will lead to negative returns of social systems, with many accompanying consequences. The solution would be the reduction of fixed cost and earlier, not later, retirement age. While it is politically difficult to implement these policies, we hope our theory help to stimulate discussion on these important issues.

**3.4 Monetary policies and business cycles**

Fischer Black once stated,

I like the beauty and symmetry in Mr. Treynor’s equilibrium models so much that I started designing them myself. I worked on models in several areas:

 Monetary theory

 Business cycles

 Options and warrants

For 20 years, I have been struggling to show people the beauty in these models to pass on knowledge I received from Mr. Treynor.

In monetary theory --- the theory of how money is related to economic activity --- I am still struggling. In business cycle theory --- the theory of fluctuation in the economy --- I am still struggling. In options and warrants, though, people see the beauty. (Mehrling, 2005, p. 93)

Apparently, Black thought monetary theories, business cycles and options can be understood by related theories. His intuition was very insightful. In the following, we will apply the analytical theory just derived to understand monetary policies and business cycles.

The main tool of monetary policy is the setting of interest rate levels. We will examine investment decisions at countries with different levels of interest rate while keeping other parameters constant. Suppose in two countries A and B, annual consumption of total product is 1 billion dollars. Uncertainty rate is 30% per annum. Decision makers attempt to maximize the net present value of investment project, which is a common criterion in investment. Suppose discount rates are 3% and10% per annum in country A and B respectively. We will ask the following questions. How much will be the desired fixed costs and how long will be the expected project durations? What are the net present values of projects in countries A and B?

We maximize (3.13), which is net present value of an investment by changing fixed cost, *K* and duration, *T* when interest rates are set at 3% and 10% per annum respectively. With commonly used software, such as Excel, we can find the solution easily. The following table lists the calculated results.

|  |  |  |  |
| --- | --- | --- | --- |
| Interest rate | Fixed cost | Duration | NPV |
| 3% | 7.1 | 35.7 | 12.3 |
| 10% | 3.9 | 13.6 | 4.0 |

From the table, we find that when interest rate is lower, the amount of fixed investment is larger, investment duration is longer and the net present value is higher. So investors normally prefer low interest rate environment. However, the net present values are expected returns calculated at the beginning of a project. The actual returns depend on future environment. Suppose after the projects are built, the actual level of uncertainty becomes 80% per annum instead of 30% per annum as previously expected. We can recalculate the net present values from (3.13) to find the net present value of the first project, built in the low interest rate environment of 3%, becomes -6.2 billion dollars while the net present value of the second project, built in the high interest rate environment of 10%, becomes -2.0 billion dollars. Both projects suffer losses. But the first project suffers much more losses. When environmental conditions change, values of investment in low interest rate environment experience larger fluctuations. In other words, the monetary policy of low interest rate will generate greater business cycles. So business cycles are greatly tied to monetary policies. Our theory provides a simple and clear understanding about the level of interest rate and magnitude of business cycles.

If low interest rate policy is associated with large magnitude of business cycles, why low interest rate policy has become the common choice for most governments most of the time? When interest rates are lower, the borrowing costs are lower. This is especially helpful for high fixed cost projects, which often require substantial financing. Higher fixed cost systems have lower variable cost and hence have more significant scale economy than lower fixed cost systems. When the market size for a product or a service is large and growing, a small number of large projects provide higher returns than a large number of small projects.

In a growing economy, businesses often anticipate growth. They build up very high fixed cost projects, and with their corresponding low variable costs, take advantage of future large market size. However, high fixed cost systems require higher levels of output to breakeven. While future demand is expected to be high, the current revenue is often not high enough to support the operation of high fixed cost systems. When there is a large gap between revenue and cost in an economic system, and the capital market is unable to fill the gap, recession occurs. To stimulate demand, governments often lower interest rate to ease borrowing from consumers. Hence, low interest rate policy has been used to stimulate both investment and consumption. Since high fixed cost systems are more competitive in a large and growing economy, and a low interest rate environment helps high fixed cost systems, low interest rate policy has in the past been adopted by most governments, despite their magnifying impacts on business cycles.

High fixed cost production systems, with their low variable cost, often reduce the average cost of products in a large and expanding economy. At the same time, the expansion of the economy tends to strain the availability of inputs, which often put an upward pressure on the costs of products. In the past, the abundance of cheap oil ensured that the inflation pressure was mild most of the time. However, with the increasing scarcity of resources, rapid economic expansions often trigger rapid commodity price inflations. After the burst of the internet bubble in 2000, interest rates were lowered to stimulate the economy. For several years, the housing market, stock market and commodity market rose continuously, which seemed to suggest that the low interest rate policy worked well in stimulating the economy. However, the rise of commodity prices became unstoppable. Eventually, high commodity prices and the subprime mortgage crisis, both a result of the low interest rate policy, generated the biggest recession since the Great Depression.

Balancing the need for economic growth and a check on inflation has always been a problem in implementing interest rate policies. In the past, governments mostly focused on the goal of economic growth by keeping the interest rate low. Low interest rate environments also generate high rates of inflation in stock and housing prices. But inflation in stock and housing prices is often given a positive interpretation. Inflation in the price of commodities, the raw material of all economic activity, had been kept low for a long time by a steady increase in commodity output. However, with the increasing physical cost of extracting many commodities, it becomes increasingly difficult to keep inflation in commodity prices low while simultaneously demanding a high rate of economic growth.

Low interest rate environments reduce financing cost to build up specialized production systems. It also reduces financing cost in purchasing generic scarce commodities. The scarce commodities could be raw materials, real estate in prime locations, or shares in companies with monopoly powers. Investment in production systems needs highly specialized skills and is very time consuming. The trading of production systems is highly illiquid. The markets for trading of generic commodities are highly liquid. When resources are abundant, there is little inflation pressure on the resource. To make a profit, companies have to make long term investments in manufacturing. That was what happened to the world economy over most of the last several centuries. However, with the population growth and depletion of resources, many resources, such as land in prime locations, mineral reserves and agricultural products, have been under increasing inflation pressure. High inflation pressure in commodities increases investment opportunities in commodities and reduces investment opportunities in manufacturing, which relies on commodities as inputs. As a result, the investment capital has been migrating from specialized and illiquid manufacturing into generic and liquid commodities, real estate and financial assets. Overall, in an age of scarcity of commodities, low interest rate policy will be less and less able to stimulate economic activities and more and more able to stimulate speculative activities, generating waves after waves of economic instabilities.

Discount rate is a reflection of risk. In economic downturns, lenders naturally raise interest rates to compensate for higher risk. However, in economic downturns, central banks often lower interest rate to stimulate economy. Do not central banks and governments, or the general public have to bear the same risks? They do! In response to the bursting of the stock market bubble in early 2000, the central banks lowered interest rates to stimulate the economy. The low interest rate environment inevitably stimulated speculation, especially highly leveraged speculation, which had a higher upside potential. The massive bailout of the major banks and other financial institutions indicated that governments and the general public, like other institutions or individuals, have to bear the same risks for low interest rate loans or guarantees. In the past, long term risk inherent in low interest policy was small for many countries because of the expected long term economic growth fueled by cheap resources. However, with increasing scarcity of resources, the prospect of long term economic growth in many countries is dim. Risk associated with the low interest rate policy will increase while the stimulus impact will decrease.

Under a low interest rate environment, financial institutions, which incur low borrowing costs, will benefit. Depositors, who receive low interest incomes, are harmed. The depositors represent the general public and the financial institutions represent a small minority of people. Why do low interest rate policies prevail for such a long time? This is because in a large and expanding economy, high fixed cost systems provide a higher return than lower fixed cost systems. A low interest rate environment makes high fixed cost systems more competitive. Successful companies create large number of high paying jobs. Part of the large amount of profits brought in by the high fixed cost large firms becomes tax revenues and is redistributed to the general public. Overall, in a large and expanding economy, a low interest rate policy benefits the general public despite their loss as depositors.

With a steady increase in the cost of extracting natural resources, however, steady economic growth has become increasingly difficult. In such economic conditions, the divergence of interest between financial institutions and the general public becomes a very important issue for social stability. In a no growth and declining economy, the problem of income distribution will become a very important political issue.

With a simple analytical theory, we show that interest rate levels impact not only the amount of economic output, but also the structure of economic system. As a result, interest rate levels impact not only short term economic activities, but also long term economic activities. With a simple analytical theory, we show very clearly about the relation between interest rate and spending. While each person’s spending pattern differs, on average, there is a strong relation between interest rate and spending. Instead of demanding ordinary people stop borrowing, the central banks should stop the low interest rate policy. In a world of increasing cost of extracting natural resources, the continuation of low interest rate policy will generate wide gyration of social systems that we have witnessed in recent years. We are not proposing that central banks actively engage in a policy to maneuver interest rates. Instead, if the governments and central banks gradually reduce the magnitude of guaranteeing banks’ assets and bailing out financial institutions, interest rate will rise to reflect the risk involved in financial transactions. Individuals and organizations borrow less when interest rate is high. This will reduce the occurrence and magnitude of financial crisis.

Hyman Minsky one said, “Stability is destabilizing”. What does it mean exactly? With an analytical theory, we can obtain a very clear understanding from an example similar to the one at the beginning of this section. Suppose in two countries A and B, annual consumption of total product is 1 billion dollars. Suppose interest rate is 5% per annum in both countries. Uncertainty rate is 30% per annum in country A and 55% per annum in country B. Decision makers attempt to maximize the net present value of investment project. How much will be the desired fixed costs and how long will be the expected project durations? What are the net present values of projects in countries A and B?

We attempt to maximize (3.13), which is net present value of an investment by changing fixed cost, *K* and duration, *T* when uncertainty rates are set at 30% and 55% per annum respectively. The following table lists the calculated results.

|  |  |  |  |
| --- | --- | --- | --- |
| Uncertainty rate | Fixed cost | Duration | NPV |
| 30% | 5.8 | 25.3 | 8.5 |
| 55% | 2.1 | 12.1 | 2.3 |

From the table, we find that when uncertainty rate is lower, the amount of fixed investment is larger, investment duration is longer and the net present value is higher. So investors normally prefer low uncertainty rate environment. However, the net present values are only expected returns calculated at the beginning of a project. The actual returns depend on future environment. Suppose after the projects are built, the actual level of uncertainty becomes 80% per annum in both countries due to reasons unforeseen by decision makers, such as a global financial crisis. We recalculate the net present values from (3.13) to find the net present value of the first project, built in the low uncertainty rate environment of 30%, is -4.4 billion dollars while the net present value of the second project, built in the high uncertainty rate environment of 55%, is 0.0 billion dollars. The first project suffers heavy losses while the second project barely breaks even. When environmental conditions change dramatically, values of investment in formerly stable environment experience large fluctuations. In other words, “Stability is destabilizing”.

**3.5. Equilibrium and non-equilibrium theory**

Theories and terminologies are invented by humans to describe our world. In turn, they greatly influence our thinking.

General equilibrium theory forms the foundation of the established economic theories. From general equilibrium theory, wealthy countries have reached the desired optimal equilibrium state. So wealthy countries are called developed countries in economic literature, as well as in popular writings. Poor countries are certainly not in the desired optimal equilibrium state. But from the equilibrium theory, all economic systems, if allowed to operate in free market conditions, will always reach optimal equilibrium state. Hence poor countries are called developing countries.

Since wealthy countries are called developed countries, it is hard for most mainstream economists to imagine, let alone predict, that wealthy countries will fall into deep recessions. When deep recessions do occur in wealthy countries, given time, economy will inevitably return to the optimal growth path of the equilibrium state. Since poor countries are called developing countries, their conditions are expected to improve over time. When improvements don't occur, it must due to institutional deficiencies that distort market functions. Overall, global conditions will improve indefinitely, with only occasional temporary setbacks. Indeed, our time has been called the age of “Great Moderation”. This is why our societies were unprepared for the 2007, 2008 financial crisis and are unprepared for future crises, judging from the policies we adopted.

From a non-equilibrium theory, wealthy countries, being further away from equilibrium state than poor countries, consume more resources to sustain themselves. This is consistent with empirical observation. When resources are abundant, wealthy countries, which have developed technologies and institutional structures to utilize large amount of resources, especially energy resources, are more powerful than poor countries. Wealthy countries, whose high fixed cost structures require stable environment, devote more resources to maintain stability. As a result, wealthy countries are more stable than poor countries most of the time. However, wealthy countries, due to their higher fixed cost, could suffer more than poor countries when they could not develop effective mechanisms to deal with specific instability. Poor countries have less technology to utilize large amount of energy and other resources to work for them. So they are less powerful than the wealthy countries most of the time. But simpler technologies also require less resource to maintain them. Poor countries can endure resource scarcity and unpredicted uncertainty better than wealthy countries.

Biologists have pondered similar questions.

The constant conditions which are maintained in the body might be termed equilibria. That word, however, has come to have fairly exact meaning as applied to relatively simple physic-chemical states, in closed systems, where known forces are balanced. The coordinated physiological processes which maintain most of the steady states in the organism are so complex and so peculiar to living beings --- involving, as they may, the brain and nerves, the heart, lungs, kidney and spleen, all working cooperatively --- that I have suggested a special designation for these states, homeostasis. The word does not imply something set and immobile, a stagnation. It means a condition --- a condition which may vary, but which is relatively constant.”(Cannon, 1932, p. 25)

Cannon considered and rejected the term equilibrium in describing steady states in the internal environment of human bodies. He was aware his thoughts “would be suggestive for other kinds of organizations --- even social and industrial”. Indeed, biologists have recognized non-equilibrium theories provide better descriptions of living systems than equilibrium theories since the time of Darwin. “In the old system, each species was imagined to have been created according to some ideal type. Variation was just so much noise superimposed on the ideal type. After Darwin, the variation itself was seen as real and important, while the notion of an ideal type was recognized as a useless abstraction” (Trivers, 1985, p. 22).

Today, the non-equilibrium theory of living systems has been widely accepted in biological studies and other scientific studies. However, non-equilibrium theories have neither been accepted nor actively discussed in social sciences. Probably, non-equilibrium theories could not satisfy some emotional needs for human beings. A non-equilibrium theory tells us that life with material abundance is difficult to attain, and if attained, difficult to maintain for a long period of time. But there is a natural longing for good life. Various theories are developed to satisfy this longing. Generally, these theories promise good life in the heaven or in next life, such as in religions, or good life in the future, such as in communism. In essence, they promise a great product without having to deliver it. Furthermore, by separating heaven and earth, religions can provide realistic and practical advice on our current life. This is the key to long term success for many religions. General equilibrium theory, as an adaptive product from an age of affluence, promises not only better life in the distant future, but also great life right now and near future. This is the attraction of general equilibrium theory and standard theories in social sciences, despite their inconsistency with basic scientific principles.

A dynamic and non-equilibrium theory provides more precise description of our societies than the equilibrium theories. But equilibrium theories often precede non-equilibrium theories due to their conceptual simplicity. Furthermore, there is no chasm between equilibrium theory and non-equilibrium theory. Our production theory was inspired by the option pricing theory, which was developed as a dynamic equilibrium theory.

* 1. **Physics, mathematics and predictability**

There have been a lot of criticisms about establishing economic theory on the foundation of physical principles. But both Jevons and Walras, the main founders of neoclassical economics, were trained as scientists. “Jevons did so many things that it is difficult to classify him by occupation. … from examination of his other works we are inclined to list him rather as a physicist who wrote extensively on economics.” (Jaynes, 2003, p. 316) Both Jevons and Walras attempted to establish economics on the foundation of physics. From *The Coal Question* and other works, it is clear that Jevons understood economics from the perspective of thermodynamics. However, the mathematical methods of non-equilibrium thermodynamics were not yet developed in Jevons and Walras’ time. Jevons himself made it very clear that "I believe that dynamical branches of the Science of Economy may remain to be developed, on the consideration of which I have not at all entered." (Jevons, 1871, p. vii) Our theory can be understood as mathematical formalization of Jevons and Walras’ vision.

Furthermore our body, our mind and our environment are physical systems governed by the same physical laws. Cars move much faster and carry much heavier loads than humans because cars are designed to consume much more energy than humans. Several years ago, it was mandated in some countries that cars should burn energy grown from the fields. Food prices soared as humans have to compete with cars for available crops. Riots erupted in many regions because of the high food prices. When we understand humans and human societies as both physical and economical systems, we will be able to assess the consequences of our actions more accurately. There will be much less “unintended consequences”.

This is a mathematical theory of human societies. Some people suggest that events in human societies are not conducted in controlled environment, such as laboratories. Mathematical theories, which have been successfully applied in natural sciences, cannot be applied with the same level of effectiveness in human societies. But planetary movements are not conducted in controlled environment. It was the attempt by Copernicus, Kepler, Newton and many other pioneers to describe planetary movements mathematically that modern science was born. Some people suggest that events are human societies are not repeatable. Mathematical theories, which are supposed to describe repeatable events, cannot be applied to prediction with high level of confidence. But mathematical theories of atomic bombs and hydrogen bombs were developed before even a single atomic bomb or hydrogen bomb was detonated. Indeed, the very designs of nuclear bombs were guided by the mathematical theories of nuclear bombs.

To me, good mathematical theories are low cost experimental science. An actual global financial crisis cost trillions of dollars. But with a mathematical theory built on a solid foundation, it takes only several hours of calculation to understand the links between the types of policies and the magnitude of business cycles and financial crisis. However, the policies that generate economic instability over long term often stimulate economic activities over short term. A mathematical theory describing economic activities over time may make people more aware of the long term consequences of specific policies.

The leading economists often claim that nobody could have predicted the coming of financial crisis every time a financial crisis strikes. In times of financial crisis and large monetary losses incurred by financial institutions, it is often said that social science is not an exact science like engineering. So we will look at an example of engineering. If you climb up a play set, you may find signs like “Maximum capacity: 150 lbs”. Does that mean if you are 151 lbs, the play set will collapse? From our own experience, we know the play set will not collapse. In general,

When an engineer estimates the weights which a bridge or beam must support, or the pressures to which a boiler will be subjected, he does not provide merely for those stresses in building the structure. The engineer multiplies his estimates by three, six or even by twenty, in order to make the structure thoroughly reliable. The greater strength of the material, above that calculated as necessary, measures what is known as a “factor of safety.”(Cannon, 1932, p. 231)

By regulation, financial institutions are required to have their capital level above capital adequacy ratios. By law, many governments are required to restrict spending or deficit at certain percentages of income or GDP. In most cases, financial institutions and governments would keep their financial ratios barely above regulatory or legal requirements. Rarely any financial institutions or governments would strengthen their financial ratio “by three, six or even by twenty”, in order to make their financial structure thoroughly reliable. It is the low level of “factor of safety” in financial institutions and governments, and not that social sciences are not exact that causes many financial crises and large monetary losses.

If the factor of safety in financial system is increased to the level of that in engineering system, the financial system will be as safe as engineering system. At the same time, the income levels of top earners in financial firms will be reduced to that of top earners in engineering firms. This is the most important reason why the factor of safety in financial system is low. In engineering system, the payoff for reducing the factor of safety is low. But the punishment for engineering accidents is high. People responsible for injuries or deaths from low quality engineering projects will face legal responsibilities. By contrast, the payoff from reducing the factor of safety in financial system is high. But there is virtually no punishment for causing financial crises. Despite the systematic frauds in financial institutions that caused the 2007, 2008 financial crisis, not a single person was legally responsible for the frauds. There has been a great shift of attitude towards financial frauds. Many people got jail terms for savings and loans frauds in 1980s, which were at much smaller scale than the frauds that led to the 2007, 2008 crisis.

There are many ways to increase the “factor of safety” or lower the leverage of our financial system. A natural method is to allow the interest rate to reflect the actual market risk. Currently, by implicitly and explicitly guarantee financial assets, governments lower the risks of financial transactions by transferring their risk to the government and hence to the general public. This produces the low interest environment, which encourages borrowing. Policymakers are aware that people will borrow heavily in a low interest rate environment. They often warn people against heavy borrowing. Yet they keep the interest rate low to stimulate the depressed economy. Suppose policymakers often warn the public against drug use. Yet they keep the drugs at very low cost to stimulate the demand from the depressed people. Do you think policymakers should bear any responsibility? Macroeconomics and microeconomics are generally taught in different courses. However, micro behaviors are influenced by macro policies. By lowering the interest rate to encourage heavy borrowing and at the same time warning public against heavy borrowing, policymakers hope to receive credit for short term economic recovery and assign blame to the general public for the long term economic turmoil. This is the heart of the problem in economic policy making.

By keeping social sciences “inexact”, many people can take advantage and avoid blames of the murky situation. But some prominent economists think that economic theories are already too mathematical. Paul Krugman claimed that economists mistook beauty for truth. Then he managed to dig up two examples of the supposed beauty. One was CAPM theory, a beauty half century old. From his description of CAPM, he apparently forgot what CAPM was. Even Krugman, renowned for his mathematical prowess among economists, could not properly name two mathematical theories in economics. Indeed, he didn’t find any beauty in mainstream economic literature. All he found were “impressive-looking mathematics” “gussied up with fancy equations”, which are generously applied in research papers as cosmetics, to cover up their lack of substance. Statistics shows that the influence of theoretical works has dropped sharply in the last several decades. But the deterioration of mathematical standard is not limited to economic theory. From college admission to textbooks in many areas of science, the standard of mathematics has been declining over time. Mathematics is abandoned because it provides objectivity and clarity. Mathematics becomes a politically convenient target because there are few, instead of many, mathematical thinkers in social science.

**3.7. Concluding remarks**

The mathematical theory of social and biological systems is derived from basic economic and physical principles. It provides simple quantitative simulations of many important relations in biological systems and in our society. It can be applied to very broad areas. Since the theory was developed, it has been applied to project investment, corporate finance, trade and migration, resource and social structures, language and cultures, evolutionary and institutional economics, fiscal and monetary policies, firm size and competitions, software development economics and many other problems.

As an application of the theory, we use it to study monetary policies and business cycles. We showed that when interest rate is low, investment with high fixed cost and long duration will generate high expected profit. At the same time, high fixed cost, long duration investments are more sensitive to the increase of uncertainties. So the low interest rate policy has the potential to stimulate economic growth and to increase amplitude of business cycles. In an environment of abundant resources, the stimulation impact is strong and the destabilization impact is weak. In an environment of scarce resources, the stimulation impact is weak and the destabilization impact is strong. Our analysis provides a clear understanding how policies and social environment influence not only short term economic activities but also structures of our society and long term economic activities. When we evaluate policies and social structures, we need to consider not only short term performances but also long term consequences.

**Chapter Four: Languages and Cultures: An Economic Analysis**

### 4.1. Introduction

#### Economic theories usually are applied to understand concrete social systems, such as businesses and nations. But they also can be applied to understand “soft” social systems, such as languages and cultures. Languages and cultures are mediums of transmitting information. Their value rests on their ability to lower variable costs in communication. Lower variable cost systems in general entail higher fixed costs. Each culture or language has evolved to adapt to the local environment. As environment changes, however, different cultural systems may fare differently under new conditions. We apply the economic theory developed in the last chapter to study the co-evolution of languages, cultures and social systems. Many problems about the evolution and diffusion of cultures and languages and how they co-evolve with the social and economic systems can be understood in a very consistent way.

There is a long debate about the impact of differences in languages on human societies. (Whorf, 1956; Pinker, 1994; Devitt and Sterelny, 1999) The ability to acquire languages is innate and universal (Chomsky, 1988). Written languages, however, appeared very late in the history of human evolution and independently originated only in very few places (Diamond, 1997). Our discussion will be confined to how the differences in written languages are related to other aspects of human societies.

Words in English, an alphabetic language, are linear combinations of twenty-six letters. Characters in Chinese, a logographic language, are two-dimensional pictures. Several thousand of these complex characters need to be memorised before one can read articles reasonably well. Therefore Chinese is much more difficult to learn than English (Hanley, Tzeng and Huang, 1999). For the same reason, Chinese is more spatially compact and visually distinct than English. A Chinese document is much shorter than an English document of the same content. The speed of reading in Chinese is higher than in English (Lu and Zhang, 1999). From an economic point of view, the fixed cost (learning a language) of Chinese, a logographic language, is high and the variable cost (using a language) of Chinese is low. The opposite is true for English, an alphabetic language. Its fixed cost is low and its variable cost is high.

The classification of cultures by context is analogous to the classification of languages by fixed cost. “In general, high context communication, in contrast to low context, is economical, fast, efficient and satisfying; however, time must be devoted to programming.” (Hall, 1977, p. 88) This means that high context cultures are of high fixed costs and low variable costs. The opposite is true for low context cultures. The lowest-context cultures are probably Northern European and North American cultures, while “China, Japan, and Korea are extremely high-context cultures.” (Anderson, 2000, p. 266, 267) It is easy to note the link between the context of a culture and the fixed cost of learning a written language. Chinese language is a logographic language. Japanese and Korean languages are mixture of alphabetic and logographic languages. The logographic composition of these languages makes them more difficult to learn than alphabetic languages. Since learning and using written languages are such important parts of our lives, and since written languages are relatively stable and written records accumulate over time, they have a strong influence on many aspects of cultures.

Cultures and languages are mediums of transmitting information, which is the reduction of entropy (Shannon, 1948; Bennett, 1988). The value of cultures (and languages) rests on their ability to lower variable costs in communication. Lower variable cost systems entail higher fixed costs. In a homogeneous and densely populated society, people share common background and engage in communications frequently. It pays to spend more time to build up the context of the culture to reduce the variable cost of communication. In a sparsely populated society or a society with members of diverse background, it is more economical to keep the fixed cost of a culture low. Each culture has evolved to adapt to the local environment. As environment changes, however, different cultural systems may fare differently under the new conditions.

The performance of an economic system with respect to its fixed cost and variable cost can be studied with the analytical framework presented in the last chapter. From this framework, it can be derived that as fixed costs increase, variable costs decrease rapidly in a low uncertainty environment and decrease slowly in a high uncertainty environment. The main insight from this theory is the trade-off between efficiency of high fixed cost systems in a stable environment and flexibility of low fixed cost systems in a fast changing environment. One of the major purposes of the researches on culture is to find out the relation between culture factors and economic development. However, the mixed statistical results often puzzle cultural researchers. One type of culture that is linked to high economic growth in one period often performs badly in another period (Hofstede, 1980). With this analytic framework, it can be analyzed in a straightforward way. Lower context cultures have advantages in fast changing environments and higher context cultures have advantage in stable environments.

Cultures are often called multidimensional phenomena. Some of the primary dimensions of cultures are context, individuality, power distance and uncertainty avoidance (Hall, 1977; Hofstede, 1980). From our analytic framework, it can be derived that all these dimensions are linked to one single factor, the fixed cost or context of a culture. Although cultures are often expressed in colorful ways, at core, they display highly consistent patterns. This is not surprising, for the function of culture is the same across different cultures, to reduce the cost of communication.

This analytical framework helps understand some of the long standing puzzles in linguistics and many problems in the evolution of language and societies. Why the ancient Egyptian language didn’t take the “natural” step to evolve into an alphabetic language? (Diamond, 1997, p. 226) Was the logographic Chinese writing created independently or diffused from somewhere else? (Diamond, 1997, p. 231) Why are there so many low fixed cost alphabetic languages, but so few high fixed cost logographic languages, although all the earliest written languages were logographic? Is it a coincidence that Chinese, a logographic language, has the most native users? Why are most of the original Chinese characters, used three thousand years ago, still used today, with many of these characters having retained their original meaning, whilst over the same period of time, most alphabetic languages have changed considerably? Why was China the wealthiest country in the world during the long period of the stable agricultural society? Why China could not initiate industrial revolution despite its immense wealth? Why do democratic systems have a long tradition in the environment where alphabetic languages are used, while they rarely developed in regions that use logographic languages?

Many of these questions have been answered by many people in many different ways with various level of confidence. However, this theory will answer all these and other questions in a very simple and consistent way. The simplicity of the answers is not accidental. This is because the language, cultural and economic activities, which are thermodynamic processes, are directly modeled with an analytical thermodynamic theory. For any language, cultural or economic system, if it can help its hosts to extract more low entropy resources from the environment than the amount that dissipates, it will expand. Otherwise, it will contract.

This chapter is organised as follows. In Section 4.2, we apply the economic theory to give a unifying analysis of different dimensions of cultures. In Section 4.3, we make a detailed analysis of the evolution of written languages and how they affect cultural and economic development. Section 4.4 concludes.

**4.2. An application to cultural analysis**

The value of the fixed asset rests on its ability to reduce the variable cost in production or information transmission. In general, the variable costs decrease when the fixed costs are increased. However, the rate of decrease is a function of uncertainty. As the fixed cost increases, the variable costs decrease rapidly in a low uncertainty environment and decrease slowly in a high uncertainty environment (Figure 3.1). So the payoff of building up the context of a culture is high in a stable environment. In a fast changing environment, low context culture, being more flexible and innovative, have more advantages.

Suppose the fixed cost of understanding a culture is *K*, the variable cost is *C* from (5). The volume of information transmission is *Q*. Assume the value of each transmission is *S*. Then the total cost of information transmission is

while the total value of the information transmitted is *QS*. The rate of return of the information transmission is

Figure 3.2 is the graphic representation of the above formula for different levels of the fixed costs or the contexts of cultures. Two properties can be observed from Figure 3.2. First, it takes higher volume of communication for the high context culture to break even. So it is more economical to keep the fixed cost of a culture low in a sparsely populated society. Second, higher context cultures have lower variable costs. In a densely populated society, the volume of communication is high. The return of a high context culture is higher than a low context culture.

Cultures are often called multidimensional phenomena. Some of the primary dimensions of culture are context, individuality, power distance and uncertainty avoidance (Hall, 1977; Hofstede, 1980). We will discuss the relationships between these dimensions with the above framework.

“In general, high context communication, in contrast to low context, is economical, fast, efficient and satisfying; however, time must be devoted to programming.” (Hall, 1977, p. 88) This means that high context cultures share large amount of common fixed assets among their members. Because of this, members in a high context culture will value collectivism more than those in a low context culture, who will value individualism more. From Figure 2, the return curve is steeper in a high context culture than in a low context culture. Power and wealth is more unevenly distributed in a high context culture than in a low context culture, which means high context cultures have higher power distances. From Figure 1, high fixed assets systems perform better in low uncertainty environment. So high context cultures entail a higher level of uncertainty avoidance than low context cultures.

Hofstede (1980) and Hall’s (1977) works indicated that the values of these four dimensions of culture variation are positively correlated. Their correlation may even be higher than the statistical results because some formulas that calculated the indices may not adequately represent the defined meanings. For example, Singaporeans are generally considered very cautious. However, Singapore scored lowest on the Uncertainty Avoidance Index (Hofstede, 1980, p. 165). This is because this index is derived partly from employment stability and Singapore has a high job turnover rate. In a small and densely populated city state such as Singapore, changing jobs cause very little uncertainty. First, changing jobs rarely requires changing homes. Second, there are many firms crowded in a small place. It is often easy to find similar jobs nearby.

This analytic framework also makes it easy to analyze the relation between cultural factors and economic growth. For example, Hofstede (1980, p. 205) found that the Uncertainty Avoidance Index is negatively correlated with economic growth in the volatile period from 1925 to 1950 while it is positively correlated with economic growth in the stable period from 1960 to 1970. From Figure 1, a system high in fixed assets, (and hence a high need for uncertainty avoidance) performs well in a stable environment and performs poorly in a volatile environment.

Most cultural and economic indicators are highly correlated with latitude (Hall, 1977; Hofstede, 1980; Parker, 2000). High latitude areas receive less solar energy. So bio-densities in general and human population densities in particular are lower in the cold high latitude areas than in low latitude areas, where there is abundant solar energy. In high density areas, people interact more. Therefore it is more efficient to develop high fixed cost communication systems that have lower variable costs. People in the warmer low latitude areas, such as Southern Europe, generally develop high context cultures. In low density areas, people interact less. It is more economical to keep the fixed cost associated with communication small. People in the cold high latitude areas, such as Northern Europe, generally develop low context cultures.

## The evolutionary patterns of languages

It is more intuitive to map what we see into pictures than into alphabets. All the independently created written systems, from Sumer, China and Mesoamerica, are logographic languages. “The first Sumerian writing signs were recognizable pictures of the object referred to. … The earliest Sumerian writing consisted of nonphonetic logograms.” Gradually, phonetic representation was introduced to write an abstract noun “by means of a sign for a depictable noun that had the same phonetic pronunciation”. (Diamond, 1997, P. 220)

The first solution to a problem is often very complex. When a problem is known to be solvable, later solutions tend to be much simpler. As a writing system diffused to other regions, many logograms were gradually simplified to easy to write alphabets, especially the abstract nouns and grammatical items that are best represented phonetically instead of logographically. The earliest alphabets can be tracked to ancient Egyptian languages, which is probably influenced by the Sumerian language. But they kept the logograms in their language. This has puzzled some people. “The Egyptians never took the logical (to us) next step of discarding all their logograms …” (Diamond, 1997, p. 226) However, if we look at the patterns of innovation, the puzzle can be easily resolved.

Logograms, which are difficult to learn, contain high information content. They are of high fixed cost and low variable cost. To the people who already learned the logographic languages and hence already invested on the fixed cost, there is little incentive to convert the logograms into alphabets. It is for the similar reason that Japanese language, which is influenced by logographic Chinese language, retain many of the logograms. Only when the writing systems diffused further into new environments, the low fixed cost alphabetic system were gradually established. Semites familiar with Egyptian languages discarded all logograms and transformed the language into a pure alphabetical language. (Diamond, 1997, p. 226-227) This pattern is very similar to the innovation of new products. Many of the new ideas are initiated inside established large companies. However, it is often the new and small companies that implement the novel ideas.

The spread of writing systems is also an evolutionary process from logographic to alphabetic systems. This observation helps us resolves the puzzle whether the logographic Chinese writing was created independently or was diffused from Sumerian writing. (Diamond, 1997, p. 231) If the Chinese writing was diffused from Sumer, it would have been evolved into an alphabetical language over such a long distance and over such a long time.

Next we will apply this analytical framework to understand several properties of languages. First, since alphabetic languages are of low fixed costs, they are very easy to spread out. That is why there are many more people using alphabetic languages instead of logographic languages. Because of the simple structure of alphabetic languages, it is easy for them to mutate into new forms to adapt to local dialects. That is why there are many more alphabetic languages than logographic languages and most of the alphabetic languages have a relatively small number of users. Because of the high fixed cost, a logographic language is difficult to get established and to sustain itself. That is why there are few logographic languages left today although all the earliest written systems were logographic.

Why has the Chinese language survived while all other logographic languages were eventually replaced by alphabetic languages? “China’s long east-west rivers ... facilitated diffusion of crops and technology between the coast and inland, while its broad east-west expanse and relatively gentle terrain, ... facilitate north-south exchanges. All these geographic factors contributed to the early cultural and political unification of China.” (Diamond, 1997, p. 331) The same geographic factors also help the Chinese language to spread out quickly and gain large number of users. A logographic language has to be used by many people so that the low variable cost can offset its formidable high fixed cost. Chinese, a logographic language, has more than one billion native users, which is the highest among all languages.

Second, alphabetic languages, with low fixed costs and simple structures, can create new words easily and absorb words directly from other languages. Logographic languages, with high fixed costs, are more conservative and change less. Most of the characters Chinese used three thousand years ago are still used today, and many of these characters retain their original meanings. Over the same period of time, most alphabetic languages have changed considerably. New words in Chinese are formed by new combinations of existing characters. Since each Chinese character carries distinct meanings that are very stable, it is often difficult to create proper words to represent really novel ideas, which delays the understanding and adoption of new ideas. When the Japanese language absorbs a new word from other sources, the sound will first be represented by the alphabetic part of the language. If the word becomes popular, people will gradually create a logographic word with a distinct meaning to represent it for easier communication. (Harigaya, 2001)

Since societies using logographic languages are more stable, people there will value past experiences and are more conservative. Since societies using alphabetic languages change fast, a forward looking perspective is more valuable.

People using an alphabetic language spend less time learning the language, their common property. So they value individualism more. People using a logographic language spend far more time learning the language, their common property. So they value collectivism more. Hofstede (1980) found that those Chinese-majority regions “Taiwan, Hong Kong, and Singapore score considerably lower on individualism than the countries of the western world.” He attributed it to a particular Chinese philosophy (p. 215, p. 231).We find that language, by itself, offers a clearer explanation.

Third, economically, regions using logographic languages will do well in a stable environment because of the low variable cost in communication. During the stable agricultural society, which occupied most of the past two thousand years, China was the wealthiest country in the world. However, the high fixed cost Chinese language makes it difficult to initiate new changes. Instead, changes are initiated in the regions of diversified alphabetic languages. An alphabetic language, with low fixed costs, is more flexible and more innovative, which enjoys advantages in a fast changing environment. Regions using alphabetic languages have been leading the innovative changes since the beginning of industrial revolution several hundred years ago.

The difference in performance is not only reflected in world history but also in industries at different stages of maturity. Regions using alphabetic languages, such as USA, are dominant in new technologies. Regions using a mixture of alphabetic and logographic languages, such as Japan, absorb and perfect the advanced technologies. China, the region using logographic languages, is the main manufacturer in the mature industries.

Fourth, in a logographic language environment, if one manages to learn this language, one can acquire a huge amount of information at high speed. But the process of learning this language is difficult. Those who fail to master it may be left behind. The difference in payoff is drastic. In an alphabetic language environment, the result is not that drastic because of the low fixed cost, or low barriers of entry (Figure 3.2). So one would expect the logographic language environment to be more elitism oriented and the alphabetic language environment to be more pluralism oriented. Indeed, democratic institutions have a long tradition in the environment where alphabetic languages are used, while they rarely developed in regions that use logographic languages.

## Concluding remarks

In this chapter, the analytical theory of economics is applied to languages and cultures to offer a unified understanding of the co-evolution of languages, cultures and social systems. Written languages and cultures, which are created by human beings, in turn, exert great influence on the path of the development of human societies.