The Informational Theory of Investment:
A Comparison with Behavioral Theories

Jing Chen
School of Business
University of Northern British Columbia
Prince George, BC
Canada V2N 4Z9
Phone: 1-250-960-6480
Fax: 1-250-960-5544
Email: chenj@unbc.ca
Web: http://web.unbc.ca/~chenj/

.

Abstract

This paper develops a new theory of investment based on a newly developed information theory. This new information theory states that information is the reduction of entropy, not only in a mathematical sense, as in Shannon's theory, but also in a physical sense. The rules of information transmission developed in Shannon's theory, as mathematical rules, apply not only to communication systems, but also to all living organisms, including human beings. We also develop a new theory of learning from this information theory. We then demonstrate that the new information theory provides the foundation to understand major market patterns. Compared with behavioral theories, the assumptions of the informational theory of investment are much simpler and the informational theory can provide much more precise understanding about market behaviors. When predictions from this informational theory and behavioral theory differ, empirical results are consistent with the informational theory.

Keywords: information, entropy, human mind, psychology, behavioral finance
JEL classification: G14

Recent empirical works questioned whether investor performance should be interpreted primarily as a behavioral phenomenon or information-driven. They conclude that information is the primary reason of investment performance (Coval and Moskowitz, 2001; Ivkovic and Weisbenner (2005); Kacpercyzk, Sialm and Zheng (2005); Malloy, 2005). The empirical evidences suggest we should explore an information based theory of investment. However, "(current information) theory offers little guidance in identifying informed investors and in distinguishing between securities with scarce information and those with widely available information" (Coval and Moskowitz, 2001). In this paper, we will provide a new theory of investment based on a newly developed information theory that is generalized from Shannon's entropy theory of information. This informational theory of investment provides a unified understanding of observed market behaviors documented in increasingly abundant empirical works.

More than fifty years ago, Shannon (1948) developed the entropy theory of information in communication theory. Since then, its technique has been applied to many different problems in economic and finance. However, the standard economic theory of information, represented by Grossman and Stiglitz (1980) was not built on the foundation of entropy theory. Recently, entropy theory of information has been successfully applied to understand more economic phenomena. Sims (2003) utilized the capacity constraint in information processing to explain price stickiness. This concept has been applied to better understand some empirical patterns in asset market (Peng, 2005; Peng and Xiong, 2006). The significance of this idea is that it introduces physical constraint of information processing into economic literature. Sims noted, "Readers may recall, if they use modems with their computers, that modem speeds are rated in 'bits per second', or bps. This is their channel capacity. Speed increases for modems plateaued at 56000 bps, because telephone connections are designed to have approximated that capacity, so no change in modem design can achieve higher speed." This indicates that physical constraints are very relevant in information processing and entropy theory of information from communication theory may play a larger role in understanding economic behaviors. Recently, a new theory of information was developed to expand Shannon's entropy theory of information into an economic theory (Chen, 2003, 2004, 2005a). The basic idea of this theory is:

> Information is the reduction of entropy, not only in a mathematical sense, as in Shannon's theory, but also in a physical sense. The rules of information transmission developed in Shannon's theory, as mathematical rules, apply not only to communication systems, but also to all living organisms, including human beings.

Several important properties can be derived from this new information theory. First, information with higher value is in general more costly to obtain. This is a direct extension from Maxwell's (1871) thought experiment on an intelligent demon. Second,

the amount of information one can receive is the amount of information generated minus equivocation. The level of equivocation, which is the measure of information asymmetry, is determined by the correlation between the source of information and the receiver of information. In general, how much information one can receive depends on the background knowledge of a person. Therefore the process of understanding information is a process of learning, which often takes long time. This result is a direct extension from Shannon's theory from technical communication systems to human cognitive systems. It is consistent with the concept that people have limited information-processing capacity and differs from the standard assumption that market distills new information with lightning speed (Sims, 2003; Peng and Xiong, 2006). Third, the value of information is inversely related to the number of people who understand it. For example, an investor who buys the shares of a company before it becomes popular often earns higher rate of return than those who buy the shares of the same company when it becomes hot. At the same time, the value of a company's investment is also affected by how much its competitors understand the technology and market potential of a product.

With the development of this new information theory, the new theory of investment can be simply stated as the following:

> Major patterns in asset markets are the result of information processing by the heterogeneous investment public.

In the following, we will briefly compare the main properties of this informational theory of investment to that of behavioral theories and leave the detailed discussion about theoretical predictions and empirical evidences of this theory to later sections.

First, the informational theory needs very few assumptions. Since many behavioral biases have been documented in the cognitive psychology literature, almost any patterns in the financial markets can be linked to one or several of these biases. However, "the potentially boundless set of psychological biases that theorists can use to build behavioral models and explain observed phenomena creates the potential for 'theory dredging.'" (Chan, Frankel and Kothari, 2002) Furthermore, many theories, while consistent with empirical patterns that they are set out to explain, are not consistent with other empirical results. For example, while Bloomfield and Hales (2002) find evidence supportive of behavioral model of Barberis, Shleifer, and Vishny (1998) in a laboratory setting, Durham, Hertzel and Martin (2005) find scant evidence that investors behave in accordance with the model using market data. Behavioral models often assume some market sectors or activities are irrational while other parts are rational. However, they often differ on which part should be considered irrational (Shefrin, 2005). For example, Shleifer and Vishny (2003) develop a new model on corporate mergers. "Mergers in this model are a form of arbitrage by rational managers operating in inefficient markets. This theory is in a way the opposite of Roll's (1986) hubris hypothesis of corporate takeovers, in which financial markets are rational, but corporate mangers are not. In our theory, managers rationally respond to less-than rational markets" (Shleifer and Vishny, 2003, p. 297).

The informational theory, on the other hand, only needs one assumption: Human activities, including mental activities, follow physical laws. Strictly speaking, it is not really an assumption. Few, if any, would claim that human activities are beyond physical laws. But traditionally, researches on social sciences rarely discuss how physical laws constraint human activities. By requiring information processing of human beings to be constrained by physical laws, this informational theory offers very clear and precise prediction about market patterns.

Second, the informational theory provides more precise predictive power than the existing theories. The link between behavioral theory and investment behavior are often vague. For example, empirical works reveal that small investors' trading activities often hurt their investment return (Hvidkjaer, 2006). This is usually thought that small investors are more prone to behavioral biases than professionals, who are better trained (Shanthikumar, 2004). Yet some empirical work suggests that professionals exhibit some behavioral biases to a greater extent than non-professionals (Haigh and List, 2005). Empirical evidences indicate that the patterns of trading by small, individual investors differ systematically from that by large, institutional investors. Yet the existing behavioral theories do not offer any particular behavioral explanations to this systematic difference (Hvidkjaer, 2006; Shanthikumar, 2004). The new information theory, however, provide very precise understanding of the systematic differences between trading patterns of large and small investors. This information theory states that more valuable information is more costly to obtain in general. For large investors, it pays to spend a lot of effort and money to research the fundamentals. For small investors, it doesn't pay to dig into the fundamentals. They depend on processed and easy to understand information that is readily available at low cost, such as news from popular media and price movement of the shares. In this work, we will show that patterns of returns by small and large investors and patterns of return and trading volumes of stocks can be easily understood from this information processing perspective.

Third, the informational theory helps provide deeper understanding to behavioral models. For example, Hong and Stein's (1999) results are built on three key assumptions. The first two assumptions are that traders are classified as "newswatchers" and "momentum traders" according to their information processing abilities. They commented that, "the constraints that we put on traders' information-processing abilities are arguably not as well-motivated by the experimental psychology literature as the biases in Barberis et al. (1998) or Daniel et al. (1998), and so may appear to be more ad hoc" (Hong and Stein, 1999, p. 2145). These assumptions can actually be derived naturally from this new informational theory. Depending on the value of assets under management, different investors will choose different methods of information gathering with different costs. "Newswatchers" are large investors who are willing to pay a high cost to collect private information and to make a deep understanding of public information. "Momentum traders" are investors who 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 price momentum signals. Cohen, Gompers and Vuolteenaho (2002) empirically confirm that institutional investors buy on fundamental news while individual investors buy on price trends. The third assumption of Hong and Stein (1999) is that private

information diffuses gradually across the newswatcher population. The gradual diffusion of private information means that people gradually learn about the background knowledge of information and understand information better over time. This new information theory provides clear understanding to all three assumptions in Hong and Stein's model.

Fourth, when predictions from this informational theory and behavioral theory differ, empirical results are consistent with informational theory. Because behavioral theories are designed to explained specific market patterns, they are always consistent with those particular market patterns. However, because of vague link between market patterns and theoretical explanations, the consistency between theory and empirical patterns often do not survive generalization. The informational theory, on the other hand, is derived from fundamental physical laws that all living and non-living systems obey. This makes the theory less prone to the problem of overfitting theories to empirical observation. The debate about equity premium offers such an example.

Mehra and Prescott (1985) observed that the large size of risk premium on US equities can not be explained by the standard general equilibrium models and called it a puzzle. Behavioral theory attributes the high risk premium to loss aversion by investors. Barberis and Thaler (2003) provided a survey of works along this line.

The equity premium puzzle is a very simple pattern, which, once discovered, can be understood very easily by the investment public. The strategy of profiting from the high equity premium is easy to implement and of low risk. From the informational theory of investment, high equity premium, as an easy to understand pattern, will not persist for a long period.

Whether high equity premium is a universal pattern generated by loss aversion by investors can be answered by more comprehensive data. "For 1921 to 1996, U.S. equities had the highest real return of all countries, at 4.3 percent, versus a median of 0.8 percent for other countries. The high equity premium obtained for U.S. equities appears to be the exception rather than the rule." (Jorion and Goetzmann,1999, p. 953) If equity premium is caused by some emotion common to all human beings, it should present in stock markets all over the world. The empirical evidence is consistent with the informational theory and not consistent with the behavioral theories.

Empirical evidences show that patterns in the security markets are mainly driven by patterns of information processing by the investment public. However, this does not mean that human emotions do not influence market behaviors. It has been shown that most psychological patterns either reflect the constraints of physical laws or evolutionary adaptation to process information more efficiently in our evolutionary past (Chen, 2003, 2005a). In a word, human emotions are really low cost, but not necessarily unbiased ways of information processing. By showing that most patterns in asset markets documented in the literature can be explained by this information theory, the focus of attention can be directed to search the links between human biases and the phenomena or magnitude of phenomena in asset markets that could not be explained by the information theory.

Furthermore, human activities, including mental activities, are constrained by physical laws. These constraints offer initial tests to the plausibility of assumptions in behavioral theories. For example, from the information theory, new information can only be understood gradually by human beings. If a behavioral theory suggests investors will generally overact to new information, we need to examine the empirical evidence with great caution.

This paper is an update from earlier works Chen (2003, 2004, 2005a, 2005b). The remainder of the paper is structured as follows. Section I presents the generalized entropy theory of information. This information theory provides natural measures of the cost of obtaining information and of information asymmetry. It offers a deep connection among information, entropy and human mind. Section II develops a general theory of knowledge and learning from the perspective of Gibbs inequality and the theory of coding. Section III explains how patterns of investor returns and patterns of security returns and trading volumes are natural results of information processing by the heterogeneous investment public. It also answers many questions on the gaps between existing behavioral theories and empirical evidences. Section IV discusses how this information based Investment theory is related to models of behavioral finance. Many assumptions in some of the recent theoretical models can be derived naturally from the generalized entropy theory of information. Section V concludes.

I. Entropy, resource, information, and the evolution of mind

The value of information is a function of probability and must satisfy the following properties:

  (a) The information value of two events is higher than the value of each of them.
  (b) If two events are independent, the information value of the two events will be the sum of the two.
  (c) The information value of any event is non-negative.

The only mathematical functions that satisfy all the above properties are of the form

$$H(P) = -\log_b P \tag{1}$$

where H is the value of information, P is the probability associated with a given event and b is a positive constant (Applebaum, 1996). Formula (1) represents the level of uncertainty. When a signal is received, there is a reduction of uncertainty, which is information.

Suppose a random event, X, has n discrete states, $x_1, x_2, \ldots, x_n$, each with probability $p_1, p_2, \ldots, p_n$. The information value of X is the average of information value of each state, that is

$$H(X) = -\sum_{j=1}^{n} p_j \log(p_j) \qquad\qquad (2)$$

The right hand side of (2), which is the entropy function first introduced by Boltzmann in the 1870s, is also the general formula for information (Shannon, 1948).

After the entropy theory of information was developed in 1948, its technique has been applied to many different problems in economic and finance. However, the standard economic theory of information, represented by Grossman and Stiglitz (1980) was not built on the foundation of entropy theory. Economists often feel that "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) The Shannon measure actually carries weight of information. For example, N symbols with identical Shannon measure carry N times more information than a single symbol (Shannon, 1948). Similarly, the value of the information about the future price is higher to a large manufacturer than to a small manufacturer, other things being equal.

Authorities in information theory also discouraged the application of entropy theory to broader areas.

> 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)

This orthodox view was reaffirmed recently:

> The efforts of physicists to link information theory more closely to statistical physics were less successful. It is true that there are mathematical similarities, and it is true that cross pollination has occurred over the years. However, the problem areas being modeled by these theories are very different, so it is likely that the coupling remains limited.

> In the early years after 1948, many people, particularly those in the softer sciences, were entranced by the hope of using information theory to bring some mathematical structure into their own fields. In many cases, these people did not realize the extent to which the definition of information was designed to help the communication engineer send messages rather than to help people understand the meaning of messages. In some cases, extreme claims were made about the applicability of information theory, thus embarrassing serious workers in the field. (Gallager, 2001, p. 2694)

However, the dissonance between entropy function as a mathematical representation of information and the practical value of information has long puzzled many people and recent works have shown that our intuitive concept of information coincides with the mathematical definition of information as entropy (Adami, 2004; Bergstrom and Lachmann, 2004; Qian, 2005). In the following we will provide a more formal argument. If some decision making process is truly important and is needed again and again in life, it is highly economical that quantitative modules to be evolved in the mind to expedite the process. For example, predators need routinely to assess their distance from the prey, the geometry of the terrain, the speed differential between itself and the prey, the energy cost of chasing down its prey, the probability of success of each chase and the amount of energy it can obtain from prey to determine whether, when and where to initiate a chase. There are many other sophisticated functions, such as navigation by migrating birds over long distances, that need sophisticated mathematical capabilities. Many animals need to make precise calculations of many of these quantitative problems many times in life. To reduce the cost of estimation, mathematical models must have evolved in their mind so that many decision making processes are simplified into parameter estimation and numerical computation. It is highly likely that, if some function is very important for the survival of the animal, in the process of evolution, this function will be genetically assimilated.

From the second law of thermodynamics, a system has a tendency to move from a state of low entropy to a state of high entropy. This tendency of directional movement is the source of useful energy that drives, among other things, the living organisms. Mathematically, the increase of entropy is equivalent to the decrease of useful energy, which is defined as free energy in physics (p. 85, Avery, 2003). Since all living organisms need to extract low entropy from the environment to compensate continuous dissipation and entropy is the only mathematical function to measure scarcity of resources (Chen, 2005), it is inevitable that information, which we collect for our survival, is largely about entropy. It is not a mere coincidence that our intuitive concept of information and the mathematical definition of information as entropy largely overlap. The intuitive concept is really a simplified evaluation of a mathematical computation.

If Shannon's entropy theory of information is purely a mathematical theory with little connection with the physical laws, it would be a miracle that information defined as entropy turns out to have the magic properties that handle technical problems in communication so well. However, once mathematical theories are thought to be a natural part of our evolutionary legacy, it would be natural for entropy theory of information to possess this property.

From the above discussion, an entropy theory based economic theory of information can be simply stated as:

> Information is the reduction of entropy, not only in a mathematical sense, as in Shannon's theory, but also in a physical sense. The rules of information transmission developed in Shannon's theory, as mathematical rules, apply not only to communication systems, but also to all living organisms.

In the following, we will discuss some distinct properties of this new information theory. First, information that is more valuable is in general more expensive to obtain. From the second law of thermodynamics, Maxwell (1871) concluded that information of higher value is of higher physical cost. Since economic cost is highly correlated to physical cost (Georgescu-Roegen, 1971; Chen, 2005), more valuable information is in general more expensive to obtain.

Second, the amount of information one can receive depends on the person's background knowledge about that particular information. The most important result from Shannon's entropy theory of information is the following formula

$$R = H(x) - H_y(x) \tag{3}$$

where R is the amount of information one can receive, H is the amount of information a source sent and $H_y(x)$, the conditional entropy, is called equivocation. Formula (3) shows that the amount of information one can receive would be equal to the amount of information sent minus the average rate of conditional entropy. Before Shannon's theory, it was impossible to accurately assess how much information one can receive from an information source. In communication theory, this formula is used to discuss how noises affect the efficiency of information transmission. But it can be understood from more general perspective. The level of conditional entropy $H_y(x)$ is determined by the correlation between senders and receivers. When x and y are independent, $H_y(x) = H(x)$ and $R = 0$. No information can be transmitted between two objects that are independent of each other. When the correlation of x and y is equal to one, $H_y(x) = 0$. No information loss occurs in transmission. In general, the amount of information one can receive from the source depends on the correlation between the two. The higher the correlation between the source and receiver, the more information can be transmitted.

Since conditional entropy is closely associated with correlation, it is often used as a measure of correlation. Let ρ represent correlation between x and y. From p. 45, Cover and Thomas (2006), it can be defined as

$$\rho = 1 - \frac{H_y(x)}{H(x)}$$

With this definition, it can be shown that

$$R = \rho H(x) \tag{4}$$

Therefore the amount of information one can receive is proportionally to the level of correlation between information senders and receivers.

The above discussion does not depend on the specific characteristics of senders and receivers of information. So it applies to human beings as well as technical

communication equipments, which are the original focus in information theory in science and engineering. However, the laws that govern human activities, including mental activities, are the same physical laws that govern non-living systems, such as communication equipments.

$H_y(x)$ in Formula (3) offers the quantitative measure of information asymmetry (Akerlof, 1970). Since different people have different background knowledge about the same information, heterogeneity of opinion occurs naturally. To understand the value of a new product or new production system may take the investment public several years. To fully appreciate the scope of some technology change may take several decades. For example, the economic and social impacts of cars as personal transportation instruments and computers as personal communication instruments were only gradually realized over the path of several decades. This is why individual stocks and whole stock markets often exhibit cycles of return of different lengths. This property is very different from Grossman-Stiglitz information theory, where economic agents can recognize the value of information instantly and pay according to its value.

Third, the same information, when known to more people, becomes less valuable. Figure 1 is a graph of (1), where H is a function of P, the probability of any given event. From Figure 1, value is a decreasing function of probability. In the standard information theory, P represents the probability that some event will occur. In this theory, P is generalized to represent the percentage of people or money that is controlled by informed investors. When $P = 1$, -log P =0. Thus the value of information that is already known to everyone is zero. When P approaches zero, -log$P$ approaches infinity. Therefore, the value of information that is known to few is very high. The following example will illustrate this point. Figure 2 shows overnight rate of return and trading volume of shares of WestJet, a Canadian airline, surrounding the announcement of the bankruptcy of Jetsgo, the main competitor of WestJet. Jetsgo announced bankruptcy at the evening of March 10, 2005. If one bought stock at March 10, he would have made a return of 40% overnight. Judging from the trading volume of March 9, some people did buy WestJet stock before information was released to the public. After the announcement made the information public, trading volume was very high and the rate of return is near zero. Figure 2 neatly illustrates the relation between value of some information and the number of people who know the information.

Unlike Grossman-Stiglitz information theory, this information theory is a non-equilibrium theory. It does not assume a company possesses some intrinsic value waiting to be discovered by the investment public. Instead, the process of understanding the value of a company by the investment public is accompanied by the process of understanding the technology and market potential by its competitors, which generally reduce the value of that particular company. Empirical evidences that we will present later support this statement.

Since this new information theory can be applied to much broader fields than Shannon's theory, it may be called the Generalized Entropy Theory of Information. Empirical evidences show that Grossman-Stiglitz information theory cannot explain observed

investor behaviors (Barber and Odean, 2000). Other models extended from Grossman-Stiglitz theory often have little explanatory power outside a limited scope. We will show later that the generalized entropy theory of information will provide a unified understanding of the empirical market behaviors documented in the literature.

II. A general theory of knowledge and learning

What is knowledge and learning? The following result from statistical physics helps answer this question.

If $\{p_1, \ldots p_n\}$ and $\{q_1, \ldots q_n\}$ are two sets of probabilities, then

$$-\sum_{j=1}^{n} p_j \log(p_j) \leq -\sum_{j=1}^{n} p_j \log(q_j) \tag{5}$$

with equality achieved if and only if each

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

This result is called Gibbs inequality. (Isihara, 1971) Since the expression

$$-\sum_{j=1}^{n} p_j \log(q_j) \tag{6}$$

will be used frequently, we call it extended entropy, which is a more general than the entropy formula. In Gibbs inequality, p_j can be understood as the probability of event j in nature and q_j is the subjective probability of our assessment of that event, which we often call knowledge or learning. The left hand side of formula (4) is the average uncertainty of events and the right hand side is the uncertainty of our subjective assessment of those events. In general, the difference between the left hand side and right hand side of (4) is smaller when q_j is closer to p_j.

From Gibbs inequality, the level of uncertainty in understanding a type of events can be represented by extended entropy

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

where p_i and q_i are objective and subjective probabilities respectively. Suppose this type of events has two possible outcomes, state 1 and state2. The probability of state 1 is 90% and the probability of state 2 is 10%. An expert on this type of events may correctly estimate these probabilities and for her the uncertainty in prediction is

$$-0.9\ln 0.9 - 0.1\ln 0.1 = 0.33$$

A novice, who has no priori knowledge on these events, may assign 50% probability to each outcome. For her the uncertainty in prediction is

$$-0.9\ln 0.5 - 0.1\ln 0.5 = 0.69$$

It is clearly that the expert, who has accumulated knowledge through long time experience, has better estimation than novice in a stable environment.

Now assume the environment experiences some fundamental change and the new probabilities of state 1 and state 2 become 10% and 90% respectively. This time, the uncertainty of the prediction by the expert, who still uses the same probability, is

$$-0.1\ln 0.9 - 0.9\ln 0.1 = 2.08$$

while the uncertainty of prediction by a novice is

$$-0.1\ln 0.5 - 0.9\ln 0.5 = 0.69$$

This shows that when environment changes suddenly, novice actually perform better than experts, whose priori knowledge often cause severe biases in prediction. This is why successful investors cannot eliminate other investors continuously and why financial markets will not become progressively more efficient over time. This also explains why new industries are often pioneered by newcomers or outsiders.

The state of knowledge can be defined by the level of relative entropy

$$D(p \parallel q) = \sum_{j=1}^{n} p_j \log(p_j) - \sum_{j=1}^{n} p_j \log(q_j) \tag{7}$$

which is always nonnegative and is zero if and only if $p = q$. p represents the objective distribution of reality and q represents the subjective understanding. The lower the value of relative entropy, the better our understanding of reality.

The effect of learning and knowledge can also be understood from the theory of coding. A major purpose of coding is to reduce the average length of codes from knowledge about the probability distribution of events. In general, more frequently events are represented by shorter codes. Human language, a coding system we created to represent

our world, offers an example of such economic principle. 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. 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 world to the abstract symbol. It is a highly structured coded mapping. This coding reduces the average length of the message, or equivalently, reduce the entropy of the message. However, the change of environment often makes the original optimal code worse than a generic code. We will use a coding example adapted from Cover and Thomas (2006) to illustrate this.

Suppose in a communication system, information is transmitted with four symbols, 1, 2, 3 and 4. 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

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

Assume that, through experience, we learned that the probabilities of four symbols 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 above probability distribution is

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

which means that the shortest possible average length of binary code per symbol is 1.75. Such code can be designed with the following mapping:

$$C(1) = 0$$

$$C(2) = 10$$

$$C(3) = 110$$

$$C(4) = 111$$

Therefore, learning or knowledge can help reduce the average length of coding, or reduce the uncertainty in understanding. Indeed, knowledge itself can be regarded as a coding that makes understanding of the world easier (Sims, 2003).

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

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

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

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

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

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

$$\frac{1}{8} \times 1 + \frac{1}{8} \times 2 + \frac{1}{4} \times 3 + \frac{1}{2} \times 3 = 2\frac{5}{8}$$

which is higher than 2, the average length of generic code that can be designed without any knowledge about the probability distribution of the symbols. Therefore, knowledge gained from past experience may hinder instead of help information processing when environment changes substantially.

The average length of a coding can also be calculated with extended entropy.

$$-\frac{1}{8}\log(\frac{1}{2}) - \frac{1}{8}\log(\frac{1}{4}) - \frac{1}{4}\log(\frac{1}{8}) - \frac{1}{2}\log(\frac{1}{8}) = 2\frac{5}{8}$$

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

III. Theoretical predictions and empirical evidences

A. Differences in informational advantages of investors of different sizes

From this information theory, information with higher value is in general more costly. The amount of time and resource an investor will spend depends on the value of her portfolio. It can be predicted that the rate of return for large, institutional investors will be higher than small, individual investors because large investors will spend more time and resources in research. As the market rate of return is the average rate of return of all investors, we expect large investors outperform market average and small investors underperform market average. This is supported by empirical evidences. (Wermers, 2000; Barber and Odean, 2000; Cronqvist and Thaler, 2004)

If it is not economical for small investors to spend a lot of effort to study individual stocks, can they expect superior returns by investing in mutual funds with talented managers? The information theory predicts that the value generated by the talented managers will be mainly retained by the managers, who spend time to collect and analyze information, and not be distributed to mutual fund investors. Wermers (2000) documented that during the 1977 to 1994 period, mutual funds did generate higher raw rate of return than the market indices, which indicates researches by mutual fund managers do uncover valuable information. However, the net rate of return to investors is an average of 13.3 percent per year, which is the same as the Vanguard fund, the largest index fund. As the average transaction costs decrease over the years, average expense ratios increase, making the sum of these two costs remaining relatively constant over the years. (Wermers, 2000) This shows that it is the mutual fund managers, not the average investors, who capture the benefit of declining transaction costs. This is consistent with the model of Berk and Green (2004), where mutual fund managers capture the entire rents from their performance. In essence, for small investors without special information advantage or people whose cost of collecting and analyzing information outweighs its benefit, which are the majority of the population, they cannot expect a higher rate of return on stocks than the general market. The above result may be called the generalized efficient market hypothesis: Investors without informational advantage can not outperform the general market.

Odean (1999) documented that the shares individual investors sold outperform the shares they bought. He attributed this to some unknown biases of investors' judgment and stated, "What is more certain is that these investors do have useful information which they are somehow misinterpreting." (Odean, 1999, p. 1296) The real reason may be a lot simpler. Chen, Jegadeesh and Wermers (2000) documented that shares bought by mutual fund managers outperform shares they sold. The differential performance of shares bought and sold by individual investors is partly due to informational advantages of some of their trading counterparties. Even if individual investors randomly select buy and sell orders, most trades that get executed are not in their favor, because some of their counterparties, such as mutual fund managers, possess valuable information and select to fill the orders that are in their favor. For example, if an individual investor randomly select two stocks to sell with limit orders, it is more likely that the stock which will have a higher rate of return in the future will be bought by more informed investors. Later, we will demonstrate that small investors' dependence on low cost, low value information does make them prone to bad trading decisions.

Barber and Odean (2000) documented that there is very little difference in the gross performance of households that trade frequently and those that trade infrequently. On average, shares individual investors sold outperform the shares they bought (Odean, 1999). If all else are same, more frequently traded investors should earn lower rate of gross return. So this indicates that more frequently traded investors have better investment skills than less frequently traded investors. This prediction can be directly verified from trading records. Barber and Odean cited Carhart (1997) to show that frequent trading also hurts the performance of mutual funds. However, a more recent and more detailed study by Wermers (2000) shows that the most traded mutual funds

outperform least traded mutual funds by a wide margin. Barber and Odean attribute trading activities to overconfidence, a behavioral explanation. It may be simpler and more consistent with empirical evidences to explain trading as a type of learning activity. Like all other kinds of learning, trading is costly. The cost of learning is compensated, at least partially, by knowledge gained from the experience.

B. Differences in informational advantages of investors of different backgrounds

From (3), the ability to understand information depends on the background knowledge of investors. This indicates that investors will earn higher rate of return if they choose to invest in securities that they are familiar with. This is supported by some recent empirical investigations. Professional managers' and individual investors' local investments outperform their remote investments (Coval and Moskowitz, 2001; Ivkovic and Weisbenner, 2005). Malloy (2005) provides evidence that local analysts are significantly more accurate than other analysts in forecasting and recommendations. Mutual funds with high industry concentration, where fund managers can focus on particular industries they are familiar with, are more successful in selecting securities than diversified funds (Kacpercyzk, Sialm and Zheng, 2005). Investors take longer time to understand information from sources they are less familiar with. Hong, Lim and Stein (2001) empirically confirm that information from small firms, from firms with low analyst coverage and from firms with bad news, which managers are reluctant to release, generally diffuse slower. From Hvidkjaer (2006), the selling pressures on losers generally are stronger and last much longer than buying pressures on winners, suggesting information processing is less efficient on bad news.

Since insiders understand information much better than others, they can and will take advantage of this information asymmetry to pursue certain corporate activities. Shleifer and Vishny (2003) develop a new model on corporate mergers, in which mergers are a form of arbitrage by rational managers operating in inefficient markets. They show that their model has better explanatory power than the behavioral based models and is supported by new empirical evidences (Ang and Zheng, 2002). While assuming managers are rational, they continue to assume financial markets are less-than-rational. However, form the information theory, there is no need to assume financial markets are less-than-rational. In the next subsection, we will show how information processing by investors of different sizes with different background knowledge will generate cycles of undervaluation and overvaluation.

C. Information processing by heterogeneous investors and market patterns

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 are based on some ad hoc assumptions and could not explain other patterns (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)

Lee and Swaminathan (2000) and Hvidkjaer (2006) presented many empirical results regarding return, trading volume and patterns of investment decisions by traders of different sizes. Many of these results, especially those in Hvidkjaer (2006), are not presented in a systematic way because of the lack of a proper theoretical framework. In the following, we will show how the generalized entropy theory of information offers a coherent and comprehensive understanding of observed market patterns and the underlying dynamics.

To an investor, the choice of information gathering is a matter of cost and the ability to understand some information depends on her background. More valuable information is more costly to obtain in general. For large investors, it pays to spend a lot of effort and money to research the fundamentals. For small investors, it doesn't pay to dig into the fundamentals. They depend on processed and easy to understand information that is readily available at low cost, such as news from popular media and price movement of the shares. Cohen, Gompers and Vuolteenaho (2002) confirmed that institutional investors buy on fundamental news while individual investors buy on price trends. Whether an investor will be able to understand certain information also depends on her particular background, which determines her level of equivocation in receiving that information. Information processing by investors of different sizes with different backgrounds will generate the return and trading volume cycle that is similar to Lee and Swaminathan's (2000) momentum life cycle. In the following, we will illustrate the patterns of return and trading volumes of a stock of a typical company from the information processing perspective.

Suppose a company develop a new technology, which is expected to bring the company high profit in the future. From (3), those who are familiar with the technology and company will have low equivocation in receiving the information. They understand the significance of the information first and buy the company shares. Since they are a small number of people, the buying is of low volume. This corresponds to the beginning of the low volume winner stage in momentum life cycle. From table IX of Lee and Swaminathan (2000), the return on equity does improve over the next three years for low volume winner, which shows that the investors in this stage do have accurate perception about the future. From Figure 3 of Hvidkjaer (2006), the buying pressures from both large trades and small trades in this period trend up gradually, signaling the gradual diffusion of information. The buying pressure from large trades are higher than the buying pressure from small trades, which shows that large traders as a group are better informed than small traders.

As the technology goes through various stages from R&D to production, the potential becomes clearer to more people. This means that the level of equivocation gradually reduces to more people, which sustains buying interest and share prices increase gradually. As the technology becomes adopted in production and profit figures become public, the level of equivocation decrease further and the pool of investors increases further. Eventually, both the sustained increase of stock price and stable pattern of profit increase, which are very easy to understand by the general public, attract large amount of buyers, which results in high trading volume and pushes the stock prices further up. This corresponds to the high volume winner stage in momentum life cycle. From Figure 2,3 of Hvidkjaer (2006), there is a steady and higher buying pressure among large traders than in low volume winner stage, signaling a consensus of bullish sentiment from informed investors. Because of this consensus, the return of this stage is extremely high. (Lee and Swaminathan, 2000, Table IV) From Table IX of Lee and Swaminathan (2000), the return on equity is very high for high volume winners. However, the high return of the company will attract the attention of not only investors but also competitors, which will try to produce same or similar products for this high profit market.

From Figure 1, the value of some information that is known to everyone is zero. As the good news reaches most investors, the security is probably already fully or over priced. Among the increased pool of investors, more and more investors understand very little of the fundamentals behind the technology and depend on easy to understand signals such as coverage from popular media and stock price movement to make trading decisions. For this group of investors, they will stop buying only when the opinion of public media changes and the trend of price increase reverses significantly. As stock price keeps increasing, momentum trading becomes highly profitable, which will eventually push the share prices higher than the fundamental value. Since large investors spend more resources in investment, they are generally better informed than small investors. As share prices become highly overvalued relative to fundamentals, large investors start to unload positions while small investors keep buying. As the selling pressure from large investors becomes greater than the buying pressure from small investors, the trend of price increase reverses to price decrease. (Hvidkjaer, 2006, Figure 2) This is the period of high volume losers in momentum life cycle. In the period of high volume loser, as competition intensifies, return on equity drops sharply from previous years (Lee and Swaminathan, 2000, Table IX).

As the pattern of price drop becomes clear, more and more people joined the selling. After large investors and some of the small investors have finished unloading the positions, the volume of trading will decline, which is the period of low volume loser in momentum life cycle. This period is characterized by active selling of small investor. (Hvidkjaer, 2006) Since small investors are typically slow to understand information, their active selling, after the selling by large investors, signals the selling is overextended, which indicates the low volume losers will rebound and earn high future return in general. From the operation point of view, this is the worst time for the company. Overcapacity of a once high profit margin industry pushes down the return on equity further from the high volume loser stage. There are probably some layoffs of labor and

write off of capital. But the return on equity will gradually regress toward normal level (Lee and Swaminathan, 2000, Table IX).

The above paragraphs describe the patterns of information processing and trading when the initial news is positive. When the news is negative, a similar pattern exists at opposite directions. Since there is a cost shorting stocks and there are many institutional constraints on shorting stocks, short selling is much more difficult than buying stocks. With good news, there are many potential buyers. With a bad news, the sellers are largely confined to existing share holders. So overreaction is less strong on bad news. The statistical results, which are the average of all phenomena, mainly reflect the action from good news instead of bad news. With this observation in mind, we can discuss the following:

"The Hong and Stein (1999) model predicts that momentum profit should be larger for stocks with slower information diffusion. If we make the assumption that scarcity of trading leads to insufficient diffusion of information, then the Hong and Stein model would predict a greater momentum effect among low volume stocks. Our result indicate this to be true among winners but not among losers. That is, low volume winners have greater momentum, but low volume losers actually have less momentum." (Lee and Swaminathan, 2000, p. 2062)

The information theory indicates that low trading volume may reflect either a lack of understanding of some new information or a lack of information. From the above discussion, there are two types of low volume losers. The first, which is an average representative from Lee and Swaminathan's statistical results, is part of a cycle that is triggered by some good news. It has experienced the cycle of low volume winner, high volume winner, high volume loser and low volume loser. The low volume loser period represents the end of an information processing cycle. This is why it exhibits less momentum. The other type is the low volume loser period at the beginning of an information processing cycle that is triggered by a bad news (Hong, Lim and Stein, 2001). These low volume losers do exhibit strong momentum in low rate of return. So the apparent inconsistency of both views can actually be reconciled with a more detailed analysis from information theory.

The above analysis shows that securities often experience cycles of underreaction and overreaction as the result of investors' information processing. What determine the level of underreaction and overreaction? It depends on how much we understand the fundamentals. If the fundamentals are easy to understand by many people, both initial underreaction and eventual overreaction will be small. If the fundamentals are difficult to understand, mispricing can be substantial. We can have a look at glamour stocks. Glamour stocks are from companies with high earning growth. This means they have very few potent competitors, which generally indicates the lack of deep understanding about the particular products or production systems. That is to say, there is a high level of information asymmetry between the companies of glamour stocks and the general public. Initially, these types of companies are underpriced because few people understand them. However, the solid earning growth of these companies makes the share prices grow

continuously, generating clear technical signals. The clarity of technical signals and vagueness of fundamental information will eventually cause high level of overreaction. Statistical results show that stocks undergoing price momentum over longer period will exhibit higher level of reversal. (Lee and Swaminathan, 2000, Table I) Economy wide, great speculative bubbles are generally associated with "new era" or "new economy", when the general public is touched by the profound influence of technology breakthroughs while having little understanding of the underlying mechanisms (Shiller, 2000).

In the following, we will answer the three questions posed by Lee and Swaminathan at the end of their paper.

"First, the asymmetry in the timing of momentum reversals between winners and losers remains a puzzle. We show that low volume losers rebound quickly and outperform high volume losers with the next three to 12 months. However, it takes low volume winners longer (more than 12 months) to significantly outperform high volume winner. We know of no explanation for this timing difference." (Lee and Swaminathan, 2000, p. 2067)

From our analysis of information processing cycle, the low volume winner stage is the gradually understanding of fundamental news about a firm. Since the understanding of fundamentals is very costly and generally take very long time, it will take long time for low volume winners to significantly outperform high volume winners. The high volume loser stage is when large investors are already well informed about the overpricing and are active sellers. The price at this stage is supported by active buying of small investors, which mainly respond to popular media coverage and technical signals. (Hvidkjaer, 2004) Since coverage from popular media and technical signals, which are information with low cost and low value, are easier to understand than details about fundamentals, the price adjustment at this stage is much faster.

"Second, with the possible exception of the disposition effect from the behavioral literature, we know of no explanation for why trading volume should decline when firms fall out of favor."

The volume of trading reflects how many investors believe they can make profitable trades. When stocks are out of favor, few people believe they can make a profit buying these stocks. Hvidkjaer's detailed analysis shows that losing stocks do experience consistent selling pressures over a long period of time. The low volume of trading when firms fall out of favor reflects one fundamental asymmetry in security trading. For a stock, there are always more potential buyers than potential sellers, who are largely existing share holders.

"Finally, we find it remarkable that measures as readily available as past returns and trading volume can have such strong predictive power for returns. … Why this information is not fully reflected in current prices is another puzzle we leave for future research."

From (3), how much information we can understand depends on our background knowledge about the information and how much weight we assign to the information. From the efficient market theory, trading volume carries very little information. So little weight was given to the idea that trading volume might contain valuable information, which inhibited the research on this direction in the past.

D. On equity premium

Mehra and Prescott (1985) observed that the large size of risk premium on US equities can not be explained by the standard general equilibrium models and called it a puzzle. Among much research generated by this observation, two approaches are relevant to our study. One approach attributes the high risk premium to loss aversion by investors. Barberis and Thaler (2003) provided a survey of works along this line. The other is survivorship bias proposed by Jorion and Goetzmann (1999).

From the generalized entropy theory of information, how long a pattern persists depends on the cost of learning. It is often very costly to gain a deep understanding of a company or an industry, especially when the industry is new. The equity premium puzzle, however, is a very simple pattern on financial data, which, once discovered, can be understood very easily by the investment public. The strategy of profiting from the high equity premium is easy to implement and of low risk. This indicates that the pattern of high equity premium, if it does exist, is a short term one.

Whether high equity premium is a pattern or a result of selection bias can be answered by more comprehensive data. Jorion and Goetzmann (1999) documented that among all the equity markets around the world in the past century, the US market had the highest return. They argued that US market had the highest return because US was the most successful economic system in the world in the last century. "For 1921 to 1996, U.S. equities had the highest real return of all countries, at 4.3 percent, versus a median of 0.8 percent for other countries. The high equity premium obtained for U.S. equities appears to be the exception rather than the rule." (Jorion and Goetzmann,1999, p. 953) If equity premium is caused by some emotion common to all human beings, it should present in stock markets all over the world. The empirical evidence is consistent with the informational theory and not consistent with the behavioral theories.

IV. Understanding models of behavioral theories

Recently, several behavioral models provide frameworks to interpret the short to intermediate term momentum and long term reversal of return. In this section, we will discuss the informational theory help understand these models. Daniel, Hirshleifer and Subrahmanyam (1998) explain momentum in terms of both initial and delayed overreaction, while Barberis, Shleifer and Vishny (1998) and Hong and stein (1999) explain momentum in terms of initial underreaction and followed by delayed overreaction.

From the information theory, the absorbing of a new information is a gradual process, in which the equivocation gradually reduces. So stock prices generally undereact to new information initially, which is confirmed by empirical evidences (Hvidkjaer, 2006). This is consistent with the models of Barberis et al. (1998) and Hong and stein (1999). In the following, we will make further analysis of these two models.

Barberis et al. (1998) utilize the concept of conservatism to understand underreaction. Conservatism states that individuals update their beliefs slowly in the face of new information. This property is a natural result from formula (3). Barberis et al. (1998) attribute overreaction to representativeness heuristic. "People rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (Tversky and Kahneman, 1974, p.1124). Many investors don't want to spend tremendous resource to research fundamentals. They rely on a limited number of heuristic principles, such as technical signals and opinions from popular media, which reduce the complex tasks of assessing probabilities and predicting values of stocks to simpler judgmental operations with low cost. As we analyzed in the last section, this reliance on simple heuristic principles leads to overreaction in the asset market.

Hong and Stein's (1999) results are built on three key assumptions. The first two assumptions are that traders are classified as "newswatchers" and "momentum traders" according to their information processing abilities. They commented that, "the constraints that we put on traders' information-processing abilities are arguably not as well-motivated by the experimental psychology literature as the biases in Barberis et al. (1998) or Daniel et al. (1998), and so may appear to be more ad hoc" (Hong and Stein, 1999, p. 2145). These assumptions can actually be derived naturally from the entropy theory of information. Depending on the value of assets under management, different investors will choose different methods of information gathering with different costs. "Newswatchers" are large investors who are willing to pay a high cost to collect private information and to make a deep understanding of public information. "Momentum traders" are investors who 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 price momentum signals. Cohen, Gompers and Vuolteenaho (2002) show that institutional investors buy on fundamental news while individual investors buy on price trends. The third assumption of Hong and Stein (1999) is that private information diffuses gradually across the newswatcher population. The gradual diffusion of private information means that the number of people who enjoy low level of equivocation on some information gradually increases.

Both the reduction of equivocation of a representative investor and the increase of number of investors who have low level of equivocation on information contribute to the gradual reduction of underreaction, which generates momentum. Both representativeness heuristic and "momentum trader" can generate overreaction, which will lead to eventual reversal. The information theory can further distinguish the models of Barberis et al. (1998) and Hong and stein (1999). In Barberis et al. (1998), a representative investor make trading decisions. In Hong and stein (1999, 2003), investors are heterogeneous.

Investor heterogeneity can be understood naturally because of the different background of the investors and different cost that different investors are willing to pay to gather information. Empirical evidences show that investor heterogeneity exists in financial markets and plays an important role in the formation of trading patterns (Hvidkjaer, 2006).

From (3), the understanding of information depends on the background knowledge. Investors take longer time to understand information from sources they are less familiar with. Hong, Lim and Stein (2001) empirically confirm that information from small firms, from firms with low analyst coverage and from firms with bad news, which managers are reluctant to release, generally diffuse slower. From Hvidkjaer (2006), the selling pressures on loser generally are stronger and last much longer than buying pressures on winners, suggesting information processing is less efficient on bad news.

After discussing the existing behavioral models, Lee and Swaminathan summarized, "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). Trading volume, on the other hand, is an integral part of the informational theory of investment, which answers many questions on the gaps between existing theories and empirical evidences.

V. Conclusion

The emergence of the behavioral theories is a response to the failure of the efficient market theory, which states that financial markets can process information instantly. However, the links between behavioral biases and market patterns are often vague and arbitrary. We show that the new informational theory of investment provides a simple and unified understanding of the empirical patterns documented in the literature. The informational theory also offers deeper understanding to existing behavioral theories.

References

Adami, Christoph, 2004, Information theory in molecular biology, working paper.

Akerlof, Geroge, 1970, The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics* 84, 488-500.

Applebaum, David, 1996, *Probability and Information, an integrated approach*, (Cambridge University Press, Cambridge).

Arrow, K. (1983 (1973)). Information and economic behavior, in Collected Papers of Kenneth J. Arrow, Volume 4, Cambridge, MA: The Belknap Press of Harvard University Press.

Avery, John, 2003, Information theory and evolution, World Scientific, Hackensack, NJ

Baierlein, Ralph, 1999, *Thermal physics*, Cambridge University Press, Cambridge

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

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

Barberis, Nicholas and Robert Thaler, 2003, A survey of behavioral finance, in *Handbook of the economics of finance*, edited by G. Constantinides, M. Harris and R. Stulz.

Bennett, Charles H., 1988. Notes on the History of Reversible Computation. IBM Journal of Research and Development, 32, 16-23.

Bergstrom, Carl and Lachmann, Michael, 2004, Shannon information and biological fitness, Working paper.

Berk, Jonathan, and Richard Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269-1296.

Bloomfield, Robert, and Jeffrey Hales, 2002, Predicting the next step of a random walk: Experimental evidence of regime-shifting beliefs, *Journal of Financial Economics* 65, 397-414.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

Chan, Wesley S., Richard M. Frankel and S.P. Kothari, 2002, Testing Behavioral Finance Theories Using Trends and Sequences in Financial Performance, working paper.

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

Chen, Jing, 2005b, Information theory and market behavior, *The Institute of Chartered Financial Analysts of India (ICFAI) Journal of Behavioral Finance*, 2, No. 4, 25 - 45.

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.

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.

Coval, Joshua, and Tobias Moskowitz, 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* 54, 1-39.

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.

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

Durham, Gregory, Michael Hertzel, Michael, and Spencer Martin, 2005, The market impact of trends and and sequences in performance: new evidence, *Journal of Finance*, 60, 2551-2569.

Fama, Eugene, 1998, Market efficiency, long-term returns and behavioral finance, *Journal of Financial Economics* 49, 283-307.

Gallager, R. 2001, *Claude E. Shannon: A Retrospective on his life, work, and impact*, IEEE Trans. IT, vol 47,#6,

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

Grossman, Sanford and Joseph Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review*, 70, 393-408.

Gulko, Les, 2002, The entropy theory of bond option pricing, *International Journal of Theoretical and Applied Finance*, 5, 355-383

Haigh, Michael, and John List, 2005, Do professional traders exhibit myopic loss aversion? An experimental analysis, *Journal of Finance*, Forthcoming.

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

Hong, Harrison and Jeremy Stein, 2003, Differences of opinion, short-sale constraints and market crashes, *Review of financial studies*, 16, 487-525.

Hong, Harrison, Terence Lim and Jeremy Stein, 2000, Bad news travel slowly: Size, analyst coverage, and the profitability of momentum strategies, Journal of Finance 55, 265-295.

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

Isihara, A. 1971, Statistical physics, New York: Academic Press.

Ivkovic, Zoran, and Scott Weisbenner, 2005, Local does as local is: Information content of the geography of individual investors' common stock investments, *Journal of Finance*, 60, 267-306.

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

Jorion, P. and Goetzmann, W. 1999, Global Stock Markets in the Twentieth Century, *Journal of Finance*, 54, 953-980.

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.

Lee, Charles and Bhaskaran Swamminathan, 2000, Price momentum and trading volume, Journal of Finance 55, 2017-2069.

Malloy, Christopher, 2005, The geography of equity analysis, *Journal of Finance*, 60, 719-755.

Margulis, L., 1998. Symbiotic Planet, A new view of evolution, Basic Books, New York.

Maxwell, James, 1871, *Theory of heat,* (Green & Co., Longmans, London)

Morowitz, H., 1992, Beginnings of cellular life: metabolism recapitulates biogenesis, New Haven: Yale University Press

Odean, Terrance, 1999, Do Investors Trade Too Much, *American Economic Review* 89, 1279-1298.

Peng, Lin, 2005, Learning with information capacity constraints, Journal of Financial and Quantitative Analysis, 40, 307-330.

Peng, Lin and Wei Xiong, 2006, Investor attention, overconfidence and category learning, Journal of Financial Economics, 80, 563-602

Pinker, S., 1997, *How the mind works*, (W. W. Norton. New York).

Prigogine, I., 1980. From being to becoming : time and complexity in the physical sciences. W. H. Freeman, San Francisco

Qian, Hong, 2005, Relative entropy: Free energy associated with equilibrium fluctuations and non-equilibrium deviations, Working paper.

Roll, Richard, 1986, The hubris hypothesis of corporate takeover, *Journal of Business* 59, 197-216.

Schrodinger, E., 1944. What is Life? Cambridge University Press, Cambridge

Shannon, Claude, 1948, A mathematical theory of communication. *The Bell System Technical Journal,* 27, 379-423, 623-656.

Shannon, Claude, 1956, The Bandwagon (Editorial)*, IEEE Transactions Information Theory,* volume 2, p.3

Shannon, Claude and Weaver, Warren 1949. *The mathematical theory of communication*, Urbana: The University of Illinois Press.

Shanthikumar, Devin, 2004, Small Trader Reactions to Consecutive Earnings Surprises, working paper.

Shefrin, Hersh, 2005, A behavioral approach to asset pricing, Elsevier Academic Press

Shiller, Robert, 2000, *Irrational exuberance*, Princeton, N.J. : Princeton University Press.

Shleifer, Andrei, and Robert Vishny, 2003, Stock market driven acquisitions, *Journal of Financial Economics* 70, 295-311.

Sims, C. 2003, Implication of rational inattention, Journalof Monetary Economics, 50, 665-690.

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.

Figure Captions

Figure 1: Information value and probability

Figure 2: Overnight rate of return and trading volume of WestJet stock
          surrounding the date when Jetsgo announced bankruptcy