System Catastrophe: A Distributive Model for Collective Phenomenon

Chang, Fu and Luo, Jar-Der, 2002, "System Catastrophe: A Distributive Model for Organizational Collective Phenomenon." ***Journal of Mathematical Sociology Vol 26:1-17.***

Authors
Chang, Fu
Associate Research Fellow, Institute of Information Science, Academia Sinica

Luo, Jar-Der
Associate Professor, Graduate School of Information Sociology, Yuan-Ze University

Corresponding to:
Luo, Jar-Der
Graduate School of Information Sociology, Yuan-Ze University, Far East Rd. 135,
ChungLi, TaoYuan, Taiwan, R.O.C
e-mail: jdluo@saturn.yzu.edu.tw

Word count: 5523

Abstract

Key words: system catastrophe, collective efficiency, transaction cost, dynamic routing

The phenomenon of system catastrophe often occurs in a system with a network structure. A system’s resources can be utilized in two different modes: efficiently or inefficiently. When actions with inefficient mode pose no threat to other users or, in other words, when they employ resources that would otherwise be idle, they do not waste the system's resources at all. But when critical levels of inefficient uses of system's resources are reached, there is a sudden decrease in the capacity of the system due to the multiplication effect of inefficient factors. This collective inefficiency results in everyone getting worse in average. The common theme behind the catastrophe phenomenon demonstrates a possible explanation for the famous question about the choice between market and hierarchy. That is, when all firms pursue their own individual interests, resulting in a collective breakdown, they turn to consolidated ways of carrying out transactions.

I. Introduction

What is a phenomenon of system catastrophe? Taking a telephone network as an example, direct connection of a call is the most economic way to service it in terms of resources utilization. However, for a telephone user, it may not the only way to get service in terms of instant connection, especially when the direct route is busy. In general, the telephone network allows the call to be re-routed through an indirect route that is available at the time. For example, if the direct route from New York to San Francisco is busy, a call from NY to SF can choose the route from NY via Chicago to SF. Even though the indirectly routed call employs double resources in the network, the routing strategy is still the most efficient, given that the network is lightly loaded. However, when the network becomes busy, this call will "squeeze" out another call from its direct route. For example, a call from Chicago to SF may have to go through Columbus to make the connection, and utilize double resources in that network, also. As the network becomes busier, more calls are forced to go through indirect routes, and eventually an alleged "catastrophe" occurs, i.e. the collective efficiency of the network declines. That is to say, a network with a capacity of two million calls is now able to carry only 1.2 million calls. Waste of network resources may not be very prominent at first, but when system catastrophe occurs, there is a sudden decrease in network capacity. This collective inefficiency results in every call getting worse service in average.

We live in a world in which resources are shared. Resources, however, are not available to all the people at all the same cost. Geographic distance can be one factor that renders a type of resource (rice, for example) cheap to some people and expensive to others. We can model the world with a network whose nodes represent the places where goods are created (the origins) or where they are sent (the destinations). The links of the network, on the other hand, represent the availability of resources from place to place (origin to destination). The cost asymmetry as stated above can then be expressed by the fact that a one-link path (or direct path, i.e. a path consisting of only one link) represents the cheapest means for conveying resources, and that a two-link path is the next cheapest means, and so on.

A peculiar aspect of the network environment is that each direct path can overlap with many two-link paths. For example, in Figure 1, the direct path between A and B overlaps the two-link path connecting A and C via node B, and it also overlaps the path connecting A and D via B.

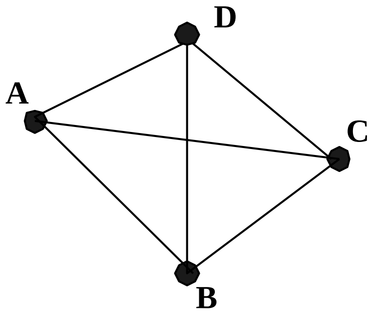

Figure 1. A network model.

Thus in a system with a network structure, the resources of communication channels is also shared among all possible users of the network. The resulting situation is rather complex. For goods to be transported from one place to another, say from A to C, the cheapest rate would be to go through the path that directly connects A and C. If, however, this path is fully occupied while links AB and BC are not, then it would be worthwhile to go through the two-link path that passes through B.

This strategy, however, is optimal to users *only* when they do not have any idea about the traffic conditions in the network. If all users applied this strategy, ignoring what others are doing, then the resources of that network will not be used efficientally.

Just think about what would happen if too much traffic poured into the network blocking up all direct paths. Then all individual users who are unable to cross their direct paths would be competing to use some two-link path. Let us assume that one such user wants to ship its goods from A to C, and is lucky enough to find the indirect path ABC open at the moment. Having sent away its goods on that path, this aggressive user will immediately block the way of any potential user who wants to ship their goods from A to B, as well as blocking the way of any other user who wants to ship goods from B to C.

Inefficient utilization of resources does not stop at this point. The users who are unable to follow their direct paths will proceed to make an attempt at securing some other indirect path. Each successful action as such will further increase the waste of the system’s resources, in the sense that two units of resources are used to serve one customer instead of two. The chain reaction goes on, resulting in a tremendous waste of system resources.

Indeed, such inefficient use of a system’s resources has been shown to take place in a paper entitled "The Overload Performance of Engineered Networks with Nonhierarchical and Hierarchical Routing" (Akinpelu, 1984). This paper notes that in a telephone network capable of nonhierarchical (dynamic) routing, a telephone call can either take a direct path or an indirect path. When a call arrives and its direct path is blocked at that moment, the network allows the call to be re-routed through an indirect path. In this simulation study, the total load being introduced into the network was set at a sufficiently high level so that, with the slight addition of a few calls, the network became overloaded. When it did, the study shows a catastrophic phenomenon occurred. Waste of system’s resources may not be evident in the beginning, but when critical levels are reached, there is a sudden decrease in the capacity of the network.

This type of congestion not only occurs in telephone traffic, but also in highway traffic. When automobiles want to change lanes, they behave like indirectly routed calls: they essentially take away both the road space they intend to leave and also the space they intend to settle in. In highway traffic, reduction of the system’s efficiency is likely to occur when two automobiles want to compete for the same space. In this scenario as well, a chain reaction can easily be set off. When one car changes lanes, it causes neighboring cars to do the same, which in turn affects their neighbors, and so on. It is changing of lanes that can decrease the traffic flow. The simple fact that there is too much traffic on the road can have the same effect. An automobile, even without changing lanes, requires a safety distance between it and the nearest vehicles to maintain its speed up at a certain level. When this safety distance is encroached upon by another automobile, the car’s speed will also be lowered in order to avoid unanticipated events. When one car slows down, it also slows down the cars behind it. Thus, if the automobiles move along in lines on a road, the diminished speed of any automobile will cause a diminished speed in the automobiles behind it.

These two samples illustrate one common theme. A system’s resources can be utilized in two different modes: efficiently or inefficiently. In the telephone network, a call that is put through a direct path makes efficient use of the system’s resources. When a call is put through an indirect path and takes away the ability of some other users to make a direct connection, it makes inefficient use of the system’s resources. In the highway system, when an automobile shifts to occupy a new space, hindering other automobiles from keeping up their normal speed, it is making inefficient use of that system’s resources.

Having said this, we do not imply that indirect telephone calls or automobiles that change lanes always act inefficiently. When their actions pose no threat to other users or, in other words, when they employ resources that would otherwise be idle, they do not waste the system's resources at all. Thus, for users of a system to act optimally, they are required to switch modes dynamically between an efficient mode and an inefficient mode. How to switch modes without curbing the overall performance constitutes the reason why these systems as they appear are difficult to regulate.

The theoretical implications behind the above reasoning are also significant and provocative in terms of social phenomena. They illustrate a scenario in which the pursuit of maximal interest of individuals involved in an inefficient mode for getting resources may result in a system breakdown, despite the fact that the same behaviors may bring benefits under a different set of circumstances. If we consider a telephone network as a special case of a transaction system, then the conclusion one draws for the former can be generalized to the latter. In terms of the problems that arise in dynamic routing, an effective means of avoiding the catastrophe is to alter the way system resources are distributed. We will argue that this is exactly the approach that many transaction systems have used upon realizing severe competition for production resources was yielding similar consequences.

II. Modeling System catastrophe

Some good analytical works have been established (Markbukh 1981, 1983; Mitra and Gibbens 1992; Mitra, Gibbens and Huang 1993) to model and analyze telephone networks with the aforementioned routing capability. These works involve complex mathematical modeling and focus on highly technical issues. In this article, we want to consider the problem from a broader viewpoint. We thus develop a model, which we believe provides better intuition about the social aspect of the phenomenon, and also demonstrates the fact that over-competition for system resources can drive the whole system to collapse.

Since the social aspect of the phenomenon is the core of our concern, we will use factories and suppliers as the basic elements in our model. Thus, we assume that there are J factories and J up-stream suppliers in a transaction system. Each factory F_j is *associated* with a supplier S_j (F_j and S_j may either belong to the same hierarchical system, or S_j is a subcontractor of F_j in a network) for j = 1, 2, ..., J. Here in our model, we do not assume that S_j and F_j hold such a rigid relationship that F_j obtains its complete supplies from S_j. We only assume that S_j is more efficient in processing orders from F_j than those from F_k, for $k \neq j$.

For this reason, we say that the resources of S_j is divided into two separate lines, the *shared* line, rendering service to all factories, and the *consolidated line*, rendering service to F_j only. Since the shared line of each supplier is open to all factories, we consider that there is *only one* shared line in the whole system, although its resources are drawn from different suppliers. In quantitative term, our first assumption can be expressed in the following statement.

(1) Supplier S_j requires one unit of resources to process the orders of factory F_j, and E units for processing the orders from F_k, for $k \neq j$. Moreover, $E > 1$.

Note that it is *not* assumed in our model that the shared line and consolidated lines are physically or functionally separated. Their difference lies in the efficiency of fulfilling orders. Thus, in the telephone network, the connection of a call via direct path (consolidated line) employs one unit of system resources and is more efficient than the connection via an indirect path (shared line) that employs double resources.

Likewise, meeting the demands of familiar (or trusted) customers can be more efficient than meeting the demands of average customers, due to the acquaintance with the partners' ways of doing business and handling legal matters of transactions.

The rest of our assumptions are given as follows.

(2) Each factory F_j first sends its orders to S_j. But, if the order has not been completed by S_j, F_j will appeal to the open market (the shared line) to complete its unfulfilled orders. Those orders that are sent to the consolidated lines will be referred as consolidated-line demands and those that are overflowed to the open market as shared-line demands.

(3) The open market provides better profits to induce suppliers to set higher priority towards its demands. Thus, suppliers devote their resources to shared-line demands first, and only employ their leftover capacities to meet consolidated-line demands.

(4) Each supplier evenly assumes the burdens of shared-line demands.

(5) Each supplier processes orders at a rate of P at each time period.

(6) The orders issued from each factory form a uniform and continuous stream, at the rate of O at each period.

Having described the general setting of the model, we now stipulate the following scenario. The orders of each retailer, in general, arrive at a rate lower than the processing speed of the factories, that is, if we set $\delta = P - O$, then $\delta > 0$. At the initial unit of time t_0, however, each supplier S_j receives P + ε orders from retailer F_j, and thereby an excessive amount ε is unprocessed by S_j at the end of time t_0 and will be redistributed to all other suppliers.

Let $U_j(n)$ denote the amount of orders unprocessed by S_j at the end of time t_n, for n = 0, 1, 2, ... Then for j = 1, 2, ..., J,

$$U_j(0) = \varepsilon ,$$

At the next time unit t_1, S_j receives 1/(J-1) proportion of unprocessed jobs from each of the other J-1 factories. It thus receives $U_j(0)$ unprocessed jobs in total. These jobs become shared-line demands and will require $U_j(0)E$ resources to process. Moreover, since supplier S_j first devotes its resources to satisfy this need, it has only P - $U_j(0)E$ resources left for the consolidated-line demands. Each consolidated-line demand requires one unit of resources, so the unfulfilled jobs of S_j at time t_1 becomes

$$U_j(1) = O - (P - U_j(0)E) = U_j(0)E - \delta = \varepsilon E - \delta ,$$

As time goes by, the quantity U_j evolves as follows.

$$U_j(2) = U_j(1)E - \delta = (\varepsilon E - \delta)E - \delta = \varepsilon E^2 - \delta(E+1) ,$$

$$U_j(3) = (\varepsilon E^2 - \delta E - \delta)E - \delta = \varepsilon E^3 - \delta(E^2 + E + 1),$$

$$\vdots$$

$$U_j(n) = \varepsilon E^n - \delta(E^{n-1} + E^{n-2} + \ldots + 1), \tag{1}$$

$$\vdots$$

Moreover, (1) can be transformed into

$$U_j(n) = \varepsilon E^n - \delta \frac{E^n - 1}{E - 1},$$

and further into

$$U_j(n) = \frac{(\varepsilon(E-1) - \delta)E^n + \delta}{E - 1}. \tag{2}$$

Note that (1) and (2) are valid only when the following conditions hold together. (i) The right-hand side of the equality is positive. If the value is negative, then $U_j(m)$ becomes zero for all $m \geq n$. (ii) $U_j(n-1)E < P$.

We now say that *system catastrophe* occurs if $U_j(n)E \geq P$ at some time period t_n. Why do we stipulate this condition? Because, when it holds, the network would use up all its resources (at moment t_n) for serving shared-line demands, leaving no resources at all for any consolidated-line demands.

Note that, in stating the above scenario, we have assumed that the processing order P is always greater than the amount of orders O, except at the very first stage t_0 when there is an excess of orders in the amount of ε. Let us now further assume that this excessive amount is so small that the following inequality holds: $\varepsilon E < P$.

We now examine the following two alternative possibilities.

Case 1: $\varepsilon(E-1) \leq \delta$.

Obviously,

$$U_j(0)E = \varepsilon E < P.$$

Furthermore, $U_j(1)E$ is either 0, or

$$U_j(1)E = (U_j(0)E - \delta)E = (\varepsilon E - \delta)E \leq \varepsilon E < P.$$

Likewise, $U_j(2)E$ is either 0, or

$$U_j(2)E = (U_j(1)E - \delta)E = (\varepsilon E - \delta)E \leq \varepsilon E < P.$$

If we go on like this, we can prove that $U_j(n)E < P$ for every n. We thus conclude that system catastrophe will never occur in this case. ∎

Case 2: $\varepsilon(E-1) > \delta$.

The right-hand side of (2) is always positive, and the value of $U_j(n)$ grows at a exponential rate of n. Thus, sooner of later $U_j(n)E$ will exceed P and the catastrophe ensues. For this reason, "$\varepsilon(E-1) > \delta$" will be referred to as catastrophic condition. ∎

We are thus facing two alternative situations (a bifurcation). When $\varepsilon > \delta/(E-1)$, system catastrophe occurs. That is, when the initial excess of order ε rises above the level of spare capacity δ to such an extent, the system resources will be gradually eaten up by overflowed jobs (i.e., the unprocessed jobs that flow back to the network). Moreover, if we look at (2), the amount of overflowed jobs grows exponentially. Thus, very quickly, the system resources will be completely used up by them. On the other hand, when $\varepsilon \le \delta/(E-1)$, no such thing happens. Moreover, the amount of unprocessed jobs will either remain constant, when $\varepsilon = \delta/(E-1)$, or decays exponentially to zero, when $\varepsilon < \delta/(E-1)$.

To obtain the same conclusions by way of a more graphical method, we note that the accumulation of unprocessed jobs behaves like the following iterative process.

$$U_j(n) = O - (P - U_j(n-1)E) = U_j(n-1)E - \delta$$

$$= F(U_j(n-1)),$$

where $F(x) = xE - \delta.$

As is illustrated in Figure 2, the "equilibrium point" of this iterative process is just the intersection point of the line y = F(x) and the diagonal line y = x, and can be found by solving the following equation.

$$x = F(x).$$

The solution is found to be $\delta/(E-1).$ Moreover, this equilibrium is *unstable* for the following reasons. If the amount of unprocessed jobs starts to be a quantity ε_1 that is greater than $\delta/(E-1)$, as shown in Figure 2, then the amount of unprocessed jobs will spiral upwards till the value P is reached. On the other hand, if the amount of unprocessed jobs starts to be a quantity ε_0 smaller than $\delta/(E-1)$, then it will spiral downwards till the value zero is reached. Thus, any slight deviation from the equilibrium point will drive the total amount of unprocessed jobs to either increase or decrease, proving that we have an unstable equilibrium. Moreover, as implied by formula (2), the rate of deviation is exponentially fast. The instability of the equilibrium has, of course, to do with the fact that the slope E of the line y = F(x) is greater than that of the diagonal line, which is 1. If E is a quantity smaller than 1, then $\delta/(E-1)$ would become a stable equilibrium.

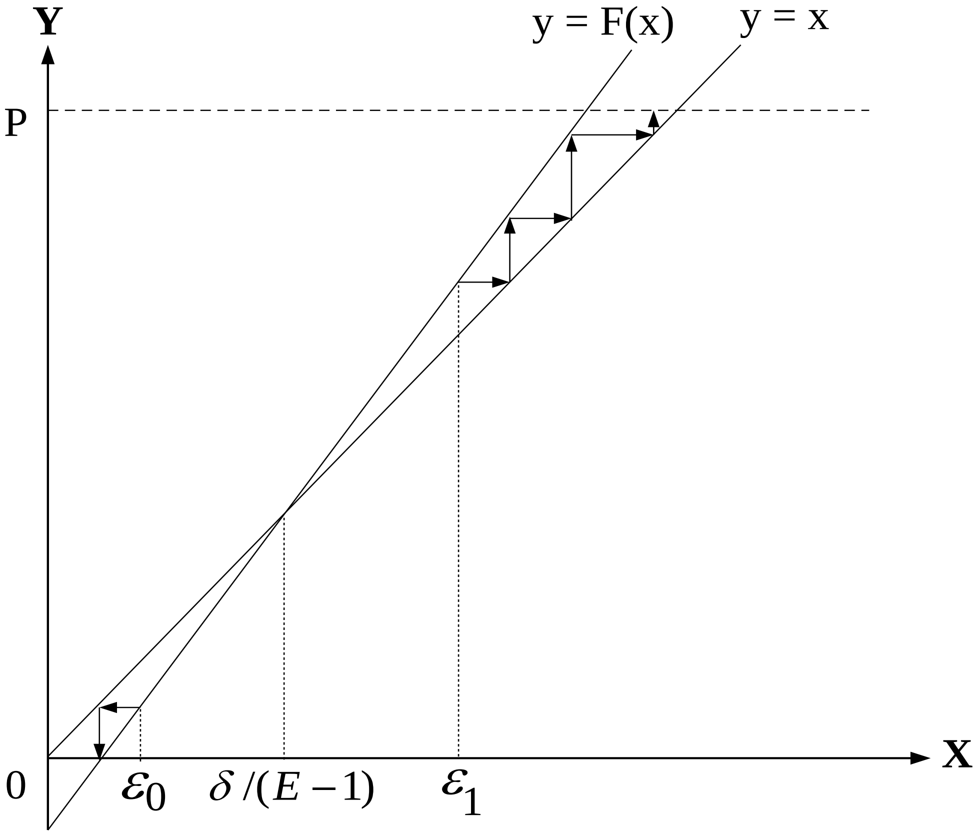

Figure 2. The iterative process.

Let us now see what would happen if a different strategy is taken by the suppliers, under the same condition that the initial amount of unprocessed jobs is ε. We assume that all suppliers are constrained to receive orders *only* from their associated factories, then the inefficiency factor does not play any role and the unfulfilled orders will be eventually absorbed, no matter how large the amount is initially.

To prove the last assertion, we note that even though there is an initial access of orders in the amount of ε, there is nevertheless spare capacity δ at each subsequent time unit. Since the amount of unprocessed jobs remains constant (it never gets multiplied), it will be absorbed at time unit t_n with $n\delta > \varepsilon$. So, it becomes obvious from the above argument that when initial amount of unprocessed jobs exceeds the critical point $\delta/(E-1)$, a better strategy for suppliers to take is to switch their strategy from an aggressive mode (taking jobs from any factories) to a conservative mode (taking jobs only from their alliance factories).

Let us consider one more scenario. We assume that orders from factories arrive at the same rate as the processing speed of suppliers, namely, P = O. At the initial time

t_0, however, there is an excess of orders in the amount of ε. From (1), it follows that

$$U_j(n) = \varepsilon E^n ,$$

for every n > 0, since $\delta = P - O = 0$ under this new assumption.

Thus, the quantity U_j gets multiplied at each time unit. On the other hand, if we assume that all suppliers receive orders only from their associated factories, then a totally different outcome would ensue: the number of unprocessed jobs remains constant ($=\varepsilon$) for all the time! Here, again, a great dichotomy lies between the instance in which unprocessed jobs flow to the open market (shared line) and that in which they do not.

Despite the simplicity, our model is useful for revealing the fact that “multiplication effect” in a market system is responsible for the onset of undesirable consequences. The key ingredient of this model is the asymmetry between the cost of doing business with familiar (or trusted) customers and with average customers. Moreover, the cost asymmetry can be multiplied when the suppliers become overloaded.

In view of this result, the best strategy during times of congestion is for individual firms to restrict their choices of business partners in the transaction system. Consolidation, rather than open market, is the best strategy when system’s resources become severely constrained. Thus, as suggested by this analysis, reduction of collective inefficiency constitutes the motivation for the firms to adopt consolidated behavior, in the forms of either hierarchies or networks, as observed in many industries. Note that a similar strategy, called *trunk reservation strategy*, is utilized in managing telephone networks in order to correct the ill effects caused by dynamic routing (Krupp 1982). Such a strategy has the effect of excluding overflowed calls from any one-link paths that are already heavily loaded. Thus, for example, when the direct path between NY and SF is near full occupation, it only accepts those calls between NY and SF.

Although mathematically assured, the possible occurrence of catastrophe during congestions may still take some people with surprise. It may be felt that the brief upsurge of demands should stimulate the growth of supplies that in turn helps diminish the unfulfilled orders. Our model, on the other hand, asserts the opposite. The reason this model predicts differently comes from a special assumption: there is a tightness of capacity in the network. Such tightness serves as a barrier on which the excessive demands keeps making stronger and stronger rebounds, thus generating undesirable consequences. While this assumption is certainly true of telephone network, since its capacity can not be expanded randomly upon any transient upsurge of call requests, some people may wonder: is the same hypothesis still true when applied to the business realm? Before we answer this question, we first note that our conclusion does not hinge upon the fixed amount of total resources (in our case, the quantity P). In fact, the result only relies upon the fixed amount of δ, i.e., the difference between demands and supplies. In fact, we now show that even this constancy assumption is not necessary for deriving the catastrophic result.

Let us consider the following scenario, which may better reflect what happened in the early phase of many industries.

(a) An excessive demand ε occurs at the initial time.

(b) Both the supplies P and demands O grow with time.

(c) However, the growth of P lags behind that of O. That is, if $\delta_n = P - O$ at the nth time unit, then δ_n decreases with n. This assumption, we might say, reflects the tightness of resources on the supply side.

(d) The following catastrophic condition holds: $\varepsilon(E-1) > \delta_1$. Note that this is similar to the condition of case 2, with δ now replaced by δ_1.

From the above assumptions, the amount of unfulfilled orders at the end of t_n can be derived as

$$U_j(n) = \varepsilon E^n - (\delta_1 E^{n-1} + \delta_2 E^{n-2} + \ldots + \delta_{n-1}).$$

Moreover, the quantity $U_j(n)$ increases monotonically with n, since

$$U_j(1) = \varepsilon E - \delta_1 > \varepsilon = U_j(0),$$

$$U_j(2) = U_j(1)E - \delta_1 > \varepsilon E - \delta_1 = U_j(1),$$

$$U_j(3) = U_j(2)E - \delta_2 > U_j(1)E - \delta_1 = U_j(2),$$

.

.

Furthermore,

$$U_j(n) \ge \varepsilon E^n - \delta_1(E^{n-1} + E^{n-2} + \ldots + 1) = \frac{(\varepsilon(E-1) - \delta_1)E^n + \delta_1}{E-1}.$$

Since $\varepsilon(E-1) > \delta_1$, the amount of unfulfilled orders grows at least exponentially fast. Note that, in deriving the above results, we only use the fact that δ_n is decreasing, and do not rely upon the growth of demands and supplies at all. But both conditions are what most likely happened in the early period of industries. That is, after a brief upsurge of demands, both the supplies and demands grow but the growth of the former lags behind that of the latter. We shall look for theoretical implications of this model in the next section.

III. Theoretical Implications for Organizational Studies

The common theme of the above analytical models demonstrates a possible explanation for the famous question about the choice between market and hierarchy (or network; Williamson 1975, 1981; Powell 1990). That is, when all firms pursue their own individual interests in open market, resulting in a collective breakdown, they turn to consolidated ways of carrying out transactions. This reveals an important phenomenon in transaction systems: turning to the open market may further decrease efficiency in fulfilling customer demands. The inefficiency factor comes from two sources:

(i) There is originally a cost asymmetry between open-market business relations and consolidated relations--that is, high transaction cost.

(ii) The inefficient factors can be multiplied when a system's resources become extremely tight, while competition for them remains unrestricted.

The multiplication effect is shown to exist when a market is gripped by a low supply of resources, causing the growth of supply to lag behind that of demand. When system catastrophe occurs, the theory predicts the following results.

(1) The system‘s collective efficiency will decline sharply.

(2) A competitive edge gained through consolidated relations starts to manifest itself.

(3) The higher the original cost asymmetry is, the sooner and the more significant the catastrophic effect appears.

In brief, turning to the open market is a good strategy for maximizing each individual's interests when system resources are abundant. However, when resources become tight and the multiplication effect sets in, consolidation among business partners makes for a better strategy than the open pursuit of that system's resources. Consolidated trading partnerships include two categories—those which occur in hierarchies and those in so called "network organizations." In the following sections, we will demonstrate how the system catastrophe theory can interpret the rise of network organizations.

The rise of network organizations has been seen in many industries since the 1980s. Strategic alliances and subcontracting systems have become a widespread phenomenon among US corporations. Small and medium sized firms relying on a network form of organization gained increasingly important status not only in terms of their numbers, but also in terms of the quality of goods and services produced. These organizations were responded for 19 million new jobs created in the US during the 80s, especially those in high-tech industries (Case 1992). The structure of the computer industry took a surprising turn from near-monopoly to intense competition with the formation of numerous alliances, following IBM's loss of hegemony in the 80s. Biotechnology followed a similar path, in which small research teams and biotechnology firms, which constituted roughly 50% of the industry, established strategic alliances with hospital, chemical, pharmaceutical, or energy companies in order to test, produce, and market their innovations (Barley et. al. 1992).

Why do firms prefer to establish long-lasting business relationships with trading partners rather than venture into the open market?

Several theories have sought to explain the competitive edge of network organizations. The most popular explanation for the effectiveness of network organizations is the doctrine of flexible specialization. Piore and Sabel first proposed a “dualism-like” organizational theory which argued that multi-divisional and multi-functional firms are well-suited for mass-production, whereas changing and fragmented markets are best tackled by flexible specialization (1984). Since the mass production economy has been in decline following the energy crises of the 1970s, flexible and specialized production, among other strategies, has seemed to provide solid hope towards building future economic prosperity. In particular, the contribution of subcontracting networks to flexible production has been thoroughly studied (such as Baker 1992; Hamilton and Kao 1990; Ka 1993). These studies have found that because small units have a shorter command hierarchy, flexible structure, less bureaucratic regulations, direct access to market information, and a narrow gap between “conception and execution” (Perrow 1992), they are good at responding to the demands of irregular markets quickly. Subcontracting networks provide small units with forward and backward linkages, in which small units can concentrate their limited resources on only one small phase of the production/marketing process. Consequently, network forms of organization naturally tend to breed specialization through a large number of flexible small firms (Luo 1997, 1998).

Taking a different approach from the above-stated theory, which emphasize firm-level advantages, collective efficiency offers another explanation of competitive edge of network organizations. In answering why the Hewlett-Packard Company (HP) surpassed Digital Equipment Corporation (DEC), Saxenian (1994) attributed Hewlett-Packard’s performance not to micro-level managerial or strategic factors, but rather to macro-level regional advantage. In a comparison of the development trajectories for Silicon Valley (where HP is based) and Route 128 (where DEC is based), Saxenian attributed the success of Silicon Valley to its different style of doing business. Namely, Silicon Valley forms subcontracting networks and alliances, rather than building a vertically integrated bureaucratic structure in an open-market environment. Similar regional advantages have been found not only in the high-tech industry of Silicon Valley, but also in other industries, such as knitwear production in Modena, Italy (Lazerson 1988), precision machine tool manufacturing in Baden-Wuerttemberg, Germany (Piore and Sabel 1984), and personal computer production in Northern Taiwan (Chang and Kao 1996), etc. A widely accepted explanation for regional collective efficiency points to the effect of knowledge diffusion. The influence of informal relationship networks to information flow has been observed by many sociologists (such as Granovetter 1973; Burt 1992). Everett Rogers also realized how important diffusion networks are in adopting new innovations (1995), and thus keeping regional technology state-of-the-art. He attributes the flourishing of diffusion networks in Silicon Valley to the high-frequency of personal interaction there, which facilitates inter-professional and inter-disciplinary communication (Rogers and Larsen 1984). Saxenian also pointed out the importance of Silicon Valley‘s social life, city design and professional associations in encouraging the development of such personal interactions, which in turn often stimulates knowledge exchange and entrepreneurship in the informal arena. Such networking among corporations helps to build a vibrant diffusion network (1994).

System catastrophe points to an alternative explanation based on collective efficiency of a system, too. It is an extension of William's Transaction Economics, in which high transaction cost is taken as the key to explaining why firms do business in hierarchies instead of in markets. Similarly, the cost asymmetry between networks and markets is the basic incentive that encourages firms to turn from open markets to consolidated partnerships in the catastrophe theory. However, Transaction Economics focuses on micro-level analysis. In its explanation, when the inefficiency factor (E) of a transaction is greater than one (refers to Assumption 1 in page 4; in other words, transaction cost is high), an individual firm may direct that transaction into a consolidated channel (i.e. hierarchies in Williamson's case). We found that firms actually respond to this inefficient trading method collectively when $\varepsilon(E-1) > \delta$ (refers to Case 2 in page 6), since system catastrophe heavily influences an individual's efficiency beyond this critical point. This model suggests three theoretical implications:

(a) The higher the cost asymmetry between consolidated partners and open markets is, the earlier catastrophe phenomenon will occur.

(b) Inability of a system’s resources to grow in pace with the demands of its users sets off multiplication effects pertaining to cost asymmetry, and collective efficiency of the open market system declines sharply.

(c) Based on the above two factors, the competitive edge of consolidated trading channels emerges suddenly following this critical point

In other words, Transaction Economics proposes the condition E > 1 for individual firms choosing between consolidated channels (hierarchies in its theory) and markets. But when E > 1 and $\varepsilon(E-1) \leq \delta$, the best choice for an individual firm may not be to succumb to the pressure of collective behavior. Only when catastrophe conditions $\varepsilon(E-1) > \delta$ emerge can the whole system restructure itself.

While further empirical testing against data is necessary, catastrophe phenomenon can be taken as a complementary theory to the doctrine of flexible specialization. Piore and Sabel (1984) proposed three kinds of environments which best suit for the needs of network organizations (also pointed out by Baker 1992).

1. A complex market, involved with complex knowledge content.
2. A fragmented market.
3. A turbulent and changing environment.

System catastrophe is a detailed explanation of this famous observation, since it demonstrates the cost asymmetry between networks and markets as well as system inefficiency in the three conditions:

(1) Complex and changing technology means high transaction costs in searching for information and supplies. A subcontracting network is better than open market system because it is able to reduce the cost of searching for information.

(2) A fragmented market partitions resources available for the system. Different types of inventory may pile up according to the demands of the fragmented market. This causes high demand for supplies and raises the cost of inventory, especially when system resources are not available for keeping factories' assembly lines open. Consolidated behavior can effectively reduce the cost of inventory.

(3) With rapid change, a market becomes unstable and its trends become difficult to predict in the long-term, resulting in the inability of suppliers to meet market demand, particularly when heavy demand emerges suddenly in an uncertain market. In this situation, firms need to maintain an inventory to protect themselves from the impact of market uncertainty, and consolidated networks help reduce this high inventory cost.

In order to make more precise conclusions than those we have covered in this article, this model certainly requires further elaboration. For example, we assume hierarchies and networks to be one in the same, just as Williamson did. However, Powell (1990) pointed out that networks must be considered as a separate category, distinguished from hierarchies. We chose to ignore this distinction in our model for the sake of simplicity. The model upon which we based sufficiently served our purposes, namely, to explore the causes and impacts of system catastrophe. In no way is our model intended to exclude other possible explanations of the consolidated behavior of organizations. Indeed, research in the area of collective efficiency has only just begun, and there are many other aspects of this subject that have yet to be explored. It is our hope that this paper will serve as a catalyst in sparking such further researches.

References

Akinpelu, J. M. 1984. “The Overload Performance of Engineered Networks with Nonhierarchical and Hierarchical Routing.” *AT&T Bell Laboratories Technical Journal* 63(7):1261-81.

Baker, Wayne. 1992. “The Network Organization in Theory and Practice.” Pp.:397-430 in *Networks and Organizations,* edited by Nitin Nohria and Robert G. Eccles. Cambridge: Harvard Business School Press.

Barley, Stephen, John Freeman, and Ralph Hybels. 1992. “Strategic Alliances in Commercial Biotechnology.” Pp.:311-348 in *Networks and Organizations,* edited by Nitin Nohria and Robert G. Eccles. Boston: Harvard Business School Press.

Burt, Ronald. 1992. *Structural Holes: The Social Structure of Competition.* Cambridge: Harvard University Press.

Case, John. 1992. *From The Ground Up: The Resurgence of American Entrepreneurship.* NY: Simon & Schuster.

Chang, Wei-An and Cheng-Shu Kao. 1996. “State and Business -- An Analysis of the Rise of the Semiconductor Industry in Taiwan.” Presented in International Conference of Economic Governance and Flexible Production in East Asia.

Granovetter, Mark. 1973. “The Strength of Weak Tie.” *American Journal of Sociology* 78: 1360-80.

Hamilton, Gary G. and Cheng-Shu Kao. 1990. "The Institutional Foundations of Chinese Business: The Family Firm in Taiwan." *Comparative Social Research* 12: 135-151.

Ka, Chih-Ming. 1993. *Market, Social Network, and the Production Organization of Smaill-Scale Industry in Taiwan---The Garment Industries in Wufenpu.* Taipei: Institute of Ethnology, Academia Sinica.

Lazerson, Mark. 1993. “Future Alternatives of Work Reflected in the Past: Putting-Out Production in Modena.” in *Explorations in Economic Sociology,* edited by Richard Swedberg. NY: Russell Sage Foundation.

Luo, Jar-Der. 1997. “The Significance of Network in the Initiation of Small Businesses in Taiwan.” *Sociological Forum* 12(2):297-317.

Luo, Jar-Der. 1998. “The Savings Behavior of Small Investors—A Case Study of Taiwan.” *Economic Development and Cultural Change* April, Forthcoming.

Marbukh, V. 1981. “An Asymptotic Study of a Large Fully Connected Communication Network with Re-routes.” *Problemy Peredachi Informatsii* 3: 89-95.

Marbukh, V. 1983. “Study of a Fully Connected Channel Communication Network with Many Nodes and Circumventing Routes.” *Automatika i Telemakhanika* 12: 86-94.

Mitra, D., Gibbens, R. J. and Huang, B. D. 1993. “State-Dependent Routing on Symmetric Loss Networks with Trunk Reservations-I.” *IEEE Transactions on Communications* 41(2):400-411.

Powell, W. W.. 1990. "Neither Market nor Hierarchy: Network Forms of Organization." Research in Organizational Behavior 12:295-336.

Perrow, Charles. 1992. “Small-Firm Networks.” Pp.:445-471 in *Networks and Organizations,* edited by Nitin Nohria and Robert G. Eccles. Boston: Harvard Business School Press.

Piore, Michael J. and Charles Sabel F. 1984. *The Second Industrial Divide: Possibility for Prosperity*. NY: Basic Books, Inc.

Rogers, Everett M. 1995. *Diffusion of Innovation*. 4th ed. NY: The Free Press.

Rogers, Everett and Judith Larsen. 1984. *Silicon Valley Fever.* NY: Commonwealth Publishing Co., Ltd.

Saxenian, Annalee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128.* Cambridge: Harvard University Press.

Williamson, Oliver. 1975. *Markets and Hierarchies.* NY: The Free Press.

---. 1981. "The Economics of Organization: The Transaction Cost Approach." *American Journal of Sociology* 87:548-77.