

**On the Impossibility of Equitable Continuously-Clearing Markets with
Geographically Distributed Traders**

Dhananjay K. Gode
Leonard N. Stern School of Business
New York University
427 Tisch Hall, 40 West 4th Street
New York, NY 10012
dgode@stern.nyu.edu
212-998-0021

Shyam Sunder
Yale School of Management
PO Box 208200, 135 Prospect Street
New Haven, CT 06520-8200
shyam.sunder@yale.edu
203-432-6160

First Draft January 1991
Revised February 2000

We thank the Graduate School of Industrial Administration and Information Networking
Institute at Carnegie Mellon University for financial support.

Abstract

Technology has enabled geographical expansion of markets, particularly electronic markets, by allowing traders far away from an exchange to have the same access as the traders close to an exchange. This has been possible because technology has speeded up communication speeds relative to decision-making speed. We argue that if decision making speeds also increase, for example, due to fully automated traders, then providing equal access to all traders would require moving from continuous markets to call markets. Our paper highlights an important tradeoff between the frequency with which orders are matched to determine trades and the equality of access to the market.

1. Introduction

The Internet has led to a surge in on-line trading. The ideal of a global market with instantaneous trading now appears achievable. Screen-based trading via computers connected to a fast network is viewed as providing low-cost trading with equal access to geographically dispersed traders. Since equal access encourages wider participation in the market, it increases the depth of the market and enables traders to execute their orders quickly without the presence of specialists, such as those on the New York Stock Exchange. The resulting wide participation in markets and instantaneous order execution is predicted to lead to an increase in liquidity. Accordingly, providing equal access and immediate order execution has been the goal of market designers.

The time it takes for market information to reach the traders and for their orders to reach the market is an important determinant of market access. The longer the communication delay, the less access a trader has to the market. Equal access requires that the communication delay for every trader be equal. With increases in bandwidth, it seems imminent that traders in remote areas will have the same access to markets as traders in financial centers.

We argue that absolute increases in communication speed alone do not ensure equal access to remote traders. Instead, it is the ratio of the communication speed to the speed with which traders make decisions that determines the extent to which remote traders have equal access. Even small differences in access times are economically significant if traders act quickly.

We point out that until now technology has dramatically improved the ratio of communication speeds to decision-making speeds. The communications are now much faster and better, but the decision-making speed has not increased as much. This is because although computers are used extensively for decision support, the bulk of the actual trading decisions are still made by human beings who are much slower than the underlying network.

As technology begins to automate trading decisions, the decision-making speeds rise, and the ratio of communication speed to the decision-making speed falls. Although currently only a few large trading houses use automated trading to a limited extent, it is now conceivable that the extent of automated trading, especially by software agents acting on behalf of traders, will increase dramatically.

With automated traders, small differences in access times are significant. This makes the “location” of a software agent relative to the market important. Software agents that are far away from the market are at a disadvantage relative to traders close to the market.

In this paper we examine market design issues when traders are as fast or faster than the network. We question the basic assumption underlying much of the current market designs that faster communications will allow geographically dispersed traders to have equal access to a market that is cleared continuously, i.e., a trade occurs as soon as a bid equals or exceeds an ask. We point out that when traders are fast, market designers have to trade-off providing equal access to geographically dispersed traders against clearing the markets continuously.

With geographically dispersed traders, the ideal of a continuous market with equal access is unachievable because transmission delays are sufficiently large even at high

bandwidths. Providing equal access then places a lower bound on the interval between markets can be cleared. The lower bound depends on the maximum transmission time to any of the traders in the market. These limitations suggest that in applications sensitive to equal access an alternative market organization – a call market – may exploit the advantage of electronic technology better than the current attempts that are directed at imperfect recreation of the continuous environments of physical exchange floors. Electronic implementation of continuous markets must necessarily deviate in significant ways from floor trading, and potentially induce different trader behavior and strategies, and price patterns in the market.

Section 2 discusses the evolution of communications and computing technologies. Section 3 describes the challenges faced in designing a market when traders are fully automated. Section 4 compares continuous markets and call market and shows why call market is the only way in which equal access can be assured. Section 5 summarizes and concludes the paper.

2. Evolution of Communications and Computing Technologies

Traditionally, the geographical scope of a market has been thought to depend on the speed and cost of the transporting goods and information among the market participants. When the objects being traded are financial claims that can be transported electronically, the speed and cost of transporting goods is the same as the speed and cost of transporting information. From this perspective, increases in communication speed and decreases in its cost should unambiguously increase the geographical scope of a market.

This has indeed been true in the past. Technologies such as the telegraph, telephone, and digital communications have expanded the geographical scope of markets by allowing traders who are physically far away from the exchange to trade in a timely way [Garbade and Silber, 1978].

The increasing speed and reliability of digital communications have facilitated the popularity of markets such as the NASDAQ and on-line auctions such as e-Bay. In these electronic markets, the physical proximity of traders to the exchange is no longer an issue – they are dispersed geographically. In contrast, physical presence on the exchange floor is still highly valued on traditional exchanges such as the New York Stock Exchange and the Chicago Board of Trade. The success of electronic exchanges has led to the speculation that these traditional exchanges will eventually disappear. It is argued that once computers can process speech reliably, the last remaining advantage of the traditional exchanges - the ability of traders to shout orders quickly – will disappear. The centralized exchanges can offer a deeper market with better price discovery because of consolidation of orders at a single point. Technology can consolidate orders without requiring the traders to be physically present on the exchange floor.

These technologies have increased the communication speed and the speed with which orders can be matched but have not increased the decision-making speed of traders. This is because, although computing speeds have risen and computing costs have fallen, computers have been used largely for decision-support. Computers rarely make trading decisions without final human approval except in index arbitrage and dynamic hedging in financial markets. Thus, the ratio of the communication speed to decision-making speed has steadily risen. This has enabled the expansion of geographical scope of markets.

There is now, however, a growing interest in software agents that execute strategies programmed by traders. Simple examples of such traders are the automated bidding agents (“Bid-clicks”) available at auction sites such as Amazon. While we do not suggest that proliferation of software agents is imminent, we believe that it is important to understand the design issues that could arise as we progress towards automated trading.

Computerized decision making reverses the trend of increases in communication speed relative to decision-making speed. This is depicted in figure 1. The spread of computerized trading will depend on advances in our knowledge of the cognitive psychology of trading. This knowledge is, at best, in its infancy. Our purpose is not to advocate computerized trading, but to document the market design issues that will arise if such replacement does occur. We ran into these issues when we designed Market 2001 at Carnegie-Mellon University, a trading system in which human traders were replaced by computer traders. This trading system is described next.

3. Fully Automated Traders: The Carnegie Mellon Experience

The Market 2001 project, which was started in 1989 at the Graduate School of Industrial Administration at Carnegie Mellon University, had two objectives. To understand how human traders translate their trading strategies into computer programs, and to analyze market design issues that arise when computer traders displace human traders. We designed and implemented a system that allowed MBA students to submit computer programs to trade on their behalf. Once the trading session began these programs traded with each other with no human intervention. In this paper we focus on the market design issues that arise in the design of such trading networks.

The design of these networks is challenging because the time it takes for traders to make decisions is comparable, if not shorter, than the time it takes for information to travel from the central exchange to the traders, and vice versa. The traders are very concerned if the communication delay to their machine is longer than the delays for other machines. Several students expressed concern about this issue in Market 2001. The design is also challenging because there is a clustering of network traffic, which makes protocols that rely of decentralized conflict resolution, such as the Ethernet, inefficient. In these markets, all traders may act at the same time, which necessitates a sequencing mechanism. We now present two major ways in which markets can be organized and then discuss which market organization is better suited in an environment where many fast traders need equal access to the exchange.

4. Continuous Versus Call Markets

Trading activity on the trading floors of New York, American, and most other U.S. stock exchanges as well as the trading pits of commodity exchanges in Chicago and elsewhere, may occur in spurts. However, there is no lower limit on the amount of time that must pass between two consecutive market actions. In other words, a trade can, and does take place as soon as a bid and an ask match, without waiting for any further bids and asks. These markets, in which a transaction is finalized as soon as a match occurs, are called continuously-clearing markets. For brevity, we will call them continuous markets.

Within the limits of the information processing capacity of the humans who operate them, most exchanges in U.S. operate as continuous markets.

Call markets, on the other hand, accumulate bids and offers for a pre-specified interval. When the market is called, a single price is computed based on the accumulated bids and asks to maximize the number of units traded. All possible transactions occur at this single price and the price and allocations are announced. Immediately following the call, accumulation of bids and asks for the following call resumes. This process determines the opening price and transactions in the New York Stock Exchange.

Continuous markets provide immediate execution of market orders and allow prices to adjust quickly to changing market conditions. However, these prices are subject to volatility arising from the sequence in which orders arrive at the exchange. Since call markets accumulate orders before matching them, they are less prone to such volatility. However, call markets do not have instantaneous price discovery.

An example illustrates the point. Suppose there are two buyers (B1 and B2) and two sellers (S1 and S2) in a market. B1 and B2 are willing to buy a unit for \$16 and \$10, respectively. S1 and S2 are willing to sell a unit for \$4 and \$8, respectively. For simplicity, assume that these traders do not act strategically; the buyers merely submit their values as bids and the sellers merely submit their costs as asks and the clearing price equals the midpoint of the bids and asks when bids exceed asks. In a call market, the clearing price after all these four orders have been received will be \$9. This price will not depend on the sequence in which these orders are submitted to the market. The clearing prices in the continuous market will, however, depend on the sequence in which the orders are submitted. If orders are submitted in the sequence: B1, S2, S1, and B2, then the clearing prices will be \$12 ($= (16+8)/2$) and \$7 ($=(4+10)/2$) respectively. If the price of the bid or ask submitted first is taken to be the clearing price, then the clearing prices will be \$16 and \$4 respectively. Thus, the sequence in which orders arrive at the market can cause prices in the continuous market to be volatile. This effect is likely to persist even if traders act strategically.

The above example makes it appear that the call market is superior to the continuous market. This may indeed be true if the demand and supply are static. However, if the demand and supply are changing constantly, then a call market has the disadvantage of slower price discovery. The market prices in the call market are unable to reflect the effect of changes in demand and supply until a new clearing price is set at the end of the call.

5. The Model

Let us assume the simplest of network configurations, a star network, shown in Figure 2. Each of the N trading stations on the periphery is connected to the exchange located at the center by a direct dedicated communication link. The star configuration and the assumption of direct dedicated lines abstract away from the complicating problems of traffic loads, and consequent variability and dependencies across channels. It also abstracts away from any correlation between market activity and the communication delay on any given channel because of congestion.

Associated with each link i is a deterministic communication delay d_i . In any global network, some trading stations will be closer to the hub than the others,

introducing dispersion in d_i across trading stations. The generality of the results presented here does not depend on the assumption of this dispersion. The delay is assumed to be deterministic for simplicity.

Implementation of a continuous market requires that each bid or ask be processed in the exact order in which it is received at the exchange. Any discretization of time into slices for network operation can create the possibility that more than one bid/ask will arrive at the exchange in a single time slice, making it impossible for the exchange to assign a time priority to them. Thus, the first requirement for implementing a continuous market on a network is that the time slicing be smaller than the smallest possible interval between two consecutive bids or asks. If no constraints on the time slice during which bids and asks originate is feasible or desirable, the strict implementation of a continuous exchange on computers is impossible. Something must give.

Things that can give fall into two categories. First, we give strict time priority. If two or more bids or asks arrive at the exchange within the same time slice, the exchange will process them in some order to be determined by, for example, the highest-bid/lowest-ask first rule or by randomization.

Second, a continuous market may be abandoned in favor of a call market. As described above, all bids and asks will be accumulated for a number of time slices, either specified in advance, or determined by some other rule (e.g., when a pre-specified number of bids and asks have been accumulated). Call markets usually operate on price and time priority, and discretization of time will place the same limits on the implementation of time priority in call markets as it does in the continuous markets discussed in the preceding paragraphs. However, in call markets, time is the second sort key after price, while in the continuous markets, time is the first sort key. Call markets do not have to sort by time, they may, instead, ration or randomize among all buyers/sellers at the margin.

Frequency of calls could be fixed in advance, or determined endogenously by a rule. As the frequency is increased, the market organization converges closer to the continuous form. We therefore parameterize the market organization by c , the amount of time between two consecutive calls.

Trader i who operates in a call market that runs at intervals of time c , with communication delay d_i , has a total of $(c - 2d_i)$ available for making decisions between two consecutive calls after subtracting the time needed to receive information about the results of the earlier call, and the time it takes for the orders for the next call to get to the exchange in time.

If the market design must permit a minimum of δ decision time to every trader, $(c - 2d_i) > \delta$ for all i , or $c > (\delta + 2d_{max})$. Both components of this lower bound on call durations are dependent on technology. To the extent that the mechanism for making trading decisions is computerized, or supported by computerized decision aids, δ may be reduced to a small amount of time. The minimum possible communication delay for the most far flung traders on the globe, d_{max} , is of the order of one second. There is not much room for further reduction in this delay, given the physical distance on the globe. This means that the minimum possible value of call duration of a global market, c_{min} would be about two seconds. If call markets are seen as an approximation of the continuous market, approximation closer than 2-second calls is not achievable. If traders on existing physical exchanges are used to faster response times (as they no doubt are in Chicago pits where

hand signals and eye contact are much faster) they may find this alternative too slow for their what they already have. Whether the advantages of this alternative outweigh the disadvantages requires further analysis.

We have already discussed above some of the advantages and disadvantages of call markets. Calls of longer duration create deeper markets that yield more precise price discovery because such markets are able to accumulate a larger set of orders before they determine the price. Our analysis here suggests a new and different reason for increasing the duration of calls. Traders who are farthest (communication) distance away the exchange have less time available to them for making decisions between calls than the traders located closer to the market do. Suppose we use the following expression as a measure of the "fairness" or "equity" of the market:

$$(c - 2d_{max}) / (c - 2d_{min})$$

This measure would be 1 if and only if communication delays for all traders are equal and less than 1 otherwise. However, as the value of c is increased progressively, this fairness measure also converges to 1. In other words, in a call market with large inter-call durations, relatively small differences in communication delays can become inconsequential.

However, depending on how close one wishes to get to perfect fairness in market design, this method may require a large increase in call duration. Longer call duration has its own costs. First, while the accumulation of more bids and offers over the longer durations yield a more precise discovery of price, this discovery takes place less often, and therefore, the more precise discovered price is also more stale. In fast moving markets, especially in the financial sector where the information role of price discovery is crucial, the trade off among consequences of duration is of critical importance.

An alternative to increasing the call duration when delays are not stochastic

Instead of increasing the call duration c , the fairness measure can be made to approach 1, by making d_{min} equal to d_{max} . This can be done by adding delays to the transmission of information and the receipts of orders from traders who are "close" to the exchange. This can be implemented as follows:

1. Synchronizing clocks of all trader machines. Of course, there are limits to the precision to which the clocks can be synchronized.
2. Embedding a time lock into the transaction information transmitted to all traders so that the information is readable by the traders only after time lock is deactivated at the time shown. The time lock should be set so that there is sufficient time for the information to reach all participants before the deactivation of the time lock.
3. Delay the transmission of orders from traders who are "close" to the exchange so that they are at par with traders who are "far" from the exchange.

It is difficult to make this scheme work if the delays are stochastic. In that case we can only ensure fair access on average. If the stochastic process underlying these delays is non-stationary, then this scheme of adding delays can get complicated.

Is providing equal access a desirable goal?

The exchanges of today do not provide equal access to all participants. Faster market access comes at a price. The price paid reflects how close, in a geographical sense, the participant or his agent is to exchange. Why should electronic markets be any different? Why should designers of electronic markets worry about providing equal access to participants?

In fact, making traders pay for faster or more convenient access to markets may be economically efficient. One of the important roles of markets is to facilitate trade between buyers who value the objects the most and the sellers who can produce them at the lowest cost and exclude sellers who cannot produce at the lowest cost and buyers who do not value the object as much. The total profits earned by traders are maximized when markets achieve such an allocation. In fact, one of the important performance measures of markets is their allocative efficiency as measured by the ratio of actual total profits of all traders to the maximum possible total profits of all traders.

If access to the market is priced, then traders that are likely to benefit most from trading are likely to be willing to pay most for access. This is likely to increase the chances of trade among traders who stand to benefit most from trading while minimizing the intrusion from traders who do not benefit as much from trading. This, of course, depends on the assumption that traders can choose the price of market access as frequently as the changes the benefits they are likely to derive from trading. A low-cost seller is unlikely to remain a low-cost seller forever. It is optimal from an allocative efficiency point of view that traders who are likely to benefit from trading at any given point are able to purchase better market access.

The benefit of providing equal access is that "by leveling the playing field" the exchanges may be able to attract more dispersed set of market participants. Wider market participation increases the size and depth of the market, which improve allocative efficiency as well as liquidity.

Information Structures and Agent Strategies

Continuous and call markets have very different information structures and therefore call for very different agent bargaining or trading strategies that need to be explored. We shall briefly consider two aspects of these differences.

Public Information in Continuous Markets. In continuous markets, typically, the current (i.e., the highest) bid and the current ask (i.e., the lowest ask) are made public to all participants at all times. Beyond this inside spread information, the information about the entire book (bids below the current bid and asks above the current ask) may, but rarely are, made available to the participants. When a transaction takes place, the price and volume (but rarely the identity of the transacting parties) made public information.

Public Information in Call Markets. In such markets, the time of the next call, and clearing price and volume made public after each call. The number and prices of unaccepted bids and asks, and the identities of the transacting parties can be, but is rarely made public. During the interval when the bids and asks are being accumulated before the call, the prices and quantities associated with each bid and ask (i.e., the book) can be made public. Whether it is in fact made public is a crucial design variable for the call market.

The decision making by trading agents in the two kinds of markets is different in important respects. In a continuous market, it is always possible for an agent to guarantee

a transaction by simply accepting the current bid or ask. In a call market, on the other hand, there is no way for the trading agent to ensure that he or she will be sure to transact in the next market clearing (though one can always increase the probability of this event by submitting a higher bid and a lower ask).

This advantage of continuous markets is counterbalanced by the market protection call markets provide to the trading agents against errors. In a call market, the agent is protected from even large errors in submitting bids or asks because all transactions at the time of clearing take place at a single price. The deeper the market, better is the protection call markets provide against such errors. In continuous markets, on the other hand, no such protection is available against bidding errors. Agents are punished immediately for any errors in bidding because other participants in the markets immediately snap up low asks and high bids relative to the prevailing market conditions.

Given these differences between the informational and strategic environments of continuous and call markets, very different trading strategies are called for on part of the trading agents. Detailed investigation of these differences, and their links to market rules is an interesting item for e-commerce research agenda.

Concluding Remarks

Ensuring equal access can not be left to communication engineers. The specifications of the market processes dictate a particular network design. Not all market processes can be implemented to guarantee equal access. It is easier to provide equal access in call markets.

It appears that as a result of computerization markets will have a wider geographical reach. This belief is predicated on the assumption that it will be easy to resolve the conflict between geographical dispersion and equal access for all the participants concerned. There are theoretical limits on the speed of communication. New technologies have reversed the traditional relationship between the speed of trader decision making and the communication and processing of messages; the latter now takes far less time than the former. With automated decision making (in markets populated by agents) the decision making speeds will increase dramatically, making it difficult to provide equal access to the market for geographically dispersed participants.

Technology changes both market institutions and market participants. Rapid advances in communications technology have resulted in markets with geographically dispersed participants. The New York Stock Exchange and Chicago Mercantile Exchange are typical examples of exchanges located in a specific geographical location. In such markets the traders on the "floor" of the exchange have an advantage over the traders away from the exchange because they receive the trading information before others and they can execute their orders faster than others.

Guaranteeing fair access is not enough, the system has to meet the cost, efficiency and speed criteria. A system that is excessively slow but fair would not be used by the participants. In most situations however fair access can be attained only at the expense of speed and efficiency. Network designers may have to limit the scope of the networks to alleviate the problem of fair access. Thus access problems may impose an upper limit on the globalization of electronic markets. How severe these problems

With sufficient computing power and communication channels, it is possible to increase the frequency of call markets to such an extent that they appear to be close to continuous markets. However, if the decision-making speeds of computerized agents or

computer-assisted traders also increase by similar proportions, then this approximation cannot approach continuous markets. Thus, aside from other reasons that have been advanced elsewhere for organizing securities exchanges as call markets, a new, very different theoretical reason in favor of call markets arises from this analysis. Electronic global exchanges should be call markets, not merely because they have more depth, but because there is no other practical alternative. The reason Globex (a project of Chicago Mercantile Exchange to trade futures contracts electronically) was delayed for so long is because its creators chased the ideal of continuous market ignoring the impossibility of that ideal.

References

Clemons, E.K., and Weber B. W., London's Big Bang: A Case Study of Information Technology, Competitive Impact, and Organizational Change, *Journal of Management Information Systems*, 6, 4, (1990), 41-60.

Clemons, E. K., and Weber, B. W., Alternative Securities Trading Systems: Tests and Regulatory Implications of the Adoption of Technology, *Information Systems Research*, (June 1996), 163-188.

Dewan, Sanjeev, and Haim Mendelson, Information technology and time-based competition in financial markets *Management Science*; Providence; May 1998

Domowitz, Ian, Equally open and competitive: Regulatory approval of automated trade execution in the futures markets, *The Journal of Futures Markets*; New York; Feb 1993.

Domowitz, Ian, A taxonomy of automated trade execution systems, *Journal of International Money and Finance*; Guildford; Dec 1993.

Domowitz, I., and Steil, B., Automation, Trading Costs, and the Structure of the Securities Trading Industry, Working Paper, Royal Institute of International Affairs, London, February 1998.

Economides, Nicholas; Schwartz, Robert A; Electronic call market trading, *Journal of Portfolio Management*; New York; Spring 1995; Volume: 21 Issue: 3 , pp: 10-19

Garbade, Kenneth D.; Silber, William L.; Technology, communication and the performance of financial markets: 1840-1975 *The Journal of Finance*; New York; June 1978;

Gode, D. K. and Sunder, S. (1992). Some Issues in the Electronic Modeling of Continuous Double Auctions with Computer Traders. Conference on Computational Economics, IC2 Institute, University of Texas at Austin.

Puneet Handa; Robert A Schwartz; Ashish Tiwari. The ecology of an order-driven market *Journal of Portfolio Management*; New York; Winter 1998.

Schwartz, R. A., and Weber B. W., Next-generation Securities Market Systems: An Experimental Investigation of Quote Driven and Order-Driven Trading, *Journal of Management Information Systems*, 14, 2 (fall 1997), 57-79.

Weber, Bruce, Next-Generation Trading in Futures Markets: A Comparison of Open Outcry and Order Matching Systems, *Journal of Management Information Systems*, Fall 1999, Vol. 16, No. 2, pp. 29-45.

Weber, Joseph, The Pits Are Becoming a Relic, Business Week; New York; March 6, 2000.

Vickers, Marcia, Getting Off the NYSE Floor: Specialist LaBranche is on the Web and Hot for electronic trading. Is this the Big Board's future? Business Week; New York; February 7, 2000 pp: 80-81

Figure 1: Evolution of communications and computer technology

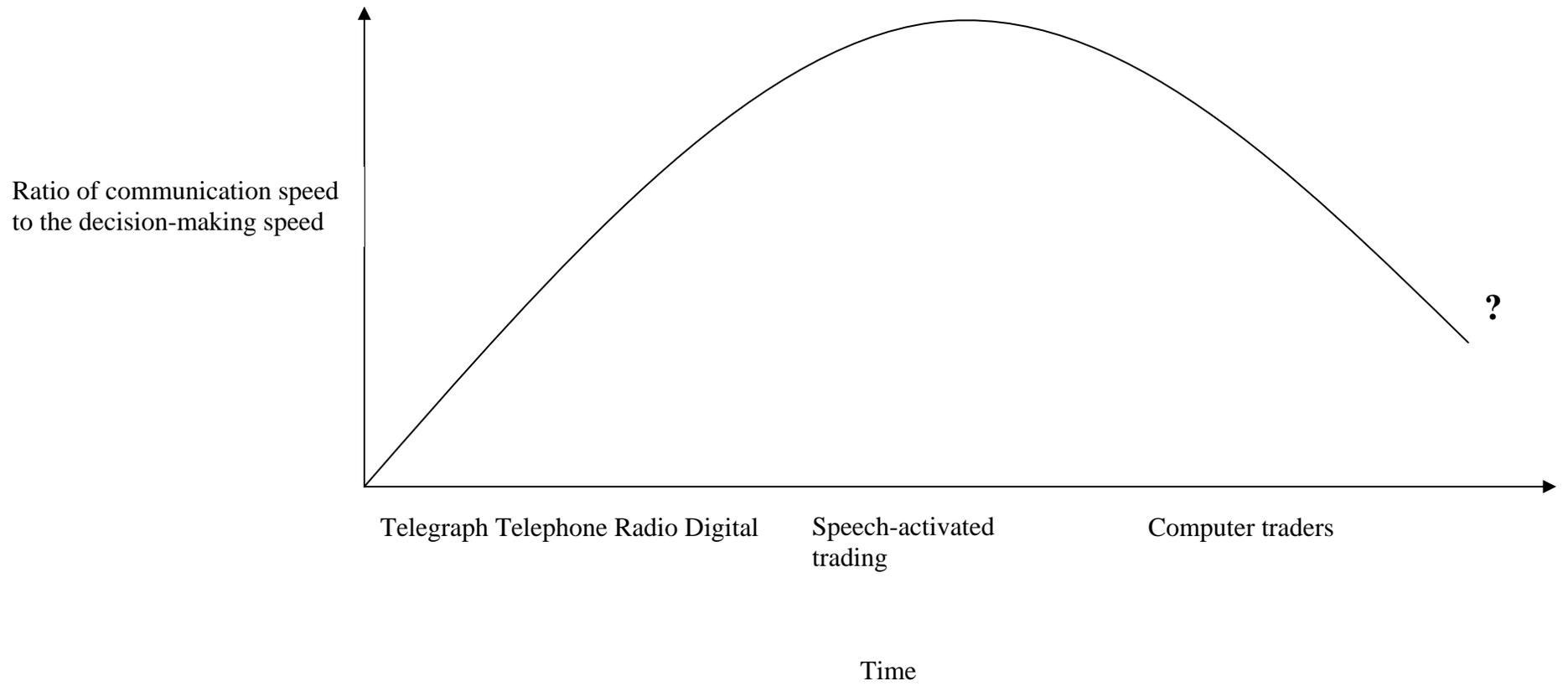


Figure 2: Star Network

