NETWORK EXTERNALITIES AND PUBLIC GOODS IN PAYMENT SYSTEMS

John A. Weinberg

Research Department, Federal Reserve Bank of Richmond
P.O. Box 27622, Richmond VA 23261, 804-697-8205

November 1996

ABSTRACT: As a network, a payment system is likely to exhibit network externalities and perhaps some public good characteristics. Such properties may be more pronounced in an electronic payment system, because of its greater reliance on communication infrastructures with high fixed and low variable costs, for instance. This paper presents the basic economics of network externalities and reviews some basic principles regarding public goods. It then asks what these phenomena imply about the role of the Federal Reserve in emerging payment systems. The general conclusion is that there is reason to be skeptical that network externalities and public goods will be significant sources of market failure in electronic payment systems. These phenomena, by themselves, give rise to no particular, essential central bank role in these markets.

In a large modern economy, there is a vast and constant movement of funds in the conduct of commerce and finance. In dollar value, the bulk of this movement is not in cash but in the form of instructions for the crediting and debiting of accounts held with public or private financial institutions. In using a non-cash payment instrument, a business or consumer relies on the process through which the payment is ultimately made final. That is, parties to a transaction rely on the network that connects the point of the transaction (in space and time) to the parties' accounts with financial institutions. The value of a payment instrument, then, depends on the extent of the network with which it connects, as well as on the cost of moving instructions over that network.

As a network for sending and receiving instructions, the payment system bears a resemblance to transportation and, especially, communication systems. Accordingly, many of the issues and questions that arise in discussions of markets for payment services have parallels in discussions of these other markets. Does private market interaction result in an efficient network composition? How do multiple participants in network service markets coordinate their actions to ensure a level of compatibility that allows all network users to communicate effectively with one another?

One might actually think of a payment system as consisting of two networks. One is the network that communicates payment instructions. The other is the network of debt obligations created in the process of making transactions with deposit money. This second network aspect to a payment system raises issues that may be distinct from those that apply to other networks. If an extensive payment network is created to take advantage of the resource cost benefits of network size, does the resulting network of debt obligations produce risks that are harder to monitor and contain than would be the case if payments were executed by some other means? Is there some sense in which such a system tends to create too much risk?

There is a long history of public sector intervention in network industries. Indeed, Peltzman (1989) identifies network characteristics as being the single most important common thread in government owned enterprises in developed economies. For instance, in many advanced economies, transportation and communication industries are served by government-owned enterprises. Even where these industries have been largely served by private enterprise, those private firms have traditionally been subject to substantial regulation. As with other network services, public sector intervention has traditionally been common in payment services industries. This intervention typically takes the form of central bank involvement in the settlement (and in some cases clearing) of payments.

The last two decades have seen dramatic changes in the role of the public sector in many network industries in the developed economies. Government sponsored enterprises have been privatized. Markets formerly reserved for protected monopolists have been open to competition, and other efforts have been made to deregulate the behavior of private market participants. This change in the public sector role has, in part, been driven by changing technology. Does changing technology in payment systems indicate a need for a similar rethinking of the public sector role in these markets? Discussions of this question will no doubt focus some attention on the network characteristics of payment services. Section 1 of this paper examines the network characteristics of payment systems and electronic payment systems in particular. Section 2 then provides a brief primer on the analysis of economic behavior in network markets. Two approaches are outlined; one focuses on market performance under particular assumptions about institutions (or rules) through which market participants interact, while the other assumes that institutions and rules of behavior will adapt to the conditions prevailing in the market. While the first approach leads to results of market failure in the presence of network effects, the second approach gives one reason to be skeptical about such results.

In addition to network considerations, discussions of a central bank's or other public sector role in the payment system often focuses on whether there are public good aspects to these markets. While potential public good aspects are related to the network characteristics of these markets, some distinct issues arise. Section 3 explores the extent to which certain components of an electronic payment system are public goods and argues that the analysis of the behavior of private markets with certain public good characteristics is analogous to the second approach to the analysis of network markets. Section 4 then discusses the implications of economic analysis of network effects and public goods for the role of the Federal Reserve in such systems.

1. Networks and Network Externalities

For the purposes of the present discussion, we can define a network as a set of economic agents (business firms, households, financial institutions) that are able to interact with one another. In particular, we might wish to specify that a network allows agents to interact at a lower cost than otherwise. I can communicate with my sister in California by travelling to California and speaking with her. If we are both connected to a postal network (system) that we see as reliably delivering messages, I can send the same message more cheaply than carrying it myself. Our communication is even less costly if we are both connected to a telecommunications network.

Clearly, the value to me of being connected to a communication network depends on whether people with whom I wish to communicate are also connected. Hence, one important characteristic of a network is how extensive it is; how big is the set of connected points? Other important characteristics of a network have to do with its physical design. What route does a message follow in going from sender to receiver? How long does it take? What happens if there are difficulties with the network's capacity for carrying messages?

One notable characteristic of network services is that there is a discrete component to the demand behavior of the individual. While an individual household's total demand for communication services may be a continuous variable, the household must also choose a discrete set of networks over which to send (and receive) communication. Each network to which it chooses to be connected may present the household with certain fixed costs that are independent of the intensity with which it uses that network.

The services of a network are inherently intermediary services. They are services that allow two individuals to interact. The ultimate end users of a network service rely on the inputs of one or more intermediary agents to carry out the necessary communication. Related to this intermediary nature of network services is the fact that a network service is a composite good made up of two or more components. In a communication system, the two components are the sending and receiving of messages; one of each is necessary to constitute one "unit" of communication. In a market with multiple sellers, market performance will depend on the compatibility of the services of competing sellers. Can a message sent by seller A's customer be received by seller B's customer? Here, end users are relying on the services of multiple intermediaries, and the value of a service composed of pieces purchased from multiple intermediaries will depend on the relationship (as well as the buyers' beliefs about the relationship) among the sellers involved.

The value to an individual user of a network service depends on the number of other users to whom the individual connects through the network. While this "network externality" is an important characteristic of network industries, it can appears in other industries as well. In general, a network externality exists whenever an individual's willingness-to-pay for a product is increasing in the aggregate amount of the product purchased. Some authors have suggested the presence of such an externality in the market for personal computer operating systems; from a given set of alternatives, buyers prefer to have the system that is more widely chosen. In this case, the externality comes from the indirect affect that a system's popularity has on the likely availability of application software. The same sort of indirect network externality is often said to exist in any market in which the ultimate service to the user is compsed of the complimentary services of hardware and software.

In the case of a communication network, the externality effect is more direct and can, perhaps be a bit more specifically defined. It is probably useful to think of an economic agent as placing a fixed value on sending a message to any particular other agent; this value is fixed in the sense of being independent of the method of sending the message, as long as the message is sent reliably and within a desired time-frame. The value of a network is not so much in the volume of communication on the network (the aggregate amount of communication services sold) as in the number of other agents with whom one can communicate. That is network externalities often have their most immediate effect on a "subscription decision" (deciding whether to connect to a network) rather than on the choice of how much of a service to buy, given subscription.[1]

There is an important difference between network externalities and some other forms

---

[1] In this respect, a network externality is similar to what some refer to as a "thich markets" externality that makes an asset more liquid the greater the number of potential traders participating in its market.

of externality. Perhaps the most common textbook externality is pollution; the economic activity of some individuals may produce pollutants that affect a much broader set of people. While individuals make choices about participation in the activity that generates pollution, they may have little choice as to whether to be affected by pollution. In the extreme case of, for instance the greenhouse effect, it may be impossible to avoid incurring the costs. Hence, pollution is an external cost that has effects beyond the set of people engaged in the polluting activity. Network externalities, on the other hand, are more self contained. An individual's decision to subscribe to a network creates external benefits for other subscribers by increasing the size of the network. Notice, however, that in order to enjoy the effects of the externality, one must join the network. Similarly, the more indirect "network" benefit that comes from using a popular piece of hardware for which a lot of software is available is only enjoyed by those who purchase the hardware. Another way of putting this difference is to say that the benefits of a network externality interact directly with the decisions that generate the externality. This difference can have important consequences for the analysis of economic behavior in the presence of network externalities.

Since a payment system is an explicit network, we should expect network externalities to be present. Certainly, the value to a business or consumer of adopting a new method of payment depends on the extent to which that payment instrument will be accepted by sellers of goods and services. Similarly, a financial intermediary's willingness to pay for a new technology for clearing payments will depend on how many other financial intermediaries are "connected" through the new technology. There may also be a network effect in the settlement function of a payment system. If payment between two individuals or institutions is settled in account balances with a third institutions. Payers and payees may prefer dealing with settlement institutions that deal with large numbers of endpoints.

Another possible network effect in some electronic payment systems arises from the need for security. A party to a transaction needs to be able to protect payment information from unauthorized eyes while, at the same time, being able to establish the authenticity of a transmission to the intended counterparty. The digital signature and certification authority procedures that have been proposed for dealing with these problems create a role for an intermediary between the parties to a transaction. This intermediary serves to verifies the identity of a counterparty. The value of such an intermediary service may well depend on the number of counterparties with whom an intermediary (or network of intermediaries?) does business. This effect may arise from the need to trust the intermediary; trust, in turn, may require that end users monitor intermediaries' reputations. The enhanced value of an intermediary that does business with a large number of end users, then, would arise from the end users' ability to economize on such monitoring costs.

2. Economic Analysis of Network Markets

The presence of network externalities in a market will certainly affect the structure and performance of the market. In the economics literature, the analysis of these effects has tended to follow one of two approaches. One approach takes certain aspects of the industry's structure and the behavior of firms as given and seeks predictions about market equilibrium prices and quantities. The equilibrium allocations of production and consumption of the product can then be evaluated against an efficiency standard. The second approach begins by characterizing efficient allocations. This approach then asks if there is a pattern of prices that makes the efficient allocation sustainable in the face of competing pricing arrangements.

Both approaches are based in the game theoretic analysis of economic behavior. The first approach falls into the category of noncooperative game theory. Noncooperative game theory studies the strategic behavior of decision makers under a fixed set of rules. Broadly defined, the rules may specify the number of "players" and the set of actions from which they are allowed to choose. For instance, describing a duopoly (two firm) market as a game, one might say that the game has two players who choose fixed prices for their products and stand ready to sell as much as is demanded at the specified prices. Note that this example places a strict limitation on the types of strategies firms are allowed to choose; more general pricing strategies that, for instance, make the price paid by any buyer depend on the amount purchased are not allowed. The noncooperative approach analyzes market behavior under a fixed set of rules governing allowable interactions among agents. The predictive tool of the noncooperative approach is Nash Equilibrium. In such an equilibrium, each player (agent) chooses his own strategy to maximize his own well being, taking as given the strategies chosen by all other agents.[2]

An alternative to the noncooperative approach seeks to analyze behavior in economic settings without placing a priori restrictions on the ways in which agents can interact. This "cooperative" approach makes the central assumption that economic agents can form coalitions to act together. With this assumption, one can seek resource allocations that leave no opportunity for any coalition (subset of agents) to form and make its members better off by doing something different from that prescribed in the status quo allocation. Such a sustainable allocation is said to be in the "core" of the economic setting under study. Note that a core allocation is an efficient allocation; if an allocation is inefficient then a coalition consisting of all interested parties (all of the agents in the economy) can make its members better-off by moving to an efficient allocation. The core is the key predictive tool of the cooperative approach.

The noncooperative approach

The noncooperative approach to the analysis of network markets, as surveyed recently by Economides (1996), examines certain fairly standard questions about market performance in the face of network externalities. For example, is competitive equilibrium efficient? Does the price and output chosen by a nondiscriminating monopolist yield greater or lesser social benefit (surplus) than perfect competition? In addition, this approach examines the strategic behavior of competing sellers of network services. Will such sellers choose to make their components compatible with those of their competitors? At what price will a vertically integrated seller choose to sell key intermediate inputs to a competitor in a final goods market? How do network externalities affect the willingness of buyers and sellers to adopt technological innovations? How do the answers to these questions depend on the number of

---

[2] In what follows, I include the case of perfect competition within my discussion of the noncooperative approach. This is not quite right. The analysis of competitive equilibrium is usually treated separately from noncooperative game theory. On the other hand, competitive analysis studies market performance under a fixed set of behavioral rules; agents take price parametrically and choose their quantities supplied and demanded.

sellers and other characteristics of market structure?

The approach to the basic questions, such as the evaluation of competitive equilibrium, can be demonstrated by means of a simple example. Consider a market populated by a large number of potential buyers. Each buyer has a demand for at most one unit of a particular good. An individual's demand is characterized by a "reservation price," a maximum price that he or she is willing to pay for a unit of the good. A buyer's reservation price is a function of his own tastes and of a network effect. For instance, one might represent a reservation price by $\theta v(x)$. Here, θ represents the individual's preference for the good; individuals with higher θ have a relatively greater desire for the good and are willing to pay a higher price. The function $v(x)$ captures the network effect, where x represents the number of individuals who purchase the good. A (positive) network effect is captured by assuming that $v(x)$ rises as x rises. Hence, the more people that buy the good (the larger the network), the greater is any individual's willingness to pay. While this example abstracts from the multiple component nature of network services, it does capture the sense in which these services exhibit an externality.

With individual demand described as above, the aggregate quantity demanded at any market price is simply the number of buyers with reservation values greater than the market price. However, since individuals condition their demand decisions on an assumption about network size(x), market demand will also depend on such assumptions. Hence, we might write a market demand function as $D(p,x)$, where p is the price of the good. As usual, this demand function is decreasing in the market price, p. Due to the externality, however, demand is increasing in participants' assumption about network size. Here, however, is a possible inconsistency. The actual network size that results from this market is simply the quantity demanded at the market price, $D(p,x)$. What happens if this quantity is different from the assumed network size (x) on which buyers base their decisions? This potential problem is resolved by requiring that the market outcome be consistent with buyers' assumptions. It is possible to define a set of price-quantity combinations (p,x) such that $D(p,x)=x$, so that consistency is maintained. The equilibrium in this market will come from this consistent set. In figure 1, this set is represented by the curve $P(x)$.

Given the demand structure described above, the equilibrium price and quantity will be determined by the costs and behavior of sellers. Suppose, for instance, that the good is produced under constant returns to scale, so that a seller's average and marginal cost are both equal to a constant number, c, independent of quantity produced. Then, the competitive equilibrium price is equal to c, and the equilibrium quantity that level of x such that D(c,x)=x (price is equal to marginal cost and total demand is consistent with buyers' assumptions).[3]

The simplicity of the characterization of competitive equilibrium masks a potential complication. Competitive equilibrium assumes that there are multiple sellers. How is the market divided among the sellers? The answer to this question depends on whether network externalities are specific to the buyers of an individual seller's product or are instead enjoyed by buyers across all sellers. That is, the answer depends on the compatibility of the products produced by competing sellers. If each seller's product is fully compatible with all others' products (so that the network externality is a function of aggregate quantity), then the distribution of sales across sellers is indeterminate and of no great concern. If instead sellers' products are not compatible with one another (so that network externalities are enjoyed only among those buyers who buy from the same source), distribution across sellers is more consequential. The preferred allocation from an efficiency point of view is for one firm to service the entire market. This allocation will be a competitive equilibrium in the traditional sense of all participants treating price as parametric and setting their own quantities demanded and supplied. There may, however, be other such equilibria with more than one active firm. These other equilibria generate fewer network benefits than does the equilibrium with a single active seller, although in all competitive equilibria, price is equal to marginal cost.

One response to this multiplicity might be to imagine the behavior of the market if the "price taking" assumption of perfect competition is relaxed. One might instead imagine that a fixed number of potential sellers plays a specific price-setting game. The Nash equilibrium outcome of this game will be sensitive to various assumptions about the game's structure, including: the number of players (potential sellers); the strategies available to the players, or

---

[3] Although figure 1 is not drawn this way, it is possible for there to be multiple network sizes satisfying D(c,x)=x (for instance, if c is greater than the vertical intercept of the curve P(x)).

what types of price (and perhaps quantity) offers they are allowed to make; and the order in which players make their moves (choose their strategies). For instance, one possibility is to examine a game that is analogous to Bertrand's price-setting competition between homogeneous-good duopolists. Here, one imagines that two sellers simultaneously announce prices and stand ready to sell to all comers at the announced price. Unfortunately, this game is subject to the same multiplicity as is price-taking competition. If there is a price-taking equilibrium with two active sellers, then that same allocation will result from the two-seller price-setting game. Of course, the outcome with a single active seller will also be an equilibrium of this game.

The failure of this simple pricing game to resolve the multiplicity of equilibrium outcomes arises, in part, from the severe limitations that this game's structure places on the types of strategies from which sellers choose. This fact highlights a drawback to the noncooperative approach in studying economic behavior in network markets; this approach requires that one make a priori assumptions about the scope of market participants' strategic thinking and behavior. This theme will prove important below, in the discussion of the alternative, "cooperative" approach.

Although the multiple competitive equilibria discussed above can be ranked according to efficiency, even the best such equilibrium, that with a single seller, is inefficient. A competitive equilibrium quantity is determined by setting the marginal private value of the good to its marginal cost. The marginal private value is simply the "last" buyer's reservation value. Because of the network effect, however, the marginal social value of the good is greater than the marginal private value. Marginal social value is the last buyer's reservation value plus the increase in the value received by other buyers due to the last buyer's addition to the network. Since an efficient allocation would equate marginal social value to marginal cost, the competitive equilibrium network size is inefficiently small.

In some economic models in which externalities cause competitive equilibrium to yield an inefficiently low output, one might guess that a monopolist may be able to internalize the externality, thereby producing a larger output. Weighing against this possibility, however, is the natural tendency of a monopolist to reduce output below the competitive level. With the network externality demand structure and constant returns cost structure described above, a nondiscriminating monopolist will set a price greater than marginal cost. Hence, the quantity sold will be less than would be produced in competitive equilibrium. One partial exception to this general rule is in the case of multiple, incompatible sellers. The monopoly solution may be more efficient than a competitive equilibrium with more than one active seller, but it is less efficient than the competitive equilibrium with a single active seller.

Note that the comparison made above is between competition and non-discriminating monopoly. Hence, this comparison, like the discussion of price-setting games, requires an a priori restriction on the set of pricing strategies available to sellers. Depending on the information available to a monopolist seller about individual preferences, it may be possible to find a discriminating pricing policy that allows the monopolist to capture the benefits from network expansion and produce more of the good than is produced in competitive equilibrium. There are other sets of conditions under which a nondiscriminating monopolist can capture the network benefits and produce an efficient level of output. For instance, if the above conditions are modified to make all buyers identical (all have the same θ) and to make production subject to decreasing returns to scale (increasing marginal cost) then the monopolist's profit-maximizing output also maximizes social surplus from the market. This case is represented by figure 2.[4]

The forgoing discussion of equilibrium in a market subject to a network externality assumed a fixed set of "networks" available to consumers. In the basic case, that set consisted of a single network. In the case of multiple sellers, the compatibility or incompatibility of sellers' products was taken as given. The case of perfectly compatible sellers amounts to multiple sellers of access to a single network. The case of perfect incompatibility amounts to each seller offering a distinct network. In this case, it was argued above that equilibria are possible in which multiple, incompatible networks are active. There is, however, something fragile about such equilibria. In the absence of a dynamic theory that explains how individual choices drive network sizes over time, static equilibrium theory simply looks for a set of conditions (prices and allocations) such that, markets clear (supply equals demand), and no

---

[4] This example is discussed by Liebowitz and Margolis (1994).

individual participant can do any better. Consider an equilibrium with two active networks. If the only possible difference between two sellers' products is in the size of their networks, then both networks must have the same size in order for buyers to be indifferent between them. One might imagine that a slight perturbation away from this balanced condition would drive the market away from the two seller equilibrium. As soon as one network becomes bigger than the other, it becomes preferred by all buyers. Katz and Shapiro refer to this characteristic as the "tippiness" of network markets.

A dynamic theory of how buyers choose among incompatible networks could refine the set of competitive equilibria predicted. Another direction of refining and extending the noncooperative analysis of network markets is through modelling sellers' choices with regard to the compatibility of their products. Compatibility choices amount to choices by multiple producers about whether to adhere to a single set of standards in their products' design. Certainly, if there is an existing standard in a network market, a new seller will most likely want to produce a product compatible with the standard. A new producer of fax machines, for instance, has little incentive to produce a machine that cannot communicate with machines already in use. What is less clear is how original standards get set when there are multiple, incompatible alternatives. If individual agents have property rights in potential design specifications, then each "owner" of a design will naturally want his to become the market standard. How does the competition between competing standards resolve itself?

An important note is that sellers will not always choose to adhere to a single standard. If sellers' products are heterogeneous and buyers' preferences are sufficiently diverse across the products, then sellers may choose to produce incompatible products even if compatibility would yield additional network benefits. The tradeoff between compatibility and diversity is discussed by Besen and Farrell (1994) and Economides (1996).

The noncooperative approach represents compatibility choices in terms of a game played by a set of sellers, in which each seller decides which standard to adopt out of a fixed set of alternatives. The equilibrium outcome of such a game depends on assumptions about: the benefits of compatibility; the benefits of variety; the way in which price competition among sellers is affected by the degree of compatibility; whether any firm has a strategic advantage due, for instance to a superior technology or a large base of buyers; and other characteristics of the market. As one might expect, noncooperative equilibria of these games are not always efficient. For instance, in a game where there is a strong benefit to compatibility, but two possible standards, there will typically be two equilibria; in each equilibrium, all firms choose to adhere to the same standard. This approach makes no prediction about which standard is chosen, even if one standard is technologically superior to the other. This feature of equilibrium is referred to in the game theory literature as a "coordination failure." Coordination failures imply that a superior new technique may be adopted too slowly. By the same token, however, coordination failures can also imply that new techniques might be adopted to quickly. Coordination failures rely on the simple logic that, in markets with strong network effects, if one individual assumes everyone else is taking a particular action, it is optimal for that individual to take the same action.

As with the determination of price and network size, the noncooperative approach to compatibility choices imposes strict a priori rules on the way in which decision makers can interact. If a given set of rules produces a "coordination failure" result, there may be another set of rules under which only the most efficient standard is chosen. The limitation of the noncooperative approach is that it treats rules as exogenous to the economic decision makers in situations where everyone could benefit by agreeing to a different set of rules.

The cooperative approach

If we imagine market participants communicating with each other in order to set the "rules" which will govern their production and consumption decisions, then we are stepping outside of the noncooperative framework. The alternative, cooperative approach, does not take a stand on what rules will actually be chosen. Rather, it assumes that economic decision makers will arrive at a set of rules that delivers an outcome that is in the "core." An outcome in the core leaves no incentive for any subset of agents to jointly deviate from the proposed outcome.

Like the noncooperative approach, the cooperative approach begins by specifying the economic environment: Who are the relevant agents? What goods are produced and consumed. What allocations are feasible given the technologies available to agents.

Additional considerations include the allocations that coalitions (subsets of agents) could obtain if they "broke off" from the rest of the market. This approach allows one to predict both the structure (e.g., size) of a network as well as the way in which costs of network services are shared among buyers. By not placing a priori restrictions on the form of pricing, the approach allows for cost sharing arrangements that make prices vary with differences in the value that individuals place on network services. Sharkey (1992) and Henriet and Moulin (1996) use this approach for studying pricing arrangements for networks, while Telser(1994) argues broadly for the use of the approach to study economic organization, especially when individuals face discrete choices (such as the choice of whether to join a network).

A simple example can demonstrate the difference between the two approaches. Consider a group of economic agents who have a desire to communicate with one another. To be precise, suppose that there are three individuals, named agents one, two and three. Each individual derives utility from sending messages to some subset of the other individuals. The pattern of desired messages is represented in figure 3. Agent one desires only to send a message to agent three. Agent two derives utility from sending messages to agents one and three, while agent three would like to send to agents one and two. There are two ways to communicate. A message can be sent by direct, bilateral communication at a cost (to the sender) of c_0. Alternatively, agents can buy access to a network. If an agent is connected to a network, then messages to any other agent that is connected to the same network are costless. The cost of connecting an individual to a network is $c_s > c_0$. If we assume that the (utility) value of sending a message is at least c_0, then each agent will send all desired messages even in the absence of a network connection. In this case, the (gross) value of network connection to an agent is the saving of bilateral communication costs. Hence, this environment has a network externality property very similar to that in the previous subsection. Here, however, the value of network connection depends not directly on the size of the network, but on the names of the other agents that have connected.

Under the additional assumption that $3c_s < 5c_0$ (costs of universal connection are less than total communication costs in the absence of a network), an efficient network is one that includes all three agents. If there are multiple potential sellers of network services, all with the same cost (c_s per subscriber), will competition for subscribers lead to an efficient network?

This may depend on the exact nature of the competition. Suppose, for instance, that sellers play the following pricing game. Each seller announces a subscription price (connection fee) and stands ready to connect all agents who want connection at that price. This game is analogous to the price competition described in the previous subsection. The equilibrium of this game has none of the agents buying access to the network (that ia, a network size of zero). Briefly, the connection fee must be at least c_s, since sellers have no incentive to sell at a loss. Agent one will not pay c_s to join the network, since joining only saves him bilateral costs of c_0 (assuming the other agents are in the network). Without agent one in the network, connection is not worth c_s to the other agents, so the equilibrium network is empty. Hence, this example illustrates the above argument that, by fixing the rules of the game, the noncooperative approach often produces the "market failure" result of inefficiently small networks.

Instead of examining equilibrium of competition under a fixed set of rules, we can examine the core of this economic environment. Since the core must be efficient, the core network is that consisting of all three agents. The remaining task is to determine how the costs of connecting the three agents ($3c_s$) should be shared among the three. As we have already seen, charging each agent c_s does not work; agent one can be charged no more than c_0. The remaining cost ($3c_s$ - c_0) must be covered by charges to agents two and three, with the restriction that neither can be charged more than $2c_0$, the cost to each of sending all messages bilaterally. Hence, one possible cost allocation is for agent one to pay c_0, agent two to pay $2c_0$, and agent three to pay $3(c_s - c_0)$. These prices give no agent an incentive to leave the network and send all communications bilaterally. The prices also leave no inventive for any pair of agents to form a separate, two-agent network and communicate with the third bilaterally. Hence, these prices support the efficient network as a core allocation.

Does there exist some set of rules under which the noncooperative equilibrium network is the efficient network? Suppose, as before, that a large number of (incompatible) network service providers competes for subscribers. Instead of requiring that competition be only in the form of nondiscriminating access prices, suppose that the sellers can make any type of price offer they wish. That is, price offers can be in the form of a distinct price for each buyer. Equilibrium prices for this form of competitive bidding correspond to core cost allocations, like the one specified above.[5]

The most notable feature of the pricing arrangement proposed above is that, in order to support an efficient network, it requires the subsidization of agent one's connection by the other two agents; agent one's connection fee must be less than the resource cost of connecting him. The other agents are willing to cover the remainder of the cost, because agent one's (social) value to the network exceeds the (private) value he places on network access.

Note that an alternative "solution" to the network externality problem in this example might involve a more centralized intervention. In particular, under a legal prohibition of or a tax on direct, bilateral communication, the efficient network will result from marginal cost pricing of access (each agent pays c_s). If there are no costs associated with implementing and enforcing such an intervention, then the difference between a core allocation and one resulting from marginal cost pricing is distributional. In the example, imposing a network priced at marginal cost works to the disadvantage of agent one, and to the advantage of agent two (agent three may be made better or worse off, depending on the precise values of c_s and c_0). The fact that there may exist a government intervention that results in the efficient network does not imply that such intervention is essential. On the other hand, intervention may be essential to achieve a particular distribution of the benefits from the efficient network, if a particular distribution is desired for some reason.

The cooperative approach can also be used to discuss issues like compatibility and standard-setting among sellers of network products. Recall the problem of multiple sellers choosing between two possible standards. With strong network effects, noncooperative equilibrium has all sellers choosing a single standard but makes no prediction about which standard, even if one is inherently superior. The cooperative approach assumes that this group of economic agents will jointly realize the benefits of coordinating on the superior standard. Since a core allocation must be efficient, this approach largely ignore the possibility of coordination on an inferior standard. While the exact method by which agents

---

[5] There are many equilibria, corresponding to many core allocations. In all of them, agent one's access price is no greater than c_0, agents two and three face prices no greater than $2c_0$, and all agents connect to the network.

communicate for the purpose of coordinating their actions is not explicitly modeled, one might imagine something like the private, standard setting bodies that exist in many industries.

A potential complication to the cooperative approach arises from the fact that in some network settings, there may be no cost allocation the "supports" the efficient network. In such a case, it is said that the core is empty. If the core is being used as the tool for predict economic outcomes, then cases of an empty core leave one with a predictive tool that doesn't always yield a definitive prediction. One approach to this problem is to adopt a stability concept that is similar to but weaker than the core. Such a concept will yield the same predictions as the core when the core is not empty and will yield predictions in cases where the core does not. Whereas an allocation in the core must be immune to deviations by any coalition of agents (must not give any coalition an incentive to break away), the weaker notions place restrictions on the types of coalitions that might reasonably be able to break away.[6]

3. Public Goods

Most goods and services people consume are purchased bilaterally and have the characteristic that if one person consumes a unit of the good, no one else can consume that same unit. Such goods are called "private goods." Some goods have distinctly different characteristics. In particular, "public goods" have two essential characteristics. First, one person's enjoyment of the services of a public good does not diminish the amount available for others. For instance, if one ship makes use of the light from a lighthouse, no other ship in the area is thereby prevented from doing the same. In this regard, public goods are said to be "nonrival" goods. Second, once a public good has been produced, it is impossible to prevent any individual from deriving benefits from it. The national security that comes from having a strong military is shared by all persons living within a nation's borders. This characteristic of public goods is referred to as "nonexcludability." A pure public good is one that is nonrival

[6] Lacker and Weinberg (1992) take this route in studying the behavior of a credit market with private information.

and nonexcludable. There are some goods that have one of the properties but not the other. Most important are some goods that are nonrival but excludable. One person's reception of a scrambled television signal, for instance has no effect on others' ability to receive the same signal (nonrival), but access to the signal is restricted to those who have the equipment necessary to unscramble the signal. Goods that are non-rival but excludable are sometimes referred to as "club goods."

The difference between pure public goods and club goods is closely related to the difference suggested above between network externalities and more general externalities. Indeed, one might think of the external benefit associated with network size as a club good. It is nonrival, in the sense that one person's enjoying the benefits of a large network does not diminish the amount of that benefit available to others. These benefits are, however, excludable; only those who choose or are allowed to join the network receive benefits.

Public goods characteristics might be present in a number of dimensions of a payments system. Can any of these dimensions, however, be considered pure public goods? There is a strong reason to believe the answer is no. Generally, the benefits to using a payment system accrue to people who use the system. Hence, people can be excluded from these benefits to the extent that they can be excluded from access to the payment system. This does not mean that excluding people from access is desirable. Rather, the importance of excludability is that it implies that access can be priced. It is important to distinguish between instances in which access to a good or service is not priced because such pricing is infeasible (pure public goods) and instances in which access is not priced because of a public policy decision. Examples of the latter include much of the interstate highway system and the Federal Reserve's sorting and replacing of worn currency.

Unlike nonexcludability, the nonrival property of public goods may be common in payment system services. In the electronic communication associated with some payment services, for instance, the marginal cost of transmitting a message is close enough to zero so that one can think of the associated communication infrastructure as nonrival. Given a greater relative importance of fixed costs, this property is likely to be more important for electronic payment systems than for those based on paper means of transmitting information.

Other characteristics of a payment system will have the nonrival property to the extent that they apply to the system as a whole and do not differentiate among uses or users of the system. For instance, actions taken to enhance the security of a system typically benefit all users in a nondiscriminating way. Hence, the reliability of an institution filling a certification authority role or serving as a repository of encryption keys is a nonrival good. It is, however, excludable, since the institution can sell the use of its services as a "trusted third party."

One last dimension of payment systems is worth mentioning with respect to public goods. As discussed above, in an industry with network effects, the establishment of commonly accepted standards is an important means for coordinating the actions of market participants. Such coordination enables diverse participants to take advantage of network effects. A technical standard, for instance, may be a design specification for hardware or software. As such, standards are similar to the design specification of a technological innovation. Romer (1990) has emphasized the nonrival nature of the new knowledge created by such innovation. Further, although intellectual property rights (e.g., patent law) attempt to make such knowledge excludable, excludability is often imperfect. There is, however, an important difference between designs that represent the creation of new knowledge and those whose primary use is in facilitating coordination. While the potential value to an individual of new knowledge may be partly or entirely independent of others' use of the knowledge, the value to an individual of a standard is primarily or entirely driven by the fact that the standard makes that individual's actions compatible with the actions of others. In an industry with strong network effects, a design that is inherently superior in technological terms has little or no value if it is not adopted by a critical mass of users.

The source of value in a standard has implications for its characteristics as a public good. Like a design representing new knowledge, the benefits created by a standard are nonrival. Excludability, however, seems likely to be more certain in the case of a standard. Adherence to a standard gives one the technical capability to interact with a network of market participants. To the extent that access to that network can be priced, the benefits of the network's standards are excludable.

In summary, while many dimensions of a payment system may have the nonrival property of public goods, it is doubtful that the nonexcludability of pure public goods is an important characteristic. Quite simply, the benefits of a payment system accrue to those who make particular choices with regard to payment instruments, and the pricing of these instruments makes the pricing of access to any related benefits possible. Excludability is important for thinking about the role of the public sector in providing a nonrival good. While it is easy to imagine difficulties in the private provision of a pure public good, club goods (nonrival but excludable) are more conducive to private arrangements. Excludability is the key here, because it allows a seller of a club good to determine the intended set of buyers and price access to the good accordingly.

The analysis of the allocation of resources to club goods is analogous to the "cooperative" approach to network externalities described in section 2 above. A core arrangement determines what sets of agents have access to what club goods and how the cost of providing these goods is allocated among their recipients. Such an allocation has the property that no subset of agents has an incentive to form their own "club," sharing among themselves the cost of the relevant club goods. As above, such an allocation is efficient, and one can often be replicated as a market equilibrium given an appropriate set of assumptions about the form of price competition.[7]

4. The Role of the Federal Reserve in Electronic Payment Systems

Do network or public goods features of payment system markets create an essential role for the Federal Reserve or other public entity? It is true that certain common forms of market behavior may fail to produce efficient outcomes in markets with these features; the noncooperative approach to network externalities makes this point. On the other hand, it is not clear why market institutions and market behavior cannot adapt to a form that yields efficient outcomes. This is the message of the cooperative approach. Because network externalities are contained within the group of individuals that purchase network services, there is reason to believe that appropriate market institutions can adequately allocate these services. Similarly with "impure" public goods that are nonrival but excludable. Hence, the cooperative approach suggests that the network and nonrival characteristics of payment

[7] The theory of club goods is surveyed in Cornes and Sandler (1986).

systems do not, by themselves imply a necessary role for the Federal Reserve.

One possible qualification to the conclusion just stated arises if there are impediments to the adaptation of market institutions and behavior. If there are legal restrictions on the forms of contracts into which private parties may enter, then pricing rules that support an efficient network might not be possible. Suppose, for instance, that there are restrictions on a seller's ability to offer general, discriminating pricing schedules. As discussed in section 2 above, an inability to price discriminate can result in networks being inefficiently small (failing to take full advantage of the network externality). If such a legal restriction could not be removed, then it could, in principle, be possible for public intervention to improve the performance of the industry.

Another possible impediment to market adaptation may come in the form of frictions which limit the speed with which market participants might take advantage of profit opportunities. If the benefits of a large network lead to a concentrated industry and if there are legal or technological barriers to entry, then the market may be subject to the exploitation of market power by incumbent sellers. The monitoring of such market power is not, of course, an inherent central bank activity although it is a role that might be usefully played by some public authority.

While network effects do not, by themselves, imply a particular central bank role, they may interact with other roles played by a central bank. Hence, if a large population of institutions providing payment services (i.e., depository institutions) hold balances on deposit with the central bank and if there is a network externality in the settlement of payments, then the central bank has a natural role in the providing settlement in reserve account balances.

The payment card industry provides a challenge to claims that network and public goods considerations justify a significant Federal Reserve role in new payment systems. The credit card associations represent extensive payment networks that have developed through the cooperative efforts of private institutions. This cooperation has included coordination in the setting of standards. As suggested above, large cooperative ventures may pose problems of market power that need to be monitored. The very fact that such problems arise, however, suggests that private industry has been effective in recognizing the gains to cooperation in developing payment networks.

In addition to monitoring potential market power problems, public sector intervention in network industries has sometimes been motivated by distributional concerns. Price regulation in such industries has often taken the form of limiting discrimination among users. Such policies have attempted to lessen the impact of differences among users in the real costs of connecting to a network. Such differences might be related to geographic location or other user characteristics. Often the key difference is between a relatively small number of high volume users and a large number of low-volume users. If costs of using the network service are dominated by the fixed cost of connecting to the network, then high-volume users might enjoy a cost advantage. The fact that the advantaged class of users is relatively small in numbers implies that regulation to offset the advantage could easily win political favored. Peltzman (1989) argues that this sort of political concern for distribution has been a significant force in shaping public sector intervention in network industries. In discussions of payment system policy, discussions of access to payment systems by small, remote depository institutions seem to be articulating the same type of distributional concerns.

Note that public sector intervention may not be necessary for pricing arrangements that favor small users. Indeed, the three-buyer example in section two shows a situation where sustainable pricing of the efficient network requires that the small user pay less than his own connection cost. The small user is subsidized, because his participation in the network creates benefits for other users. The situation of the small user in the example might be analogous to that of a bank in a small town whose customers often go to a large city to make purchases with payment instruments that need to be cleared back to the small-town bank. If the flow of payments in the reverse direction (representing customers of city banks making purchases in the small town) is significantly smaller, then the small bank's private benefits and external effects from network connection are similar to the example. The private value of connection might be low, because the institution receives relatively few deposits of payments to be cleared to out-of-town banks. On the other hand, out-of-town banks might place a greater value on having this bank in the network.[8]

While private network arrangements might produce pricing that subsidizes small users, there is no general guarantee that the direction or magnitude or magnitude of subsidy will be the same as that driven by political considerations. Hence, the possibility for politically mandated intervention remains.

Finally, does the special characteristic of a payment system as a network for the settlement of debt obligations raise issues associated with the control of risk that require the attention of the Federal Reserve? In its capacity as lender of last resort, the Federal Reserve has an interest in the risk taking activities of financial institutions and, hence, in ensuring that payment arrangements in which those institutions participate have appropriate mechanisms for monitoring and controlling both credit and operational risk. In the first instance, the risks associated with a payment network is limited to that network's members. Unless these members' risk taking incentives are distorted by the presence of an underpriced (implicit or explicit) financial safety net provided by the government or central bank, participants have an incentive to monitor and control the risks into which they enter. The presence of a safety net may increase the public sector interest in monitoring risk-taking. This, however, is not particular to institutions' participation in payment networks. Hence, as above, network externalities and public good considerations do not, by themselves give rise to any particular need for central bank participation in emerging payment systems.

---

[8] Note that the discussion in this paragraph applies primarily to a payment instrument for which the direction in which clearing and settlement messages flow is the same as for checks.

References

Besen, Stanley and Joseph Farrell, 1994. "Choosing How to Compete: Strategies and Tactics in Standardization." *Journal of Economic Perspectives*, v.8, Spring, pp.117-132.

Economides, Nicholas, 1996. "The Economics of Networks." *International Journal of Industrial Organization*, October.

Henriet, Dominique, and Herve Moulin, 1996. "Traffic-Based Cost Allocation in a Network," *Rand Journal of Economics*, v. 27, Summer, pp. 332-345.

Katz, Michael, and Carl Shapiro, 1994. "Systems Competition and Network Effects," *Journal of Economic Perspectives*, v.8, Spring, pp.93-116.

Lacker, Jeffrey and John Weinberg, 1992. "Coalition-Proof Equilibrium for a Private Information Credit Economy," *Economic Theory* v.3, pp.279-296.

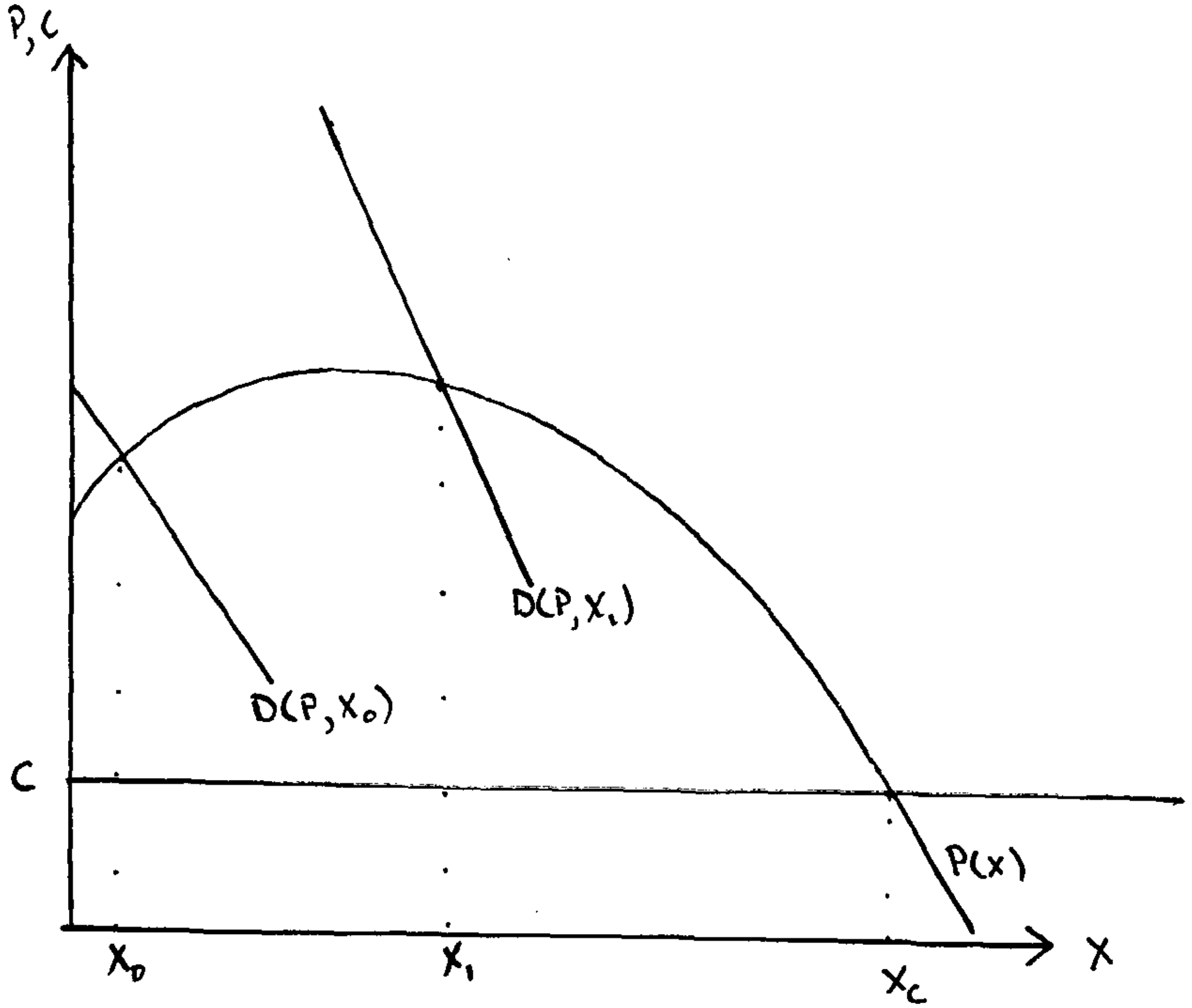
Liebowitz, S. and Stephen Margolis, 1994. "Network Externality: an Uncommon Trgedy," *Journal of Economic Perspectives*, v.8, Spring, pp.133-150.

Peltzman, Sam, 1989. Comments on "The Control and Performance of State-Owned Enterprises," in MacAvoy, Paul, et al. (eds.) *Privatization and State-Owned Enterprises*, Kluwer Academic Publishers, Boston.

Sharkey, William, 1992. "Network Models in Economics," in Moder and Elmaghraby, (eds.), *The Handbook of Operations Research*, Van Nostrand Reinhold, New York.

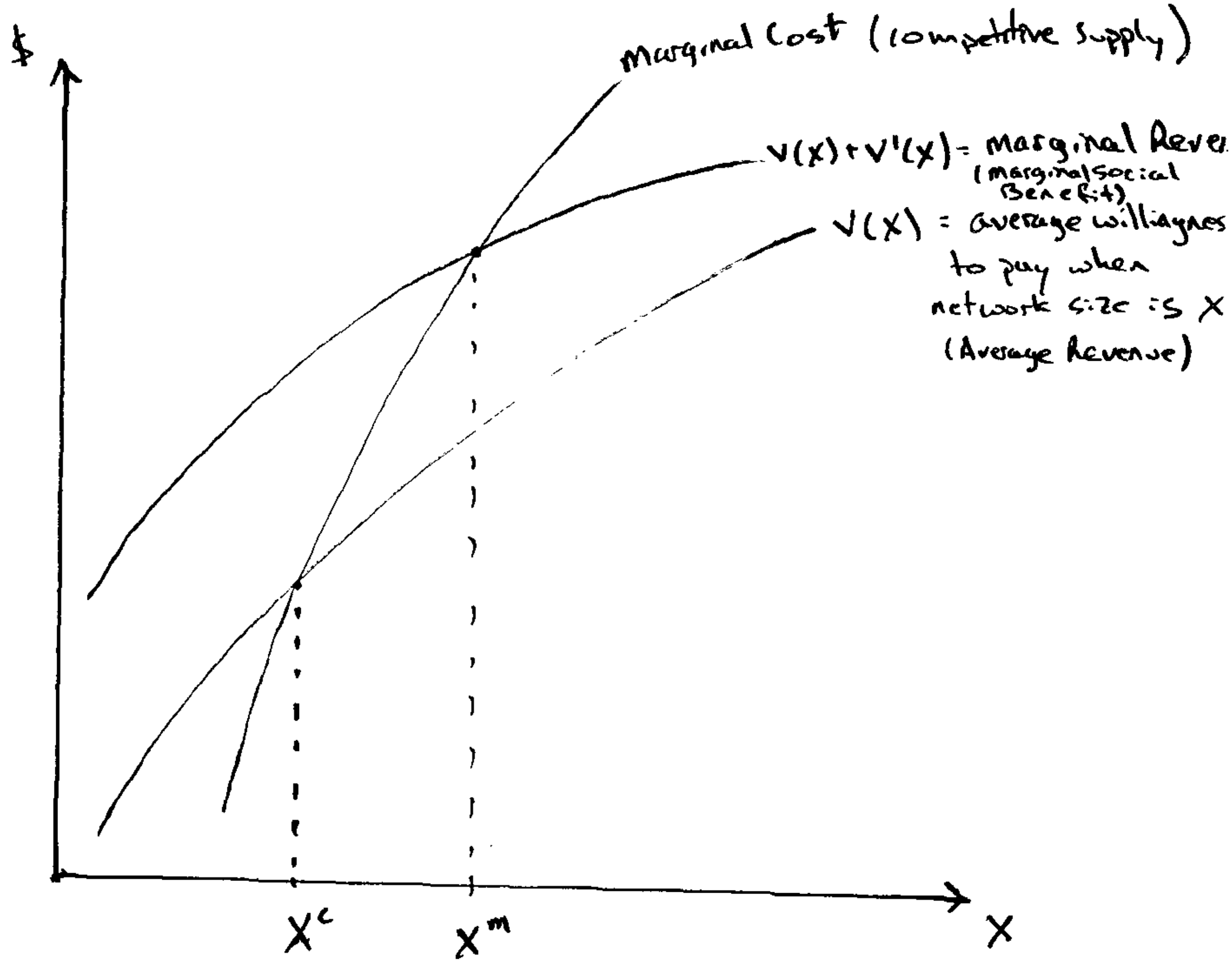
Telser, Lester, 1994. "The Usefulness of Core Theory in Economics," *Journal of Economic Perspectives*, v.8, Spring, pp. 151-164.

Figure 1



D(P, X_0) is quantity demanded at price P if assumed network size is X_0. D(P, X_1) is same when assumed size is X_1. X_c is competitive equilibrium quantity (network size).

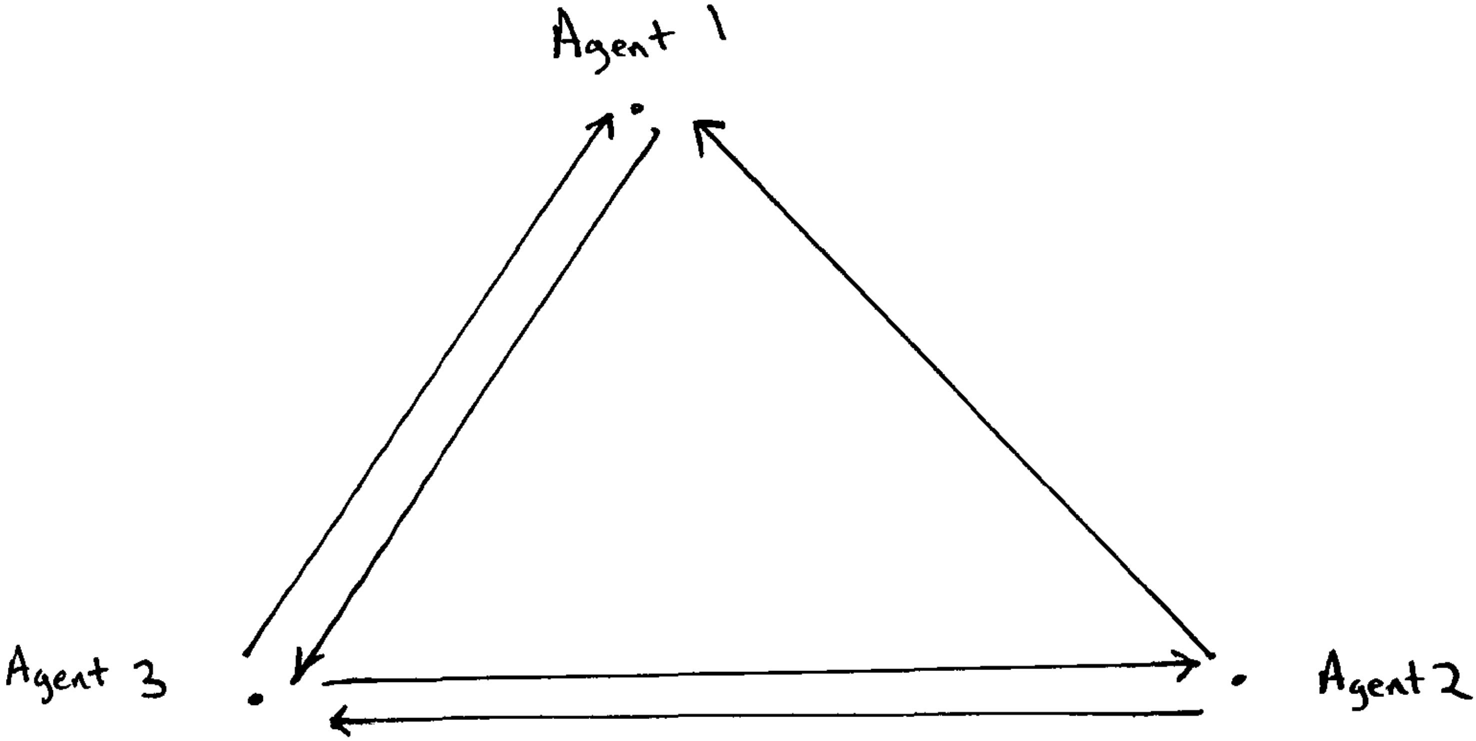
Figure 2



X^c is competitive quantity

X^M is monopolist quantity.

Figure 3



An arrow from agent i to agent j indicates that i wishes to send a message to j.