Chapter 1

The Entropy Theory of Human Mind

1.1. Introduction

People generally think that physical laws only have limited utility in understanding human behavior because our mind is free. However, human mind is shaped by natural selection and sexual selection (Pinker, 1997). Living organisms need to extract low entropy from the environment, to defend their low entropy sources and to reduce the diffusion of the low entropy. The struggle to stay in low entropy states is called natural selection. In human societies, agriculture is the main low entropy source. Part of health care systems aim at defending our own low entropy sources to be accessed by viruses and bacteria. The military forces are established to extract low entropy from others and to defend own low entropy sources. Clothing and housing reduces the diffusion of low entropy.

Sexual selection is the struggle between the individuals of one sex, generally the males, to communicate their attractiveness to the other sex in order to form a partnership for reproduction. Human beings, as well as other sexually reproducing species, are the successful descendants of the earliest sexually reproducing species about a billion years ago (Margulis, 1998). For the system of communication to be successful in different kinds of environments over such a long time, the mode of communication has to be simple, stable and universal. Since the entropy law, which states that closed systems tend towards states of higher entropy, is the most universal law of the nature, it is natural that the display of low entropy levels evolves as the universal signal of attractiveness in the process of sexual selection.

As both natural selection and sexual selection favor low entropy state, the pursuit of low entropy becomes the main motive of human mind and animal mind. Indeed the low entropy state is the main way of advertisement for most sexually reproducing species. Large body size, colorful and highly complex feather patterns with large amount of information content and exotic structures are all different representations of low entropy states. Since a low probability event corresponds to a state of low entropy, a novel feature is often attractive in the competition for reproduction. It has been generally recognized that sexual selection is the main drive of diversity (Miller, 2000).

Besides communication with members of the opposite sex, social animals need to communicate their attractiveness and power in order to influence the behavior of others (Wilson, 1975). For the same reason as in sexual selection, the most general signal is display of low entropy. Among all social species, human beings have developed the most complex social structure. The creation of distinct art works, the demonstration of athletic prowess, the accumulation of wealth, and conspicuous consumption - all of which represent different forms of low entropy - are the major methods of advertising one’s attractiveness.

As the social groups become larger and the division of labor becomes finer, people become less familiar with each other in their daily interactions, which make it more difficult for people to judge the ability of others. The need for people to advertise their attractiveness through external accumulation of low entropy also becomes stronger. People usually signal their capability by buying more expensive houses, cars, clothes, going to more expensive restaurants and attending more exclusive schools. The great efforts human beings put into non-food activities reflect the high cost of communication in a large and complex society. Historical evidences show that the transaction costs have been increasing over time (Wallis and North, 1986).

The main function of mind is information processing. The concept of information has been intimately related to entropy for over a century. In a thought experiment, Maxwell (1871) reasoned, if information is costless, the entropy of a system can be decreased. But this would violate the second law of thermodynamics. Maxwell went on to conclude that the physical cost of obtaining information must be at least as much as the value of information. Many years later Shannon (1948) identified information as entropy formally, at least at the mathematical level. (Shannon, 1956)

The remainder of the chapter is structured as follows. Section 1.2 introduces the generalized entropy theory of information. Information theory provides natural measures of the cost of obtaining information and of information asymmetry. Section 1.3 shows that entropy theory offers a unified understanding of the patterns of human psychology. Section 1.4 concludes.

1.2. What is Information?

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

$$H(P) = -\log_b P \qquad (1.1)$$

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 (1.1) 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

$$H(X) = -\sum_{j=1}^{n} p_j \log(p_j) \qquad (1.2)$$

The right hand side of (1.2), 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. (Theil, 1967; Maasoumi and Racine, 2002 and many others) However, the standard economic theory of information, represented by Grossman and Stiglitz (1980) was not built on the foundation of entropy theory. 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 concluded that information of higher value is of higher physical cost. Since economic cost is highly correlated to physical cost, (Georgescu-Roegen, 1971) more valuable information is in general more expensive to obtain. The relation among entropy, information and economic value will be discussed in greater detail in Chapter 2.

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

$$R = H(x) - H_y(x) \tag{1.3}$$

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 (1.3) 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.

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.

$H_y(x)$ in Formula (1.3) 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.

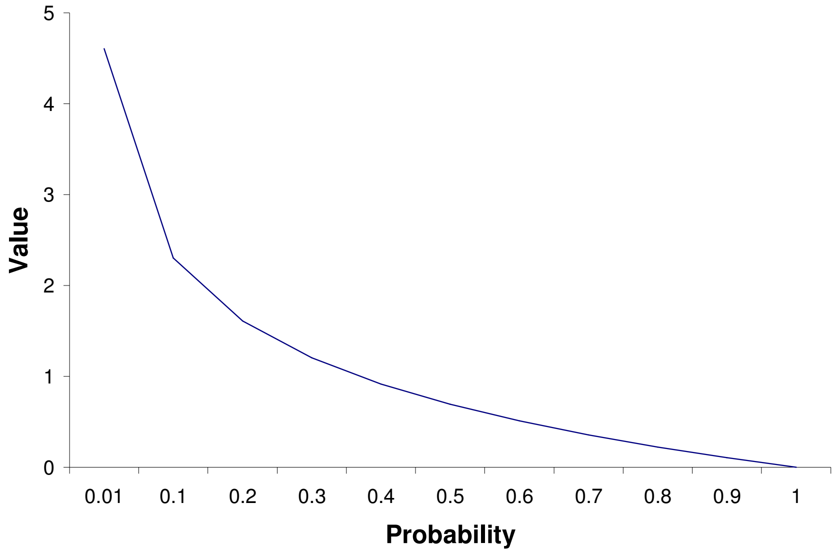

Figure 1.1 Information value and probability

Third, the same information, when known to more people, becomes less valuable. Figure 1.1 is a graph of (1.1), where H is a function of P, the probability of any given event. From Figure 1.1, 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 1.2 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 1.2 neatly illustrates the relation between value of some information and the number of people who know the information.

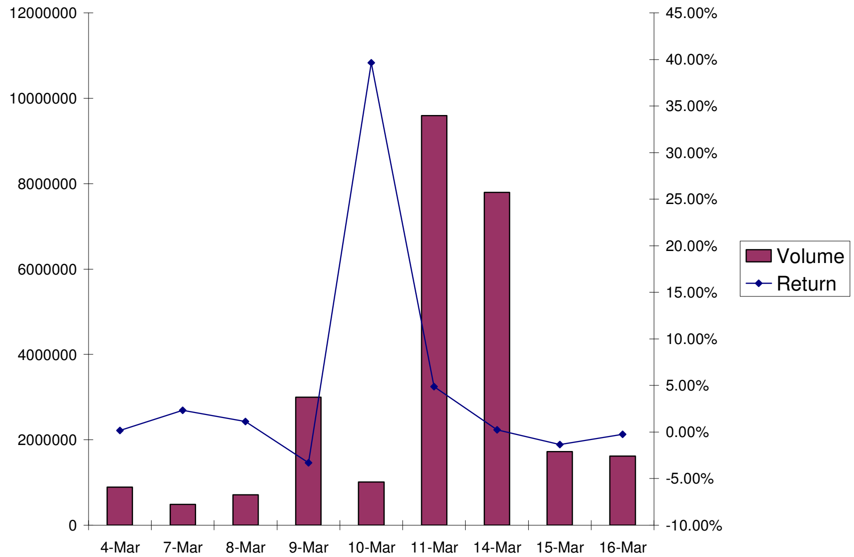

Figure 1.2: Overnight rate of return and trading volume of WestJet stock surrounding the date when Jetsgo announced bankruptcy

It is often said that the cost of information has dropped sharply over the years. But at the same time, the value of the same type of information has dropped sharply as well. Information of high value is usually carefully guarded and difficult to detect. For example, Warren Buffett, who has a very successful record for gaining and using insightful market information, would not announce to the public which stock(s) he is going to buy or sell. Animals have discovered this long ago. "In those cases where animal signals really are of mutual benefit, they will tend to sink to the level of a conspiratorial whisper: indeed this may often have happened, the resulting signals being to inconspicuous for us to have noticed them. If signal strength increases over the generations this suggests, on the other hand, that there has been increasing resistance on the side of the receiver." (Dawkins, 1999, p. 59)

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 in Chapter 6 support this statement.

When this theory was first proposed, it was treated as a natural extension of Shannon's entropy theory of information (Chen, 2002a). However, many have pointed out that Shannon (1956) would have a different view:

> 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)

Recent authority reaffirmed this orthodox view:

> 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)

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. In the next section and the rest of the book, we will show that establishing information theory on the foundation of statistical physics will yield great understanding of biological and social activities and turn "softer sciences" into a hard science.

1.3. The Entropy Theory of Human Psychology

Ideas occur at a blink of eye, which gives us impression that thinking is effortless. This seeming effortless, however, is the result of highly active brain system that consumes large amount of energy. Metabolically the brain is a very expensive organ. Representing only 2 percent of body mass, the brain uses about 20 percent of energy in humans (Pinker, 1997). Since mental activities are so costly, it is of great evolutionary advantage to have efficient information processing system to reduce the cost of thinking.

From the entropy law, we know that it is far easier for a system to disintegrate than to maintain its structure (Morowitz, 1992; Margulis, 1998). So there is a strong selective pressure for important knowledge to become genetically coded into heuristic principles to reduce the cost of learning (Tversky and Kahneman, 1974). Natural selection determines that human minds are born with many data and preferences stored (Pinker, 1997). In this section we will show that entropy theory offers a unified understanding of some frequently cited patterns of human psychology.

1. Conservatism

Conservatism in human beings may be characterized as behavior by individuals who possess a reluctance to update their beliefs in the face of new information. This property is a natural result from information theory. From Formula (1.3), the information one can receive is information sent minus equivocation, which is reduced gradually as the receiver's background knowledge about the source increases. Hence conservatism reflects the gradual reduction of equivocation by the receiver of any given information.

2. Herding behavior

From the second law of thermodynamics, a random action generally costs more than it gains. To concentrate actions into profitable ones, we, like wild animals, often learn from the experience of successful individuals and copy their behavior. It is generally very costly and impossible to repeat all of the experiences and mistakes that are possible. Therefore, we accept certain modes of behavior demonstrated by others without completely investigating the reasons behind them. Copying the actions of others directly is much easier, i.e., more efficient. Herding mentality developed because it is a cost-effective way of learning most of the time.

Human beings are social animals. Herding, or following the crowd, is good for survival. If you have walked alone in the wildness, you must have acute sense of vulnerability and powerlessness of human beings as individuals. Moose, bear and other animals can overrun human beings easily. It is only in crowds that we become powerful. So herding is essential for survival. A person who goes his own way usually cannot survive long. Herding mentality is evolved in this way.

It should be emphasized that all learning, especially institutionalized learning, are herding behavior. There are infinitely many things to explore. But we only have finite time and energy. The choice of subject to learn is from past experience. For example, when IT professionals earn high income, many people choose to get degrees in IT area.

3. Overconfidence and irrationality

"Extensive evidence shows that people are overconfident in their judgments" (Barberis and Thaler, 2003). From entropy law, any biological system, as a non-equilibrium system, faces constant dissipation of energy. Endless efforts are required to maintain a non-equilibrium system. Entropy law has been intuitively understood since ancient times. "The gods had condemned Sisyphus to ceaselessly rolling a rock to the top of a mountain, whence the stone would fall back of its own weight. They had thought with some reason that there is no more dreadful punishment than futile and hopeless labour. … If this myth is tragic, that is because its hero is conscious. … The workman of today works every day in his life at the same tasks and this fate is no less absurd. But it is tragic only at the rare moments when it becomes conscious." (Camus, 1955, p. 109) In the long course of evolution of our solar system, all life on earth will eventually go extinct in the far distant future (Lovelock, 1988). From a purely rational perspective, life is meaningless. Since human beings are self-conscious, the very question of why life is worth living lingers in many people's minds. "There is but one truly serious philosophical problem and that is suicide. Judging whether life is or is not worth living amounts to answering the fundamental question of philosophy" (Camus, 1955, p. 11). Overconfidence and irrationality are adaptive psychological traits that help us survive in this world.

The prevalence of irrationality is reflected in the prevalence of religious beliefs in various forms. A fundamental characteristic of various religions is that they are built on some miracles that are not consistent with physical or biological laws, such as virgin birth, sustainable growth or infinite human creativity (Daly, 1991). Marx (1844) once noted:

> Religion is the sigh of the exhausted creature, the heart of a heartless world and the soul of the soulless conditions. It is the opium of the people.
>
> The abolition of religion as the illusory happiness of the people is a demand for their true happiness. The call to abandon illusions about their condition is the call to abandon a condition that requires illusions.

Because of the inexorable increase of entropy in the universe, the condition that requires illusion will never leave us.

4. Loss aversion in winning and risk seeking in losses

Human beings often exhibit loss aversion in winning and risk seeking in losses. Kahneman and Tversky (1979) collected some responses to hypothetical choice problems. In one problem, the subjects were presented with two choices.

Choice A: There is an 80% probability of winning 4000 pounds and a 20% probability of winning nothing.

Choice B: There is a certainty of winning 3000 pounds.

The expected end wealth of choice A is 3200 and of choice B is 3000. Most respondents chose B, exhibiting loss aversion in winning. When the signs of the outcomes are reversed, the problems become the following:

Choice C: There is an 80% probability of losing 4000 pounds and 20% probability of losing nothing.

Choice D: There is a certainty of losing 3000 pounds.

The expected end wealth of choice C is -3200 and of choice D is -3000. Most respondents chose C, exhibiting risk seeking in losses. As money is a new invention in human evolutionary history, the preference for money must be derived from something else. Since food is the most important resource of our evolutionary past, our preference for wealth is probably derived from our preferences for food.

In the most part of the history of human evolution, we had not been able to store large amounts of extra food. If one goes without food for several days, he will starve. We translate the monetary numbers from the above four questions into days of food to obtain the following. In the case of gain, we can think of the choices of two possible strategies. In the first strategy, there is an 80% probability of getting food for 40 days and 20% chance of getting nothing. In the second strategy, there is a certainty of getting food for 30 days. It is easy to see why most people will prefer 30 days of food in certainty over a strategy that contains a 20% risk of getting nothing. In the case of loss, we can think of the choices of two possible strategies. In the first strategy, there is an 80% probability losing food for 40 days and a 20% chance of losing nothing for 40 days. In the second strategy, there is a certainty of getting no food for 30 days. Since without food for 30 days will represent sure death, people will naturally choose a 20% chance of survival. So people consistently avoid risk in both positive gain and negative loss. “Risk seeking” in loss is an unfortunate terminology borrowed from utility theory.

An important human institutional invention is the limited liability corporations, which alleviate people’s concern about unlimited loss. This greatly increases people’s willingness to explore new ideas and contributes to rapid economic growth. Human psychology has long been applied to the design of institutional structures.

5. Framing, representativeness and biases

We often frame, or sort problems into categories and assign different associated values based on the perceived relative levels of importance of each problem. Why do we do this? 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

$$-\sum_{j=1}^{n} p_j \log(p_j) \leq -\sum_{j=1}^{n} p_j \log(q_j) \qquad (1.4)$$

with equality achieved if and only if each

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

This result is called Gibbs inequality (Isihara, 1971). 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. The left hand side of Formula (1.4) 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 Formula (1.4) is smaller when q_j is closer to p_j. This means that information processing is more efficient when the subjective probabilities are closer to the objective probabilities. In particular, a mind with stored data about the natural environment is in general more efficient than a completely unbiased mind, where all subjective probabilities are to be learned from scratch. Natural selection determines that the human mind will evolve so that, "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).

Thus, "people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (Tversky and Kahneman, 1974, p.1124). What causes these severe and systematic errors? Human minds are the result of natural selection, which "operates over thousands of generations. For ninety-nine percent of human existence, people lived as foragers in small nomadic bands. Our brains are adapted to that long-vanished way of life, not to brand-new agriculture and industrial civilizations" (Pinker, 1997, p. 42). This is why we observe systematic errors in judgment by human beings, i.e., the typical framework for processing information today was developed over millennia when environmental conditions were very different. For example, most of us still have a great fear of snakes, although they rarely pose a threat to urban dwellers today. On the other hand, fear of electricity has to be instilled into children's minds with great difficulty (Pinker, 1997).

From Gibbs inequality, the level of uncertainty in understanding a type of events is

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

where p_i and q_i are objective and subjective probabilities respectively. Suppose this type of events has two possible outcomes, state 1 and state 2. 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 old probability, is

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

while the uncertainty of prediction by a novice is

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

This shows that when environment changes suddenly, novice actually perform better than experts, whose priori knowledge often cause severe biases in prediction. This is one reason why scientific revolutions are often initiated by and new industries are often pioneered by newcomers or outsiders (Kuhn, 1996; Stearns and Allan, 1996).

1.4. Concluding Remarks

From the discussion in the last section, we find that some psychological patterns, such as conservatism, reflect the constraints of thermodynamic laws. Others, such as framing and herding, are evolutionary adaptations to enable efficient processing of information, which is the reduction of entropy. Still others, such as overconfidence, irrationality and loss aversion, are mental attitudes that help us survive the constant dissipation of energy endured by all non-equilibrium systems. Therefore, entropy theory offers a unified understanding of human mind. This theory shows that mind, like matter, is governed by the same physical laws.