The Entropy Theory of Mind and Behavioral Finance

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/

January, 2011

Abstract

Rapid accumulation of empirical studies in behavioral finance calls for a unified and consistent theoretical synthesis. Instead of building up a behavioral theory of economics directly, we present the entropy theory of mind, which is an economic theory of mind. Then we integrate the value and cost of information processing into the overall picture in economic decision making. The entropy theory of mind includes a theory of judgment, which provides a common framework to integrate behavioral and informational theories of investment. The theory of judgment provides a quantitative link between investors' judgment and their trading activities. It offers a simple and unified understanding of major patterns in market activities and investor behaviors. As an application, a simple mathematical model based on the entropy theory of mind is constructed to understand many empirical patterns related to the cycles of momentum and reversals in asset markets. During various phases of the cycles, trading volumes and trading behaviors of investors of different sizes often show distinct characteristics. It has been a long standing challenge to describe the multiple patterns simultaneously from a quantitative theory. In this paper, we show that the predictions derived from the model are consistent with the multiple empirical patterns of trading volumes and investor activities at the different phases of the cycle of momentum and reversal.

We thank helpful comments from participants from a behavioral finance course at UNBC.

Rapid accumulation of empirical studies in behavioral finance calls for a unified and consistent theoretical synthesis. At the same time, increasingly sophisticated techniques employed in empirical studies enable us to test the validity of propositions derived from different theories on the human mind. In this paper, we present an updated version of the recently developed entropy theory of the human mind. The theory shows that common psychological patterns are consequences of the constraints of the physical laws and the economic principle that for any biological systems to be viable, the benefits of any activities, including mental activities, have to be higher than their costs. It contains a mathematical theory on the value and bias of judgment, which provides a common framework to integrate behavioral and informational theories of investment. The theory of judgment provides a quantitative link between investors' judgment and their trading activities. As a result, it offers a simple and unified understanding of major patterns in market activities and investor behaviors.

Before we discuss the details of human mind, we may reflect on a simple and obvious fact: the sizes of our brains are much smaller than the world we try to comprehend. This fact alone determines that our brain can only store and process a tiny fraction of information that is available in the world. Our brains also work under a tight energy budget. The energy consumption of our brains is less than a typical light bulb that lights our rooms. For comparison, Google's search engines consume more electricity than a million typical households. Mind, as a product of biological evolution, is subject to the economic principle that its cost must be lower than its benefit. It would not be economical for the mind to develop capacities to detect less general activities. Indeed, human beings have only limited capacities to detect many frequently occurring events. Our eyes can detect only very narrow ranges of electromagnetic waves. We don't have sense organs to detect electric fields, while some fish do. Our sense of smell is highly degenerated. Since it is costly to develop and maintain information processing capacity, only the most frequently occurring events that are most relevant to our survival will be detected by our senses and processed by our mind. The constraints of size and energy consumption of human brain determine that we only have very limited attention capacity, which is a major assumption in behavioral finance theories.

Among many functions of mind, the most important one is to identify resources at low cost. While the forms of resources are diverse, most resources can be understood from a unifying principle. A system has a tendency to move from a less probable state to a more probable state. This tendency of directional movement is what drives, among other things, living organisms. Intuitively, resources are something that is of low probability, or scarce. The measure of probability of a system is called entropy in physics. In a formal language, systems move from low entropy state to high entropy state. This is the second law of thermodynamics, the most universal law of the nature. The second law is often understood from an equilibrium perspective, rendering entropy an image of waste and death. However, from the non-equilibrium perspective, the entropy flow, which is manifested as heat flow, light flow, electricity flow, water flow and many other forms, is the fountain of life. Since all living organisms need to tap into the entropy flow from the environment for survival, it is inevitable that the mind, including the human mind, is evolved to identify entropy as the most important concept. Most important valuations in life are centered on entropy and related concepts. When information, which we collect for our survival, is represented mathematically by the entropy function, major problems in information theory can be resolved very easily (Shannon, 1948).

The definitions of entropy by Boltzmann and Shannon are mathematically equivalent. However, Shannon's definition intuitively represents the cost and value of information processing. As a result, Shannon's theory has been applied to many different fields. Since human brain is a physical communication and decision system, it is natural that the entropy theory of information can be applied to understand human mind (Chen, 2003, 2005). In the following, we will discuss three properties of the entropy theory of mind that are most relevant to behavioral finance.

First, information is costly and information with higher value is in general more costly to obtain. From Maxwell's (1871) thought experiment on an intelligent demon, information with higher physical value will have higher physical cost. Furthermore, physical cost is highly correlated with economic cost **(**Georgescu-Roegen, 1971 and Chen, 2005). In engineering projects, the entropy function, which is the measure of value of information, is often used to estimate the costs of projects (Kullback, 1959). Hence information of high economic value is in general of high economic costs. This result helps understand the systematic differences in the trading patterns of large and small investors. Depending on the value of assets under management, different investors will choose different methods of information gathering with different costs. Large investors are willing to pay a high cost to collect and analyze fundamental information. Small investors will spend less cost or effort on information gathering and rely mainly on easy to understand low cost information such as coverage from popular media and technical signals. Empirical works confirm that institutional investors trade on fundamental information while individual investors trade on price trends and news (Cohen, Gompers and Vuolteenaho, 2002; Barber and Odean, 2008; Engelberg and Parsons, 2009).

The differences in information processing by large and small investors generate the differences in their trading behaviors. There is a time lag between firm activities, such as R&D and project construction, and profit realization. By engaging in costly research, large investors are in a better position to estimate the values of new projects before they turn profitable and are better at separating long term components from short term fluctuation in earning data. Small investors, lacking detailed information on firm activities, have to rely on realized earning figures to assess firm values or observe the stock price movement to infer the trading activities of the informed. Since the stock transactions by individual investors are often triggered by public media, they sometimes are highly correlated (Barber, Odean and Zhu, 2009b). On average, large investors buy at an earlier stage when stock prices are rising and sell at an earlier stage when stock prices are falling than the small investors (Hvidkjaer 2006; Chen, Moise and Zhao,2009). As a result, large investors as a group make money and small investors as a group lose money from their trading activities (Wermers, 2000; Barber and Odean, 2000; Cronqvist and Thaler, 2004). Chen, Jegadeesh and Wermers (2000) documented that shares bought by mutual fund managers outperform shares they sold. Odean (1999) documented that the shares individual investors sold outperform the shares they bought. The heterogeneity of information processing and resulting trading activities by different investors is the main reason behind the observed patterns in the asset markets.

Second, the description of investor behaviors and market patterns can be refined by the theory of judgment, which is a natural extension from the information theory. Kelly(1956) developed the link between information investors received and their trading decisions. In most time, people have to make subjective assessment of events without possessing complete information. The theory of judgment provides a measure to value one's judgment. The valuation of a judgment is against a reference state, which is usually taken to be the maximum entropy equilibrium state (Jaynes, 1988). Since no additional information is required to determine the equilibrium state, the value of judgment from the decision making perspective can be naturally measured against the equilibrium state. However, the reference state can be a non-equilibrium steady state, such as a bubble state. Intuitively, if one buys a stock at two dollars and the equilibrium price is five dollars, then the value of your buying is three dollars. However, if the stock price can be momentarily moved to six dollars and you can take advantage of this high price, then the value of your buying is four dollars. Mathematically, the value of judgment is the average of profit or loss under different scenarios, which can be represented by a function generalized from relative entropy.

The value of judgment is always lower than or equal to the value of information with the same objective probability distribution and reference state. The value of judgment is equal to the value of information only when the subjective assessment of the probability distribution is identical to the objective probability distribution. Therefore, the concept of judgment is a generalization from the concept of information when a person does not have precise estimation of a random event, which is the case in most decision making processes. The difference between the values of judgment and information, or equivalently, the difference between actual cost of information processing and the lowest possible cost, is bias, which is defined by a mathematical function called relative entropy. Entropy and relative entropy are the two most important functions in information theory and statistical mechanics (Kullback, 1959; Schlögl, 1989; Cover and Thomas, 2005; Qian, 2001, 2009). Unlike the value of information, which is always positive, the value of judgment can be either positive or negative. This means that the value of active trading by investors can be either positive or negative. Trading that earn positive returns are generally attributed to information while trading that earn negative returns are generally attributed to behavioral biases. From the theory of judgment, the same judgment will have different values at different times due to changes of environmental parameters. Empirical evidences show that small individual investors often execute trades similar to those by large institutional investors but at a later stage. This could due to behavioral biases, or due to the difficulty of small investors to obtain timely information.

Under certain conditions, a judgment that is more biased may be more valuable than a less biased judgment. Suppose two investors are moderately favorable to two different stocks and each buys moderately amount of shares of respective stock. Subsequently, one stock performs very well and the other performs moderately well. Then the judgment of the investor who bought the stock that performs well is more biased. At the same time, the return from his investment is higher. This shows that value and bias of judgment are two distinct concepts. It will help clarify discussion in behavioral literature, which often identifies bias with low value of judgment. The theory of judgment bridges the chasm between the concept of information and cognitive bias. This will help provides a common framework for behavioral and informational perspectives in understanding financial market.

Investment decisions are made according to investors' judgment about returns of different assets. To establish a precise link between investors' judgment and investment return, we consider a simple market with only two assets: a risk free asset and a risky asset. Based on the subjective assessment of the return distribution of the risky asset, an investor can determine the optimal portion of the risky asset in the portfolio and calculate the expected rate of return of this portfolio. We prove that the first order approximation of the expected rate of return of the portfolios constructed from a judgment is equal to the value of the same judgment. Therefore, the theory of judgment provides a quantitative link between the value of a judgment and the expected rate of return of the portfolio constructed from the same judgment. In a broader sense, the theory of judgment provides a link between ideas and their monetary values.

Since the judgment about the performance of a stock determines the level of holding about the stock, the change of judgment about a stock determines the volume of trading in the market, which is considered as the key ingredient missing from the asset pricing models (Banerjee and Kremer, 2010). The theory of judgment provides a simple and intuitive tool to model trading volume in the asset market.

Third, the entropy theory of mind provides a simple mathematical model for a unified understanding of learning and human psychology. From the information theory, the cost of information processing depends on the relation between the structure of information sources and the structure of the coding system that transmit information. When the structure of coding system becomes more similar to the structure of the information sources that are to be transmitted, the average signal length will becomes lower. In other words, information processing is more efficient when the coding system represents the information sources more precisely. However, a more refined and specialized coding system performs poorly compared with a generic coding system when transmitting information without specific structures or with structures very different from the coding system. This tradeoff holds the key to understand human psychology and learning.

If certain events are common in the environment, it is economical to learn about them and represent them with shorter signals so mind can respond to them faster. When certain patterns persist for many generations, learning about these patterns is often transformed into more permanent structures in mind through epigenetic and genetic means so each generation does not have to relearn from scratch (Jablonka and Lamb, 2006; Rando and Verstrepen, 2007). These more permanent patterns of responses form the innate psychology. Learning and innate psychology complement each other. Learning is more costly but more flexible. Innate psychology is less costly but less flexible. Together, they provide us a coding system that lowers the average cost in information processing than an unstructured generic code in most situations that are important to us. This integrated understanding of learning and human psychology will help us understand many patterns reported in behavioral finance literature and their evolution over time. Human psychology and past learning determine that decisions by investors in particular moments may not be optimal, especially with the benefit of hindsight. Learning also determines that a particular bias, if discovered and economically significant enough, will gradually reduce due to adaptation and competition. However, the learning processes can be complex and prolonged. For example, trend following has been a popular trading strategy for a long time. But the research on momentum has kept uncovering new and sometimes surprising quantitative results (Novy-Marx, 2009; Chen, Moise and Zhao,2009). Furthermore, not all types of misevaluations of securities will decline overtime, since many misevaluations benefit major stakeholders who often are the best informed.

Many behavioral theories have been proposed to understand financial anomalies (Brav and Heaton, 2002). Instead of developing a behavioral theory of economics directly, we propose an economic theory of behavior. Then we integrate the value and cost of information processing into the overall picture in economic decision making. The entropy theory of mind has been applied to understand many empirical patterns in behavioral finance (Chen, 2003, 2004, 2007). In this paper, we will apply the theory to understand empirical patterns related to the cycles of momentum and reversal. A persistent pattern in the security market is the price continuation in short to medium run and the reversal of return in the long run (DeBondt and Thaler, 1985; Jegadeesh and Titman, 1993). Several models have been developed to explain this pattern (Barberis, Shleifer and Vishny,1998; Daniel, Hirshleifer and Subrahmanyam,1998; Hong and Stein,1999). However, these models could not explain other patterns related to the cycles of momentum and reversal (Lee and Swaminathan, 2000; Hvidkjaer, 2006). For example, the return patterns are often accompanied by distinct patterns of trading volume. However, "existing theories of investor behavior do not fully account for all of the evidence. … none of these models incorporate trading volume explicitly and, therefore, they cannot fully explain why trading volume is able to predict the magnitude and persistence of future price momentum." (Lee and Swaminathan, 2000, p. 2066)

The value of judgment provides a quantitative measure of the level of informedness of the investors. In our model, investors are classified into three groups as large, midsized and small investors according to their wealth. The proportion of people in each group is determined by the maximum entropy principle. The level of informedness of each group of investors are determined by or positively correlated to their level of the wealth. The judgments of the investors about the future movement of the prices of the risky assets determine the level of their holdings. During the cycles of the information processing, different investors, because of their different capacity in information processing, hold different portions of the risky asset. The changes of judgment about the risky asset by differently informed investors over time determine the trading behaviors of different groups of investors, the volume of trading and the direction and magnitude of asset price changes at different phases of cycles. The theoretical predictions generated by this quantitative model are very similar to empirical patterns recorded in Lee and Swaminathan, (2000), Hvidkjaer (2006) and others.

This paper is an update and extension from earlier works Chen (2003, 2004, 2005, 2007, 2008). The rest of the paper is organized as follows. Section 1 provides a reflection on the various aspects of the concept of entropy that are relevant to further discussion. Section 2 develops a general theory of innate psychological patterns and learning as means to reduce the cost of information processing. Section 3 presents the theory of judgment. Section 4 discusses how investors' judgments determine their trading decisions and the returns of their portfolios. Section 5 builds a quantitative model based on the theory of judgment and shows the predictions derived from the model are consistent with multiple empirical patterns related to the cycles of momentum and reversal. Section 6 concludes.

1. The Concept of Entropy: A Reflection

More than any other scientific theory, thermodynamics was evolved directly from the attempts to improve the efficiency in economic activities. Thermodynamics, and statistical mechanics, which is the micro foundation of thermodynamics, are the economic theory of nature and human societies. In the early days of research in thermodynamics, people focused on the quantities such as work and heat, which are intuitive to our senses. However, the amounts of work and heat transfer are path dependent, which makes it difficult to detect general patterns. In mid nineteenth century, Rudolf Clausius introduced the concept of entropy, which is a state variable and hence not path dependent. His work may remind us about path dependency in economics, which is often said to be the reason why economic activities can not be described by a universal theory. In fact, researchers in science are faced with the same problems. The challenge is to find state variables, which are path independent, so the efficiencies of different path dependent activities can be compared.

Shortly after the introduction of the concept of entropy, Boltzmann linked entropy, *S*, a macroscopic quantity, to *W*, the number of microscopic states with the formula

$$S = k \log W$$

We will derive the mathematical equivalence between the definitions of entropy by Boltzmann and Shannon, which will be important for us to understand the relation between information and physical entropy. Suppose a system has N particles, of which, n_i particles are in i^{th} state, i = 1, 2, …,k. Then

$$W = \frac{N!}{\prod n_i!}$$

where

$$N = \sum_i n_i$$

Taking the logarithm of W, we have

$$\log W = \log N! - \sum_i \log n_i!$$

Then using Stirling's theorem, we obtain

$$\log W = N \log N - N - \sum_i n_i \log n_i + \sum_i n_i = N \log N - \sum_i n_i \log n_i$$

$$= N(\log N - \sum_i \frac{n_i}{N} \log n_i)$$

$$= N(-\sum_i \frac{n_i}{N} \log \frac{n_i}{N})$$

Let

$$p_i = \frac{n_i}{N} \qquad \text{for } i = 1,2,\ldots k$$

Then

$$\log W = N(-\sum_i p_i \log p_i)$$

and

$$S = k \log W = kN(-\sum_i p_i \log p_i) \tag{1}$$

Since

$$-\sum_i p_i \log p_i$$

is defined as information by Shannon (1948), this establishes the mathematical relation between physical entropy and information. In the following, we will discuss an often cited passage from Arrow that has played a prominent role in defining the scope of application of entropy theory in the field of economic analysis.

> The well-known Shannon measure which has been so useful in communications engineering is not in general appropriate for economic analysis because it gives no weight to the value of the information. If beforehand a large manufacturer regards it as equally likely whether the price of his product will go up or down, then learning which is true conveys no more information, in the Shannon sense, than observing the toss of a fair coin. (Arrow, 1973 (1983), p. 138)

From (1), the physical and economic values of information are determined by three factors. The first factor is

$$-\sum_i p_i \log\, p_i$$

which is the Shannon information. It measures information value per symbol or per unit of output. The second factor is *N*. In physics, it represents the number of particles in a system. In information theory, it represents the number of symbols in information transmission. For example, in Shannon (1948), there is a formula

$$H' = mH$$

"where *m* is the average number of symbols produced per second. *H* or *H'* measures the amount of information generated by source per symbol or per second. (Shannon and Weaver, p. 23)" In economics, it represents the quantity of output. The third factor is *k*. It is a coefficient that converts value of one scale into another. In physics, it is the Boltzmann constant. In economics, it is the monetary value per unit of output. If the output is of high value, *k* is large. If the output is of low value, *k* is small.

It is not surprising that Arrow made such a simple mistake. Everyone makes simple mistakes. Many people must have spotted the error and some of them, such as Chen (2005), have pointed it out publicly. However, it is a telling sign that a statement containing such an obvious error has been cited or quoted as the final word on such important issue for so many years and anyone who points out the error is simply ignored.

Soon after Shannon developed the entropy theory of information, many attempts were made to give information more intuitive meaning. Kelly (1956) applied the information theory to investment decisions when investors have the knowledge of probability distribution of asset returns. He showed that the maximum exponential rate of return of an investor's asset is equal to the rate of information transmission. This result provides an economic meaning to the measure of information. Kelly suggested that the economic interpretation of information theory "can actually be used to analyze nearly any branch of human endeavor (Kelly, 1956, p. 918)". Mathematically, the entropy based criterion is equivalent to maximization of geometric mean or logarithmic utility in portfolio investment (Latane, 1959; McEnally, 1986). The theory is very intuitive and practical. The entropy based investment theory has been successfully applied to investment practice (Thorp, 1997; Poundstone, 2005). Poundstone (2005) provided a detailed account on the interaction between the information theory based investment theory and academic economists. "Calling the Kelly system a "fallacy," he (Samuelson) helped persuade most economists to reject it. (Poundstone, 2005)" As Samuelson (1969) offered no substantial argument against the theory of Kelly and others, it is difficult for many to understand why he took such a strong view about it. But Samuelson's position was consistent with his general attitude toward the use of entropy in the social sciences:

> And I may add that the sign of a crank or half-baked speculator in the social sciences is his search for something in the social system that corresponds to the physicist's notion of "entropy". (Samuelson, 1972, p. 450)

A distinct characteristic of the entropy based investment theory or geometric mean method is that it provides an objective criterion. In a scientific theory, an objective criterion is usually regarded as a good criterion. However, the conventional economic theories assume that human beings possess subjective utilities and subjective sensations, which make objective criterions less useful. So it will be helpful to investigate the concept of utility and sensation. Bernoulli's (1738(1954)) paper is regarded as the first paper on utility. The utility function derived from his paper is a logarithmic function. The relation between subjective sensation and objective stimulation has been extensively investigated by Fechner. "If Fechner's psychophysics could be said to have an inverse square law, it was the logarithmic law relating sensation and stimulus (*Reiz*):

$$S = C \log R$$

… Fechner's work was in some respects more closely related to Ohm's early investigations of current. Indeed, Ohm … had developed essentially the same empirical formula $V = m\log(1+x)$ for the relation between the loss of force of current V and length of wire x." (Stigler, 1986, p. 243)

In a section called *Can Perceptions Be Quantified?* Maor (1994) noted, "Among the many phenomena that follow a logarithmic scale, we should also mention the decibel scale of loudness, the brightness scale of stellar magnitude, and the Richie scale measuring intensity of earthquakes. (p. 113)" The specific results about the relation between sensation and stimulus can be complex (Stevens, 1961). However, empirical evidence overwhelmingly supports that subjective utilities and sensations are strongly related to objective stimulus and measurements. Furthermore, maximization of logarithmic utility or geometric mean is a dominant strategy under very general conditions (McEnally, 1986; Sinn, 2003). This is not to suggest that individual feelings and sensations are identical. The very opposite is true. We can discuss Newtonian mechanics as a parallel. All mechanical systems can be described by the basic laws of Newtonian mechanics. However, the movements of different mechanical systems can be very different and very complex. Because each person's neural system is unique by birth and uniquely calibrated by environmental stimulus, identical external stimulus may generate diverse sensations. What we suggest is that the entropy theory of mind provides a foundation to understand these complex sensations. In earlier works, we show that the entropy theory of mind provides a simple and consistent explanation to many psychological patterns (Chen, 2003, 2005). Theories of mind from similar perspectives have been developed by other researchers (Jaynes, 1988; Friston, 2010). This shows the broad appeal of understanding the mind from physical and economic perspectives.

It is interesting to note the relation between entropy and the research community. Soon after the introduction of the concept of entropy, researchers recognized entropy as a fundamental quantity. Therefore it seems natural for entropy to be used as a coordinate in understanding thermodynamic problems. However, quantities such as volume and pressure, which are more intuitive to understand, are used as coordinates most of the time. Gibbs thought:

> The method in which the co-ordinates represent volume and pressure has a certain advantage in the simple and elementary character of the notions upon which it is based … On the other hand, a method involving the notion of entropy, the very existence of which depends upon the second law of thermodynamics, will doubtless seem to many far-

> fetched, and may repel beginners as obscure and difficult of comprehension. This inconvenience is perhaps more than counterbalanced by the advantages of a method which makes the second law of thermodynamics so prominent, and gives it so clear and elementary an expression. (Gibbs, 1873a (1906), p. 11)

By using entropy as a coordinate, Gibbs was able to conceive the concept of free energy, one of the most important concepts in thermodynamics (Gibbs, 1873b (1906)). However, current textbooks on thermodynamics continue to use volume and pressure as coordinate systems instead of entropy. This is a common pattern in many research works. The concept of entropy is crucial in developing many important scientific ideas, such as Planck's quantum theory and Einstein's photon theory (Jammer, 1966; Einstein, 1905 (1989)). But in standard textbooks, the concept of entropy plays a minor role or is not mentioned when introducing these ideas. This is because entropy, which cannot be directly measured by our sensory systems, is difficult to communicate.

The above discussion shows the dual personality of entropy. It explains why the concept of entropy is behind so many scientific breakthroughs. It also explains why entropy is usually abandoned or its scope of application is restricted after the breakthroughs are established. The most prominent example is from Shannon himself. Shortly after Shannon (1948) developed the entropy theory of information, Weaver commented: "Thus when one meets the concept of entropy in communication theory, he has a right to be rather excited --- a right to suspect that one has hold of something that may turn out to be basic and important." (Shannon and Weaver, 1949, p. 13) This sense of excitement attracted a lot of attempts to apply the concept of entropy to many other areas. As it is often the case, earlier attempts to apply some promising intuition do not yield concrete results easily. In an editorial, Shannon tried to discourage the jumping on the bandwagon:

> Workers in other fields should realize that that the basic results of the subject are aimed at a very specific direction, a direction that is not necessarily relevant to such fields as psychology, economics, and other social sciences. Indeed, the hard core of information theory is essentially, a branch of mathematics, a strictly deductive system. (Shannon, 1956)

Since Shannon's statement, the entropy theory of information has been applied to many different fields. Many of these applications are now part of the standard textbooks in information theory, such as (Cover and Thomas, 2006). However, Shannon's statement is still been invoked frequently to discredit attempts to apply the entropy theory of information to new fields. The best way to honor Shannon is to repeal his statement, which has not stood the test of time, so his original ideas can be applied to broader fields.

2. Learning and Psychological Patterns as Means of Reducing the Cost of Information Processing

Researchers observe that the human mind does not process different signals at the same speed. "In general, instances of large classes are recalled better and faster than instances of less frequent classes; that likely occurrences are easier to imagine than unlikely ones; and that the associative connections between events are strengthened when the events frequently co-occur" (Tversky and Kahneman, 1974, p.1128). Similarly, in languages, not all words are of the same length. In general, more frequently used words are shorter than less frequently used words. For instance, in the sentence, "I eat sausage." the word "I" has only one letter and the word "sausage" has seven letters. This pattern develops because “I” is used much more frequently than “sausage”. When a concept becomes more frequently used, the word representing it often gets shortened. For example, automobile becomes car; bicycle becomes bike; television set becomes television and then simply TV; personal computer becomes PC. By representing high probability events with shorter expressions, we reduce the time and effort in information transmission. Therefore, language is not a purely random mapping from the concrete worlds to the abstract symbols. It is a highly structured coding system that reduces the average length of messages.

Since languages are highly structured coding systems, even after messages are enciphered, some hidden structures remain. It is through the detection of these hidden structures that people tried to decipher these messages. The modern information theory was largely born out of attempts to decipher encrypted messages (Beutelspacher, 1994; Boone, 2005). Many pioneers in information theory, such as Alan Turing, Claude Shannon and Solomon Kullback, were involved in war time effort in breaking enemy cryptosystems. It is interesting to note that many successful traders in the financial markets are originally trained in the filed of information theory (Patterson, 2010).

In the following, we will provide a more quantitative discussion on how coding systems affect the cost of information processing. We will begin with some general discussion and then present a numerical example.

Suppose a random variable, X, has n discrete states $\{x_1, x_2, ..., x_n\}$, with probability $\{p_1, ...p_n\}$. The lower bound for the cost of information transmission is the entropy of the random variable. Mathematically, the lower bound is

$$\sum_{j=1}^{n} p_j(-\ln p_j)$$

where $\{p_1, ...p_n\}$is the probability distribution of the random variable. The actual cost of information transmission by a coding system can be measured by

$$\sum_{j=1}^{n} p_j(-\ln q_j) \qquad (2)$$

where $\{p_1, ...p_n\}$is the probability distribution of the random variable and $\{q_1, ... q_n\}$is the subjective assessment of the probability distribution of the random variable when we design the coding system. Formula (2) may be called the generalized entropy function. It reaches minimum if and only if each

$$q_j = p_j, \qquad 1 \le j \le n$$

The purpose of encoding is to lower the cost of communication by accurately assessing the probability distribution of random variables. However, probability distributions of random variables may not be stationary. A coding system optimized from past data may not provide low cost communication in the future. To make the analysis more concrete, we will discuss a numerical example adapted from Shannon (1948). Suppose a random variable can generate four states, which are called state 1, 2, 3 and 4 respectively. A generic binary coding can be designed with the following mapping:

$$C(1) = 00$$
$$C(2) = 01$$
$$C(3) = 10$$
$$C(4) = 11$$

The coding has an average length per symbol of two. This corresponds to the maximum entropy of four random letters, which is

$$\frac{1}{4}\times(-\log\frac{1}{4})+\frac{1}{4}\times(-\log\frac{1}{4})+\frac{1}{4}\times(-\log\frac{1}{4})+\frac{1}{4}\times(-\log\frac{1}{4}) = 2$$

Assume that, through experience, we learned that the probabilities of four states are

$$P(x=1)=\frac{1}{2}$$
$$P(x=2)=\frac{1}{4}$$
$$P(x=3)=\frac{1}{8}$$
$$P(x=4)=\frac{1}{8}$$

The entropy of the random variable with the above probability distribution is

$$H(x)=\frac{1}{2}(-\log(\frac{1}{2}))+\frac{1}{4}(-\log(\frac{1}{4}))+\frac{1}{8}(-\log(\frac{1}{8}))+\frac{1}{8}(-\log(\frac{1}{8}))=1.75$$

From the information theory, the shortest possible average length of binary code per symbol can be as low as 1.75. Such a code exists with the following mapping.

$$C(1) = 0$$
$$C(2) = 10$$
$$C(3) = 110$$
$$C(4) = 111$$

We can confirm the average length of this coding by calculating

$$\frac{1}{2}\times1+\frac{1}{4}\times2+\frac{1}{8}\times3+\frac{1}{8}\times3=1.75$$

This shows that learning can help us reduce the cost of information processing. From the information theory, in a code with shortest average length, the code length of events with probability *p* should be close to –log*p*. In this case

$$-\log(\frac{1}{2}) = 1$$

$$-\log(\frac{1}{4}) = 2$$

$$-\log(\frac{1}{8}) = 3$$

The lengths of the codes designed are consistent with this criterion.

Now assume there is a fundamental change in the probability distribution of four states. The new probabilities become

$$P(x=1) = \frac{1}{8}$$

$$P(x=2) = \frac{1}{8}$$

$$P(x=3) = \frac{1}{4}$$

$$P(x=4) = \frac{1}{2}$$

If the coding system remains the same, the average length of code per symbol becomes

$$\frac{1}{8}\times 1 + \frac{1}{8}\times 2 + \frac{1}{4}\times 3 + \frac{1}{2}\times 3 = 2.625$$

which is longer than 2, the average length of generic code that can be designed without any knowledge about the probability distribution of the states. Therefore, knowledge gained from past experience may hinder instead of helping information processing when environment change substantially or when we attempt new tasks that are very different from earlier experiences.

The average length of code per symbol also can be calculated from the generalized entropy function as

$$\frac{1}{8}\times(-\log_2\frac{1}{2}) + \frac{1}{8}\times(-\log_2\frac{1}{4}) + \frac{1}{4}\times(-\log_2\frac{1}{8}) + \frac{1}{2}\times(-\log_2\frac{1}{8}) = 2.625$$

which is the same as calculating average length of coding directly.

The above discussion shows that the cost of information processing is low when a state with probability *p* is represented by a symbol of length about –log *p*. When *p* is large, -log *p* is small. This is consistent with the empirical regularity that more frequently occurred events are recalled faster than the less frequently occurred events. In particular, a coding system that is a good representative of the information source has a lower cost of information processing than a generic coding system which does not require specific knowledge of the information sources. According to Shannon, "The transducer which does the encoding should match the source to the channel in a statistical sense (Shannon and Weaver, p. 31)". Therefore it is economical to learn about the environment and develop coding systems accordingly. However, a more refined and specialized coding system performs poorly compared with a generic coding system when transmitting information without specific structures or with structures very different from the coding system. The choice of coding systems and their life spans depends on how persistent the environments are.

For short term patterns, we often depend on learning to calibrate our mind temporarily. If certain patterns are stable in the environment for many generations, knowledge about them is often transformed into more permanent structures in mind through epigenetic and genetic means so each generation does not have to relearn from scratch at an additional cost (Jablonka and Lamb, 2006; Rando and Verstrepen, 2007). There is no dichotomy between innate psychology and learning, or between nature and nurture. Instead, they form a continuous spectrum in phenotypic plasticity. In general, the problems that occur often in life are easier to learn than problems that occur less often; the psychological patterns that evolve earlier from our ancestors are stronger than the psychological patterns that evolve later.

"In general, ideal or near ideal encoding requires a long delay in the transmitter and receiver (Shannon and Weaver, p. 31)". The exact length of delay for ideal encoding depends on the persistence of statistical patterns. The requirement for delay in encoding presents a tradeoff between efficiency and flexibility. For coding to be efficient, a long delay is required to accurately measure the statistical distribution of random processes. However, random processes are not always stationary. The longer delay in encoding, the slower a system responds to structural changes in random processes. In statistics, this corresponds to the tradeoff between type I and type II errors. An attempt to reduce one type of error will increase another type of error. In human psychology, this corresponds to conservatism and overreaction. An attempt to reduce one type of bias will increase the likelihood of another bias. In light of this understanding, should we continue to call these psychological patterns biases?

This section discusses learning and human psychology as tools to reduce the cost of information processing. In the next section, we will measure the value and bias of our judgment, which we make based on the information we receive.

3. Value and Bias of Judgment

Suppose a random variable, X, has n discrete states $\{x_1, x_2, \ldots, x_n\}$, with probability $\{p_1, \ldots p_n\}$. The subjective judgment of a person may differ from the objective probability. Suppose the subjective judgment of the probability distribution is $\{q_1, \ldots q_n\}$, then the level of uncertainty of judgment on each q_i, is

$$H(q_i) = -\ln q_i \qquad \text{for} \quad 1 \le i \le n$$

The total uncertainty of judgment of a random event is the average of uncertainty of judgment of each state, weighted by the objective probability distribution of the random event.

$$\sum_{j=1}^{n} p_j(-\ln q_j)$$

We will compare the uncertainty of the judgment against that of a reference probability distribution of a reference state. The reference state can be the equilibrium state, or a non-equilibrium steady state. Take a simple example of binary states of up and down in the stock market. Let {p, 1-p} represent the probability of up and down of market in the next period. If on average, stocks are up 55% of the time and down 45% of the time, {0.55, 0.45} represent the equilibrium state and {0.3, 0.7} represent a non-equilibrium state, which has a higher probability to go down than to go up in the next period. Suppose the reference probability distribution of a random event is $\{r_1, \ldots r_n\}$. Then the total level of uncertainty of the reference state is

$$\sum_{j=1}^{n} p_j(-\ln r_j)$$

The value of judgment can be defined as the reduction of uncertainty from the reference state, which is

$$\mathrm{V(p,q,r)} = \sum_{j=1}^{n} p_j(-\ln r_j) - \sum_{j=1}^{n} p_j(-\ln q_j) \quad = \sum_{j=1}^{n} p_j(\ln \frac{q_j}{r_j}) \tag{3}$$

The right hand side of Formula (3) is a function generalized from relative entropy. We will call it generalized relative entropy. When each

$$q_j = p_j, \qquad\qquad 1 \le j \le n$$

The value of judgment becomes the value of information.

$$\sum_{j=1}^{n} p_j(\ln \frac{p_j}{r_j}) \tag{4}$$

From Gibbs inequality (Gibbs, 1902),

$$\sum_{j=1}^{n} p_j \ln(p_j) \ge \sum_{j=1}^{n} p_j \ln(q_j)$$

Therefore, the value of judgment is always less than or equal to the value of information with the same probability distribution and reference distribution.

In practice the reference probability distribution $\{r_1, \ldots r_n\}$ is often understood as the maximum entropy distribution under known constraints. When there is no known constraints, the maximum entropy distribution is {1/n, … 1/n} and Formula (3) becomes

$$V(p,q) = \sum_{j=1}^{n} p_j \ln q_j \ + \ln n \tag{5}$$

The value of judgment can be positive or negative. This means that active trading by investors can increase or decrease the value of their investment portfolios. The distance between the objective distribution and one's judgment is the measure of bias, which can be defined as

$$B(p,q) = \sum_{j=1}^{n} p_j \ln(p_j) - \sum_{j=1}^{n} p_j \ln(q_j) = \sum_{j=1}^{n} p_j \ln \frac{p_j}{q_j} \qquad (6)$$

which is the relative entropy function. It is always nonnegative and is zero if and only if each

$$q_j = p_j, \qquad 1 \le j \le n$$

In general, the bias will be smaller when q_j is closer to p_j.

The value of judgment, the value of information and the measure of bias are related by the following equation

$$\sum_{j=1}^{n} p_j (\ln \frac{q_j}{r_j}) = \sum_{j=1}^{n} p_j (\ln \frac{p_j}{r_j}) - \sum_{j=1}^{n} p_j \ln \frac{p_j}{q_j} \qquad (7)$$

This equation shows that the value of judgment is equal to the value of information minus the measure of bias.

We will perform some simple calculations to illustrate the properties of the value of judgment and the measure of bias. For the simplicity of exposition, we will only consider events with two possible outcomes, state 1 and state 2, in the remaining part of the paper. Intuitively, we can think state 1 as the up market and state 2 as the down market. First, we will assume the reference state is the equilibrium state, which is set to be {0.55, 0.45}. This means that it is more likely for the market to be up than down. We begin with the calculation of the value of a judgment that is the same as the equilibrium state. From (3), the value of the judgment that agrees with the equilibrium state is

$$p \ln \frac{0.55}{0.55} + (1-p) \ln \frac{0.45}{0.45} = 0$$

Hence the value of the judgment that agrees with the equilibrium state is zero, regardless of the actual probability distributions of the states. Intuitively speaking, an investor who agrees with the market does not believe he possesses valuable information and puts his money into an index fund.

Now consider two random events with different probability distributions. Assume in the first event, the objective probability of state 1 is 65% and the probability of state 2 is 35%. Someone estimates the probability of state 1 is 57.5% and the probability of state 2 is 42.5%. We assume the equilibrium state remains {0.55, 0.45}. From (3), the value of his judgment is

$$0.65 \ln \frac{0.575}{0.55} + 0.35 \ln \frac{0.425}{0.45} = 0.0089$$

From (6), the bias of this judgment is

$$0.65\ln\frac{0.65}{0.575}+0.35\ln\frac{0.35}{0.425}=0.0117$$

In the second event, the probability of state 1 is 57.5% and the probability of state 2 is 42.5%. Someone estimates the probability of state 1 is 57.5% and the probability of state 2 is 42.5%. From (3), the value of his judgment is

$$0.575\ln\frac{0.575}{0.55}+0.425\ln\frac{0.425}{0.45}=0.0013$$

while the bias of this judgment is zero. From the above calculation, the judgment that is more biased can be more valuable than a less biased judgment under certain conditions. Intuitively speaking, an investor who is moderately favorable to a stock which turns out to earn very high rate of return will perform better than an investor who is moderately favorable to a stock which turns out to earn moderately rate of return. By separating value and bias of judgment, we will be able to perform more precise analysis to investor behaviors, which we will discuss later.

Next, we will consider values of judgments when reference states are equilibrium and non-equilibrium states respectively. Intuitively, we are comparing values of investment decisions when market settles down in equilibrium state or moves into a bubble state. Suppose an investor spot a good stock with high growth potential. Assume the objective probability of this stock to move up and down is {0.6, 0.4}. The investor's own assessment of the stock is {0.575, 0.425}. We will calculate the value of his judgment if the stock settles into the equilibrium state of {0.55, 0.45} or a bubble state {0.40, 0.60}, which means that the stock will have 40% chance going up and 60% chance going down the next time period.

When the stock will settles into the equilibrium state, the value of judgment is

$$0.6\ln(\frac{0.575}{0.55})+0.4\ln(\frac{0.425}{0.45})=0.0038$$

When the stock will move into the bubble state, the value of the same judgment is

$$0.6\ln(\frac{0.575}{0.4})+0.4\ln(\frac{0.425}{0.6})=0.0798$$

The value of the judgment in a bubble state is much higher than the value of the same judgment in the equilibrium state. Intuitively, investors holding shares of a stock benefit from the high stock price.

There is much discussion about why misvaluation occurs and why arbitrage cannot eliminate misvaluation. Most of the discussion is based on the assumption of psychological biases of investors, especially small investors. Empirical evidences show that small investors are late stage momentum traders, following the trends generated by large, informed investors (Hvidkjaer 2006). The systematic biases generated by small investors are of small magnitude and are reversed over a very short period of time (Hvidkjaer 2008; Barber, Odean, and Zhu, 2009a, 2009b). However, many large scale profitable arbitrage activities, such as merger and acquisitions (Shleifer and

Vishny, 2003), require asset prices to be highly overvalued over extended periods. Most of the well informed about an asset are major stakeholders of the same asset. They often have an incentive as well as capacity to influence public opinion to move asset prices along certain direction. From the theory of judgment, the value of judgment and hence the value of investment depends on the reference state. Therefore the informed may choose to sell an overpriced asset to realize an arbitrage profit or push asset prices even higher to generate higher arbitrage profits. The above calculations indicate that it is much more valuable to generate asset bubbles than to realize profit immediately. The actual patterns of investment decisions by informed investors can be investigated in future empirical studies.

More systematic discussion on the properties of the theory of judgment can be found in Chen (2008). In the next section, we will discuss the link between investors' judgment, the expected rate of return of their portfolios and their trading decisions.

4. Value of Judgment, Expected Rate of Return of Portfolios and Investors' Trading Decisions

Investment decisions are made according to investors' judgment about stocks. To quantify the relation between investors' judgment and their trading decisions, we will consider a simple market with a risk free asset and a risky asset. The payoff of one unit risk free asset is 1+ r at the end of each time period. The payoffs of one unit risky asset can be either (1 + r)(1 + d) with probability p or (1+ r)(1– d) with probability 1-p. Investors can only assess the probabilities subjectively.

Investors aim at maximizing expected geometric return (Kelly, 1956; Latane and Tuttle, 1967). Based on the subjective assessment of the return distribution of the risky asset, an investor determines the optimal combination of the risk free asset and the risky asset in the portfolio. Then he can calculate the expected rate of return of this portfolio. Suppose an investor assesses the return distribution of the risky asset to be {q, 1-q}. Assume the portfolio he constructed contains a portion x of risky asset and the remaining portion of 1 – x is risk free asset. The expected geometric return of the portfolio over the risk free rate is

$$\begin{aligned}&((1-x)(1+r)+x(1+r)(1+d))^{q}((1-x)(1+r)+x(1+r)(1-d))^{1-q}-(1+r)\\&=((1+xd)^{q}(1-xd)^{1-q}-1\ )(1+\mathrm{r})\qquad(8)\end{aligned}$$

To determine the value of x at which the portfolio will have the maximal rate of return, we differentiate the above formula with respect to x.

$$\begin{aligned}&\frac{d}{dx}((1+xd)^{q}(1-xd)^{1-q}-1)\\&=d(1+xd)^{q-1}(1-xd)^{-q}(2q-1-xd)\end{aligned}$$

The above differentiation equals zero when

$$x=\frac{2q-1}{d}\qquad(9)$$

At this value of x, the portfolio obtains the highest expected geometric return. Plug the value of x into (8), the expected rate of return over the risk free rate is

$$((1+2q-1)^{q}(1-(2q-1))^{1-q}-1)(1+r)$$
$$=(2q^{q}(1-q)^{1-q}-1)(1+r)$$

If the objective return distribution of the risky asset is {p, 1-p} instead of the subjectively assessed {q, 1-q}, the expected rate of return of the portfolio over the risk free rate, without considering the factor 1+*r*, is

$$2q^{p}(1-q)^{1-p}-1 \qquad (10)$$

The first order approximation of (10) is

$$\ln(2q^{p}(1-q)^{1-p})$$
$$=p\ln q+(1-p)\ln(1-q)+\ln 2$$

Comparing the above result with (5), we find that the first order approximation of the expected rate of return of the portfolio constructed from a certain judgment is exactly equal to the value of the judgment. This result is an extension from Kelly (1956). As the value of judgment provides a good approximation to the rate of return on investment, it can be conveniently used to understand the relation between human judgment and patterns in investment returns and stock market.

From (9), the judgment about a stock determines the level of holding about the stock. The change of judgment about a stock determines the volume of trading in the market, which is considered as the key ingredient missing from the asset pricing models (Banerjee and Kremer, 2010). The theory of judgment provides a link between investors' judgment and trading volume in the asset market, which will be applied to understand cycles of trading in the next section. In the following, we will apply the theory to calculate several numerical examples on the level of holding of risky assets.

When interest rate is measured on the inflation adjusted basis, the risk free interest rate can be set to be zero as a good approximation to reality. The payoffs of one unit risky asset can be either 1 + d with probability p or 1– d with probability 1-p. We can calibrate the equilibrium value of *p* and *d* with the empirical data on return and standard deviation. The arithmetic mean rate of return of the risky asset is

$$pd+(1-p)(-d)=(2p-1)d \qquad (11)$$

and the standard deviation of the risky asset is

$$\{p[pd-(2p-1)d]^{2}+(1-\mathrm{p})[(1-\mathrm{p})(-\mathrm{d})-(2\mathrm{p}-1)\mathrm{d}]^{2}\}^{1/2} = 2d\sqrt{p(1-p)} \qquad (12)$$

respectively. Setting p = 0.55 and d = 0.25 results mean rate of return to be 2.5% and standard deviation to be 24.87%, which are similar to actual real rate of return and standard deviation of many stock markets in the world in the last ten years. With *p* = 0.55 and *d* = 0.25, the proportion of risky asset in the equilibrium portfolio, following formula (9), is

$$\frac{2p-1}{d}=\frac{2\times 0.55-1}{0.25}=0.4$$

These numbers will be regarded as default values of an equilibrium portfolio in the next section.

Different people at different times in different places may have different opinions about the future of the stock markets. We will calculate the proportions of asset to be allocated to the risky asset with different expectations. When p is equal to 57.5%, 60%, 62.5% while keeping d fixed at 0.25, the optimal allocations to risky asset are

$$\frac{2\times 0.575-1}{0.25}=0.6$$
$$\frac{2\times 0.60-1}{0.25}=0.8$$
$$\frac{2\times 0.625-1}{0.25}=1.0$$

while the arithmetic means of the risky asset, following formula (11), are

$$(2\times 0.575-1)\times 0.25=3.75\%$$
$$(2\times 0.6-1)\times 0.25=5\%$$
$$(2\times 0.625-1)\times 0.25=6.25\%$$

The real returns of the best performing stock markets in the worlds, such as United States, over the second half of the last century are close to or above 6.25%. This justifies the standard statement of high risk, high return and the common practice of allocating most or all assets in risky securities in long term investments. However, if the future expected real returns of risky assets are lower, the proportions of risky assets in investment portfolios should be lower as well to achieve high expected returns (McEnally, 1986).

In the above discussion, we assume a risky asset makes only two discrete movements with corresponding probabilities. In reality, a risky asset can make many different movements with corresponding probabilities. The numerical representation by (p, d), just like mean and variance, is a simplified characteristic on the movements of the risky asset.

5. Investor Heterogeneity in Information Processing and Related Market Patterns

In this section, we will apply the theory of judgment to build a simple model to study trading behaviors of the heterogeneous investors and the resulting market patterns. First we will determine the statistical distribution of investors with different levels of wealth. From earlier studies, such as those in Chatterjee, Yarlagadda and Chakrabarti (2005), wealth distribution follows exponential law as a first approximation. Investors can be classified based on their wealth. Suppose each investor in group i has i unit of wealth. Since the number of investors in each group of wealth follows the exponential law, the proportion of investors with i unit of wealth is

$$\frac{1}{2^i}$$

Since

$$\sum_{i=1}^{\infty}\frac{1}{2^i}=1$$

the proportion of investor population is normalized. The total wealth of the economy is

$$\sum_{i=1}^{\infty} \frac{i}{2^i} = 2$$

Because the investor population is normalized, the average wealth of an investor is 2.

From calculation, the Gini coefficient of this model economy is 33.3. The Gini coefficients of industrialized economies are roughly between 25 and 45. So the wealth distribution in this model economy is representative of real economies.

To further simplify discussion, we lump investors into three groups. The representative wealth for each member of the three groups of investors are 1, 4 and 9 respectively. The proportions of three groups of investors are determined by maximum entropy principle (Jaynes, 1957) with the constraints on total wealth

$$p_1 + p_2 + p_3 = 1$$

$$p_1 + 4p_2 + 9p_3 = 2 \tag{13}$$

Solving the maximum entropy problem

$$\max\{-p_1 \ln p_1 - p_2 \ln p_2 - p_3 \ln p_3)$$

subjecting to the constraints (12) gives the following answer

$$p_1 = 0.73$$

$$p_2 = 0.23$$

$$p_3 = 0.04$$

The total wealth for each group of investors are

$$w_1 = 1p_1 = 0.73$$

$$w_2 = 4p_2 = 0.95$$

$$w_3 = 9p_3 = 0.33$$

Roughly speaking, small investors with 1 unit of wealth represent individual investors. Empirical evidence shows that individual investors as a group lose money from their trading activities. So we will assume small investors do not possess private information. They base their trading decisions on past price movements and other public information. Empirical evidences show that large investors as a group make money from their trading activities. We assume the value of judgment from midsized investors with 4 unit of wealth is moderately positive and the value of judgment from large investors with 9 unit of wealth is significantly positive. The specific values of judgment by different investors will be quantified in the next paragraph.

From calculations performed in the last section, the equilibrium levels of *p*, *d* and the average proportion of risky asset held by an investor are 0.55, 0.24 and 0.4 respectively. However, from time to time, the probability of price movement of the risky asset will deviate from the equilibrium level because of various reasons. Investors with higher wealth level can detect more valuable information. Specifically, we assume large investors with 9 unit of wealth can detect all information {p, 1-p} with *p* up to 0.6 and midsized investors with 4 unit of wealth can detect all information {p, 1-p} with *p* up to 0.575.

We follow the standard literature on the assumptions of price movement of securities. The price of the risk free asset is assumed to be constant. The price movement of the risky asset is proportional to net active trading by the investors, except at the time of public release of information, when price can move without trading.

Now we consider a trading process that lasts for four time periods. At the beginning of period one, the firm underlying the risky security starts a new project. The earning from this project will become known to the public at the end of period two. This new project is expected to generate profit that corresponds to a payoff in security that is either 1 + d with probability 0.6 or 1– d with probability 0.4 in one time period if information is publicly revealed. This expected payoff is higher than the equilibrium payoff of either 1 + d with probability 0.55 or 1– d with probability 0.45. Large investors with 9 unit of wealth detect this information and purchase additional shares of the risky security. The proportion of wealth they invest in the risky security after the purchasing, according to formula (9), is

$$\frac{2p-1}{d} = \frac{2\times 0.6-1}{0.25} = 0.8$$

Since the total wealth of this group of investors is

$$w_3 = 0.33$$

The total volume of buying, which is the new holding minus the equilibrium holding at 40%, is

$$0.33\times(0.8-0.4) = 0.13$$

The purchasing by large investor increases the price of the shares and reduces the future expected returns. When the price increases to a certain level, this security will be represented by a state of future payoff that is either 1 + d with probability 0.575 or 1– d with probability 0.425 by the end of period two. This is the end of period one and the beginning of period two. In period two, midsized investors with 4 unit of wealth detect this information and purchase shares of the risky security. The proportion of wealth they invest in the risky security, according to formula (9), is

$$\frac{2p-1}{d} = \frac{2\times 0.575-1}{0.25} = 0.6$$

Since the total wealth of the midsized investors is

$$w_2 = 0.95$$

The total volume of their buying, which is the new holding minus the equilibrium holding at 40%, is

$$0.95\times(0.6-0.4) = 0.19$$

In period two, large investors with 9 unit of wealth will also keep sixty percent of their wealth in the risky asset. As a result, they will reduce the original holding. The total volume of their selling is

$$0.33\times(0.6-0.8) = -0.066$$

At the end of period two, the earning from the project becomes publicly known and the share price of the risky asset fully reflects the underlying fundamentals. Small investors with 1 unit of wealth do not possess private information. Instead, they observe that both the share price movement in the last two time periods and the earning announced at the end of period two are higher than the average. They extrapolate the past results to the future and invest accordingly. Because the share prices have moved up steadily over the last two time periods, it will be natural for small investors to base the trading decisions on the best trading decisions from two periods earlier. Specifically, in period three, on average, small investors will allocate eighty percent of their assets in the risky security. Since total wealth for the small investors is

$$w_1 = 0.73$$

The total volume of buying by the small investors, which is the new holding minus the equilibrium holding at 40%, is

$$0.73 \times (0.8 - 0.4) = 0.29$$

Now we will consider the trading activities of large and midsized investors. Their trading decisions are based on the information they received. By default, we assume no new information in the future. In this case, the movement of stock price will return to its equilibrium condition. As a result, the holdings of the risky asset by large and midsized investors will return to the equilibrium state of forty percent. The total amount they will sell is

$$(0.33 + 0.95) \times (0.4 - 0.6) = -0.25$$

The net active trading by all investors is

$$0.29 - 0.25 = 0.04 \tag{14}$$

Because of the small net active trading in period three, the price movement in this period is moderate. However, trading at the beginning of period three could be dominated by small investors who mainly depend on easy to understand information, such as earning data, which is distributed widely to the general public at very narrow time frames. So trading by small investors is highly correlated (Barber, Odean and Zhu, 2009b). Trading decisions by large and mid size investors depend more on intangible information, which is the main determinant of future returns (Daniel and Titman, 2006). But intangible information is less precisely defined and trading activities generated by intangible information is less concentrated. This means that the beginning of the period three is marked by rise of asset prices while prices decline over the rest of period three. This is consistent with the empirical evidence (Hvidkjaer 2008; Barber, Odean, and Zhu, 2009a, 2009b).

In period four, most relevant information has been acted upon and share price will finally reach equilibrium. Since share price at the end of period two has already fully reflect the fundamentals, the expected price level at the end of period four will be equal to the share price at the end of period two. Therefore, the combined net active trading of period three and four should be zero. From (14), the net active trading in period four should be -0.04. Since large and midsized investors already balanced their portfolio to equilibrium state in period three, the active trading is mainly generated by small investors who are reducing their holding from last period's buying. As there is little new information to generate extra trading, the total trading can be approximated by the net active trading.

We can summarize the trading activities in the four time periods into the following table:

	Period one	Period two	Period three	Period four
Net trading	0.13	0.12	0.04	-0.04
Trading volume	0.13	0.19	0.29	0.04

The average net trading of the four periods is

$$\frac{1}{4}(0.13 + 0.12 + 0.04 - 0.04) = 0.06$$

Since the net trading of the first two periods are higher than the average, share prices increase in the first two periods are higher than the average. They are the winner periods. In the last two periods, the net trading is lower than the average. Share prices change in the last two periods are lower than the average. They are the loser periods. Among the winner periods, the trading volume of the first period is lower than that of the second period. Among the loser periods, the trading volume of the fourth period is lower than that of the third period. The four trading periods can be summarized as

Period one	Period two	Period three	Period four
Low volume winner	High volume winner	High volume loser	Low volume loser

This is exactly the same as the empirical pattern documented in Lee and Swaminathan (2000), which they call momentum life cycle.

Hvidkjaer (2006) examined the trading behaviors of investors of different sizes at the stages of low volume winner, high volume winner, high volume loser and low volume loser. He inferred the background of investors from the sizes of the trades. In his classification, large trades are two times or more as large as small trades. In our model, the midsized and large investors are four and nine times larger than the small investors. So it is natural to merge the midsized and large investors into one group as large investors when comparing our theoretical predictions to the empirical results documented in Hvidkjaer (2006). We will examine how trading patterns predicted from our model correspond to empirical patterns. The clearest resemblance between the predictions of our model and the empirical patterns occur in high volume loser stage. From Figure 2 of Hvidkjaer (2006), small investors are active buyers while large investors are active sellers in this stage, which is exactly what the model has predicted. Our results are also consistent with Feng and Seasholes (2004), who showed that informed investors are selling while uninformed investors are buying after information release. In the low volume loser stage, from Figure 3 of Hvidkjaer (2006), small investors are more active sellers than large investors. In the low volume and high volume winner stages, from Figure 2 and 3 of Hvidkjaer (2006), large investors are more active buyers than small investors. If we interpret trading activities calculated from our model as dominate activities instead of all activities, the predictions of our model during these stages are consistent with the empirical patterns.

Alternatively, we can refine the model to make it more realistic. We had assumed the level of informedness of an investor is determined only by his wealth. To be more consistent with reality, we now assume the level of informedness of an investor is positively correlated but not determined by his wealth. Specifically, the correlation between wealth and level of informedness is represented by the following matrix

	0.55	0.575	0.60
1	0.7	0.2	0.1
2	0.15	0.7	0.15
3	0.1	0.2	0.7

This means that among investors with one unit of wealth, 70%, are uninformed, as 0.55 is the equilibrium state, 20% are informed at the level of 0.575 and 10% are informed at the level of 0.60. The level of informedness of group two and three investors can be understood similarly.

We make a further refinement about the informedness of investors who can detect the information {p, 1-p} with *p* up to 0.6. We will assume these investors detect the information but do not interpret the information precisely. To be more specific, these investors make a judgment that *p* is equal to 0.575 instead of 0.6. This is very natural since most investors underestimate the significant of new information. Note that the judgment of this group of investors is still more valuable than the group of investors who estimate *p* to be 0.575 when it is actually 0.575. With the refined model, we can recalculate the trading activities following the same procedure as before. But this time we will measure the trading activities of large investors and small investors separately. The calculation of net trading by small and large investors at period four is determined by the proportional holdings of small and large investors at the end of period three. The results are shown in the following table.

	Period 1	Period 2	Period 3	Period 4
small investor net trading	0.0145	0.0291	0.1600	-0.0076
large investor net trading	0.0742	0.1455	-0.1498	-0.0026
total net trading	0.0887	0.1746	0.0102	-0.0102
trading volume	0.0887	0.1746	0.2735	0.0102

From the above table, the trading patterns of small and large investors calculated in each period are qualitatively similar to the empirical patterns recorded in Figure 2 and 3 of Hvidkjaer (2006). However, we would not expect the patterns predicted from our model to be identical to empirical patterns collected in the literature. Our model presents an investment cycle initiated by a positive information signal. The empirical patterns are combinations of all kinds of cycle and non-cycle activities. For example, when the news is negative, a similar pattern exists at opposite directions. Different cycles have different amplitudes and length. In the future, we may conduct empirical investigations by filtering out different cycle and non-cycle components. This could help detect investment strategies with high level of returns.

This theory of judgment based model captured many stylized patterns of trading activities during the momentum reversal cycle. The mathematics involved is vey simple and the intuition from the model is very clear. However, it is still in an early stage of development. Many refinements can be made in the future, some of which are listed as follows.

First, relation between earning momentum and price momentum can be added into the model. Empirical evidence shows strong relation between earning and price movement (Lee and Swaminathan, 2000; Chordia and Shivakumar, 2006; Chen, Moise and Zhao, 2009). By modeling earning process over several periods of time, we can further clarify the trading mechanisms of small investors. If earning trends last longer, small investors, as well as other investors, will be more confident that the momentum will continue.

Second, informed investors can anticipate and influence the trading activities of uninformed investors. When informed investors can anticipate the trading behaviors of uninformed investors, they may base their trading decisions not only on the fundamental information, but also on the expected trading activities of uninformed investors. Informed investors, who usually have strong track records and are major stakeholders of publicly listed companies, can influence uninformed investors in certain ways to alter the trajectories of price movement to benefit themselves. These investor activities may be captured by more refined models.

6. Concluding Remarks

This paper presents an updated version of the entropy theory of human mind. Comparing with other quantitative theories in behavioral finance, which are mostly utility theories in one form or another, the entropy theory of mind is natural, simple, specific and intuitive. The entropy theory of mind is a natural extension from the entropy theory of information and statistical mechanics. The mathematical tools of the entropy theory of mind only involve simple algebraic functions such as logarithm functions and occasional use of calculus. The predictions derived from entropy theory of mind are more specific than from other theories. The links between investors' information processing and trading decisions are very intuitive under the entropy theory of mind.

Given many advantages of the entropy theory of mind over the utility theory, one may naturally ask why the concept of entropy only has very limited applications in social sciences. The question even attracted people from outside academia. Writer Poundstone devoted two books to examine the historical development of the concepts of utility and entropy and their relations with human mind and investment (Poundstone, 2005, 2010). Academic research, like all other social activities, is constrained by institutional structures. The concept of utility was introduced into economic research much earlier than the concept of entropy. It is always difficult to compete with a well established theory. We can only hope, through the dedicated efforts from people both inside and outside academia, the potential of entropy theory, which has catalyzed many, if not most, scientific breakthroughs in the last one and half centuries, can be further explored in the field of social sciences.

References

Banerjee, Snehal and Kremer, Ilan, 2010, Disagreement and Learning: Dynamic Patterns of Trade, *Journal of Finance*, Forthcoming

Barber, Brad, and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806

Barber, Brad, and Terrance Odean, 2008, All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *The Review of Financial Studies* . Vol. 21, 2, 785-818.

Barber, Brad, Odean, Terrance and Zhu, Ning, 2009a, Do Retail Trades Move Markets?, *Review of Financial Studies,* 22, 1, 151-186.

Barber, Brad, Odean, Terrance and Zhu, Ning, 2009b, Systematic Noise, *Journal of Financial Markets*, 12, 547-569.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307-345.

Beutelspacher, Albrecht. 1994. *Cryptology: an introduction to the art and science of enciphering*, **…** Mathematical Association of America.

Boone, J. V. 2005, *A brief history of cryptology*, Naval Institute Press.

Brav, Alon and Heaton, J.B., 2002, Competing Theories of Financial Anomalies, *The Review of Financial Studies*, 15, No. 2, 575-606.

Chatterjee, Arnab; Yarlagadda, Sudhakar; Chakrabarti, Bikas K. 2005, Econophysics of Wealth Distributions, Springer;

Chen, Jing, 2003, An Entropy Theory of Psychology and its Implication to Behavioral Finance, *Financiële Studievereniging Rotterdam Forum*, 6, No. 1, 26- 31

Chen, Jing, 2004, Generalized Entropy Theory of Information and Market Patterns, *Corporate Finance Review*, 9, No. 3, 23- 32.

Chen, Jing, 2005, The physical foundation of economics: An analytical thermodynamic theory, *World Scientific, Singapore*.

Chen, Jing, 2007, The Informational Theory of Investment: A Comparison with Behavioral Theories, *ICFAI Journal of Behavioral Finance*, 4, No. 1, 6 – 31

Chen, Jing, 2008, The Physical Foundation of the Mind, *NeuroQuantology*, 6, No. 3, 222-233

Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ Wermers, 2000,The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of fund managers, *Journal of Financial & Quantitative Analysis* 35, 343-368.

Chen, Long, Claudia Moise and Xinlei Zhao, 2009, Myopic Extrapolation, Price Momentum, and Price Reversal, working paper.

Chordia, Tarun and Lakshmanan Shivakumar, 2006, Earnings and Price Momentum, *Journal of Financial Economics* 80, 627-656.

Cohen, Randolpg, Paul Gompers, and Toumo Vuolteenaho, 2002, Who underreact to cash-flow news? Evidence from trading between individuals and institutions, *Journal of Financial Economics* 66, 409-462.

Cover, Thomas and Joy Thomas, 2006, Elements of information theory, 2nd edition, Wiley, Hoboken, NJ.

Cronqvist, Henrik and Thaler, Richard, 2004, Design choice in privatized social-security systems: learning from the Swedish experience, *American Economic Review*, 94, 424-428

Daniel, Kent, David Hirshleifer and Avanidhar Subrahmanyam, 1998, Investment psychology and Security market under and overreactions, *Journal of Finance* 53, 1839-1885.

DANIEL, K. and TITMAN, S. (2006), Market Reactions to Tangible and Intangible Information. The Journal of Finance, 61: 1605–1643.

DeBondt, Werner and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793-805.

Einstein A. 1905 (1989), On a Heuristic Point of View Concerning the Production and Transformation of Light. In *The Collected Papers of Albert Einstein*, Volume 2, Princeton University Press,

Engelberg, Joseph and Christopher Parsons, 2009**,** *The Causal Impact of Media in Financial Markets,* Working paper

Feng, Lei and Seasholes, Mark, (2004) Correlated trading and location, *Journal of Finance*, 59, 2117-2144.

Feng, Lei and Seasholes, Mark, (2005) Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Finance Markets? *Review of Finance*, 9, 305-351.

Friston, Karl, (2010) The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11, February, 127-138

Georgescu-Roegen, Nicholas, 1971. *The entropy law and the economic process***.** Harvard University Press, Cambridge, Mass.

Gibbs, J. Willard, 1873a (1906), *Graphical methods in the thermodynamics of fluids*, Collected in Scientific papers of J. Willard Gibbs, Ox Bow Press

Gibbs, J. Willard, 1873b (1906), *A method of geometrical representation of the thermodynamic properties of substances by means of surfaces*, Collected in Scientific papers of J. Willard Gibbs, Ox Bow Press

Gibbs, J. Willard, 1902 (1981), *Elementary Principles in Statistical Mechanics*, Ox Bow Press, Woodbridge, Connecticut.

Hong, Harrison and Jeremy Stein, 1999, A unified theory of underreaction, momentum trading and overreaction in asset markets, *Journal of Finance* 54, 2143-2184.

Hvidkjaer, Soeren, 2006, A trade-based analysis of momentum, *Review of Financial Studies* 19, no. 2, 457–491.

Hvidkjaer, Soeren, 2008, Small trades and the cross-section of stock returns, *Review of Financial Studies* 21, no. 3, 1123–1151.

Jablonka, Eva and Lamb, Marion J. (2006) Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life ,The MIT Press

Jammer, Max, 1966, *The Conceptual Development of Quantum Mechanics*. New York: McGraw-Hill, 2nd ed

Jaynes ET. 1957, Information Theory and Statistical Mechanics, *Physical Review*, 106: 620-630

Jaynes, E. T., 1988, `How Does the Brain Do Plausible Reasoning?', in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, **1,** 1-24, G. J. Erickson and C. R. Smith (eds.), Kluwer, Dordrecht,

Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, Journal of Finance 48, 65-91.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance*, 60, 1983-2011.

Kahneman, D. and A. Tversky, 1979, Prospect theory: A analysis of decision under risk, Econometica 47, 263-291.

Kelly, J. L., Jr. (1956), "A New Interpretation of Information Rate", *Bell System Technical Journal* **35**: 917–926

Kullback, Solomon, 1959, Information theory and statistics, John Wiley

Latane, Henry, 1959, Criteria for choice among risky ventures, *The Journal of Political Economy*, 67, 2, 144-155.

Latane, Henry and Tuttle, Donald, 1967, Criteria for portfolio building, *Journal of Finance*, 22, 359-373

Lee, Charles and Bhaskaran Swamminathan, 2000, Price momentum and trading volume, *Journal of Finance* 55, 2017-2069.
Maor, Eli, 1994, *e: the story of a number*, Princeton University press.
Maxwell, James, 1871, *Theory of heat*, (Green & Co., Longmans, London)
McEnally, Richard W., 1986, Latane's Bequest: The Best of Portfolio Strategies, *The Journal of Portfolio Management*, 12:21-30.
Novy-Marx, Robert, 2009, Is Momentum Really Momentum?, *Journal of Financial Economics*, forthcoming.
Odean, Terrance, 1999, Do Investors Trade Too Much, *American Economic Review* 89, 1279-1298.
Patterson, Scott, 2010, *The Quants: How a New Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed It* Crown Business;
Poundstone, William, 2005, *Fortune's Formula: The Untold Story of the Scientific Betting System That Beat the Casinos and Wall Street*, Hill and Wang
Poundstone, William, 2010, *Priceless: The Myth of Fair Value (and How to Take Advantage of It)*. Hill and Wang
Qian, Hong, 2001, Relative entropy: Free energy associated with equilibrium fluctuations and non-equilibrium deviations, *Physical Review E*, **63**, 042103.
Qian, Hong, 2009, Entropy Demystified: The ‘‘Thermo’’-dynamics of Stochastically Fluctuating Systems, in *Methods in Enzymology, Volume 467, Computer Methods, Part B*, Edited by Michal Johnson and Ludwig Brand, 112-134.
Rando, Oliver J. and Verstrepen, Kevin J., 2007**,** Timescales of Genetic and Epigenetic Inheritance, *Cell*, Volume 128, Issue 4, 655-668
Samuelson, Paul, 1969, Portfolio selection by dynamic stochastic programming, *The Review of Economics and Statistics*, 51, 3, 239-246
Samuelson, Paul. 1972. *The collected scientific papers*, Vol. 3, Cambridge,Mass.: MIT Press.
Schlögl, Friedrich, 1989, *Probability and heat : fundamentals of thermostatistics*, Braunschweig : F. Vieweg,
Shannon, Claude, 1948, A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.
Shannon, C. and Weaver, W. (1949). *The mathematical theory of communication*, Urbana: The University of Illinois Press.
Shleifer, Andrei, and Robert Vishny, 2003, Stock market driven acquisitions, *Journal of Financial Economics* 70, 295-311.
Sinn, Hans-Werner, 2003 Weber’s law and the biological evolution of risk preferences: The selective dominance of the logarithmic utility function, *The Geneva Papers on Risk and Insurance Theory* 28, 87-100;
Stevens SS, 1961, *To Honor Fechner and Repeal His Law: A power function, not a log function, describes the operating characteristic of a sensory system*. Science. 133 (3446), 80-6.
Stigler, Stephen, 1986, *The history of statistics: The measurement of uncertainty before 1900*, The Belknap Press.
Thorp, Edward O. 1997, *The Kelly criterion in blackjack, sports betting and the stock market*, presented at: The 10th International Conference on Gambling and Risk Taking Montreal.
Tversky, Amos and Daniel Kahneman, 1974, Judgment under uncertainty: Heuristics and biases, *Science* 185, 1124-1131.
Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance* 55, 1655-1703.