New telecommunications services

Network externalities and critical mass

David Allen

Network externalities – the requirement that there be a group of subscribers if communications are to occur – play a central role in the demand for new networks. And as telecommunications evolve, new networks have increasingly taken attention. In an effort to describe the demand for new networks, this paper investigates the critical mass phenomena that characterize network externalities. The experience to date with the Minitel information service in France serves as an informal empirical check on conclusions. With this base, the paper then draws implications for the domain appropriate to regulation and for universal service pricing under liberalization. It concludes with a speculation about the role of mixed economy, illustrated by the construction of Japanese universal service.

David Allen can be contacted at the Centre for Technology, Policy and Industrial Development, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

The author would like to acknowledge Charles Jonscher and Jeff Rohlfs, who are cited in the text. He appreciates suggestions from Hajime Oniki, and also appreciates discussions with Cristiano Antonelli, Henry Elkington, Alain de Fontenay, Marcellus Snow and Paul Verdin. Greg Duncan and Haruko Yamashita will also find that their observations have shaped this text.

This paper in its present form was first presented at the Fifteenth Telecommunications Policy Research Conference, Airlie House, VA, 27–30 September 1987.

Until about two years ago the prospects for videotex seemed theoretical at best. Numerous trials around the world had produced what might be termed an experimental response from the user – in each case the trial service penetrated to only a nominal percentage of the intended user population. Then the French videotex effort, Minitel, in the language of those who monitor new services 'took off. With terminals currently in place for roughly 15% of all French telephone subscribers, with usage running at 65 million hours annually and with some 8000 services vying for the users' favour, Minitel has become a dramatic contrast to the 1970s' unyielding effort to implant videotex among an entire society.

Although Minitel has proved a remarkable case, videotex is only one instance that may excite interest. The same phenomenon may characterize other new telecommunications services: including the wholesale reconfiguration of a national network (such as ISDN), or new forms of the local network (such as LANs). It is difficult to know, however, how to identify which of several trial efforts will take off and become a new kind of network, and indeed what distinguishes the successful new network from those that sputter and fail to become realized.

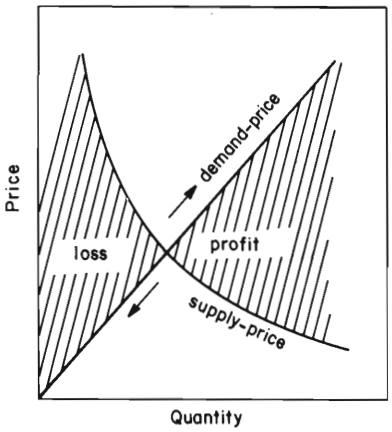
Much of our recent understanding about the demand for networks has been based on experience with relatively mature telecommunications systems. However, a startup network faces a special circumstance that differs in one key respect from the mature network. Since the fundamental *raison d'être* of a network is to connect people, a new network by its very nature requires a group of subscribers if it is to startup. This need for the critical mass group surely distinguishes the startup from the mature system.

A mature system has necessarily moved beyond that point in its development where a critical mass has initially assembled. As a consequence, new subscribers to a mature network can join one after another rather than as a group. That indicates why we should not expect to find evidence of network externalities in data from a mature system.[1]

Not all new telecommunications services have the characteristic of a

0308-5961/88/030257-15 $3.00

Network externalities.



Standard microeconomics.

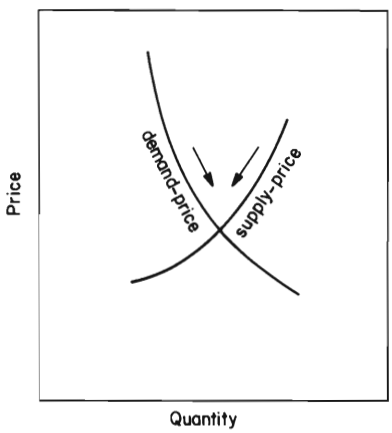


Figure 1.

The intersecting curves in these two comparative graphs are, of course, the supply and demand curves fundamental to economic analysis – the lower graph representing the standard analysis, the upper the analysis for network externalities. The arrows indicate the direction of the equilibrium, or disequilibrium, forces that are at work in the two situations. And on the upper graph, the shaded areas also indicate by their geometric size the amount of the supplier loss, then profit, that is generated in the case of network externalities.

[1]The economist's notion of externalities can be construed to denote the process whereby a group, rather than an individual, makes (or forebears to make) economic choices. Construed in this way, the new denotation for the idea would derive, albeit awkwardly, from the original framework, which had put the individual decisionmaker at the centre of the conception of externalities. With the individual as starting point, choices by a group would be 'external' to any individual.

Perhaps it is worth recollecting that it is not necessary to espouse centrally planned economies, nor for that matter is it

continued on page 259

new network. For instance, services that only replace existing mature arrangements – including so-called 'cream skimmers' which operate via the benefit of some universal service pricing umbrella – require little in the way of new behaviour by subscribers, nor do they entail so much new plant and equipment on the supply side. It seems clear that new telecommunications services may have more or fewer characteristics of a completely new network.

The analysis in this paper is most directly applicable when the new service involves significant departure from past practices on both the demand and the supply sides of the network – that is, when network 'newness' is fullblown and network externalities and critical mass play their largest role. Full motion videoconferencing is a case in point; so would be an interactive broadband service to the home.

Voicemail involves less change, and therefore the analysis here is less directly applicable. Nonetheless, there is below a consideration of how critical mass in networks may also relate to mature systems and so, by implication, to those new services which do not represent such a large departure from mature arrangements.

Factors for success of new services

Externalities and critical mass are, of course, only one factor which determines whether or not a new telecommunications service succeeds. The service must also fulfill some requirements of human communication. Obviously, whether such human needs are met – needs perhaps for visual tools, message storage, mobile reception – constitutes other vital factors in the success or failure of a new service.

What puts externalities and critical mass in an unusual role is their capacity, at least within a limited range, to determine whether these other factors are accepted in the first place by a group of prospective users. Critical mass, it devolves, is a sort of portal through which the other features must pass if those features are to make their impact.

The economics of critical mass

The existence of network externalities carries an essential message about the perception of value during a network startup: prospective subscribers perceive more value as the subscriber pool grows. Since the value of the network derives from its connectivity, and since a larger pool of subscribers means more possible connections, value increases with size, at least to the limit of the community with which any subscriber wants to communicate. This suggests a cost–benefit analysis with a demand–price curve such that, as the quantity of subscribers rises, so does price; that is, a demand–price (benefit) curve which ascends with a positive slope (see Figure 1).

In this particular case, the overriding function of price might be described as signalling among the individual members of the subscriber pool the fluctuating size of that group.

Completing the picture with a cost curve (supply–price curve) – supposing increasing economies of scale[2] – critical mass can be defined at least in terms of its economics. The case of network externalities can also be compared with the more typical analysis in standard microeconomics. The cross between supply and demand in the standard analysis has rotated around its plane to produce the cross that emerges in the network case. The implications of this bear examination.

The standard analysis watches price mediate an adjustment process between supply and demand, one that progressively eliminates any discrepancy of surfeit or shortage that separates them. That process promises convergence to a stable equilibrium at the anchorpoint of demand/supply intersection.

The network case, however, seeks a different style of analysis. Because benefits are interdependent, they signal group size among the demanders, instead of mediating supply and demand. This change from the standard case, along with the fact of supply by a monopolist, switches the focus of the analysis (in the case of the network) onto supplier incentives.

From startup to network maturity

During the startup of a network, the entrepreneur faces losses since cost exceeds price, while in maturity, when price increasingly spreads above cost, the network provider reaps burgeoning profits. *Dis*equilibrium results as the supplier incentives lead *away* from the point of intersection.

In this unbalanced arrangement, the point of intersection (which is an unstable equilibrium of just-matched revenues and outlays) defines the network's critical mass. Upon reaching that critical mass point, the network no longer faces the losses that are endemic to startup. And if the network then somehow moves off from its unstable equilibrium in the direction of a larger subscriber pool, the growing profit underwrites an internally generated snowball effect up the demand curve.

Subsidy investment in network startup

This analysis also offers some detail on the startup campaign which any new service provider will face. The effort to overcome built-in inertia and reach critical mass will require the investment of a subsidy[3] in the amount of the left-hand shaded area in Figure 1. Additionally, startup may begin with zero pricing; later, price can rise along with perceived value.[4]

This approach to the economics grew out of a suggestion from Charles Jonscher four years ago. About a decade before that, Jeff Rohlfs had derived the positive-sloped demand curve.[5] Charles Jonscher suggested the demand–supply cross, which I have dubbed the Jonscher cross.

Problems of demand

Clearly, however, there is a glaring inadequacy in this analysis of critical mass, at least as it stands. Paul Samuelson has put it succinctly.[6] In his classic text he warns, in an injunction to the neophyte, that 'the fallacy to be avoided like the plague is [. . .]: "I have disproved the law of downward-sloping demand."'

Logic seems to argue – even persuasively to some perhaps – that the above account of network externalities can but make sense since demand is interdependent: value surely grows as the favourable choices of demanders accumulate. In opposition to this reasoning, however, the increase of quantity with price apparently strains beyond credibility, at least when judged against the decades of contrary theorizing and empirical confirmation that now constitute demand theory.

Is this account even sensible? Could the events of network startup, if inspected closely, reasonably be expected to transpire even loosely as

continued from page 258
necessary to oppose a paradigm of individual choice, if one wants nonetheless to accept that groups, composed of individuals, do on occasion make choices – as occurs in the case of network externalities.

[2]Of course, the debate as to whether telecommunications is a declining cost industry is voluminous and lengthy. In particular, the rationale for the divestiture at AT&T held, as one fundamental, that the economies do *not* increase. Despite this, subsequent evidence has failed to settle the question in any conclusive way; in fact, the increasing economies proposition, at least to some observers today, seems more than plausible. And any economies from declining costs are here compounded by a Chamberlinian sort of 'local monopoly' that results from the network innovator's unique position.) Since the overarching topic in our main discussion is demand, this is not the place to pursue the debate about supply.

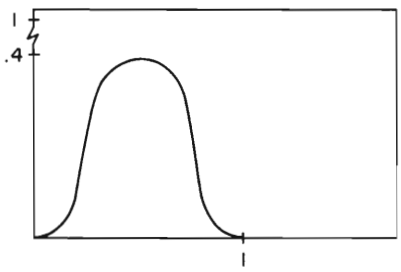
[3]This phrasing, which purposely mixes the ideas of investment and subsidy, was chosen to cast startup investment in terms of the cross-subsidy debate.

[4]Of course, the relevant pricing here is pricing for access. Since the central phenomenon of concern to us is interdependence of choice, and since that interdependence is operationalized in the decision to choose access, our interest in pricing centers on pricing for access. When the primary role of access has been accounted in the analysis, then pricing for usage becomes a second order of consideration.

[5]Jeffrey Rohlfs, 'A theory of interdependent demand for a communications service', *The Bell Journal of Economics and Management Science*, Vol 5, Spring 1974, pp 16–37.

[6]Paul A. Samuelson, *Economics*, 11th edition, McGraw Hill, New York, 1980, p 59.

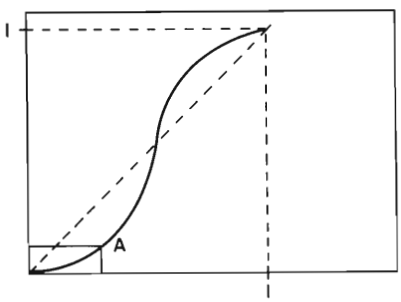
This proportion of the subscriber pool requires . . .



. . . this minimum size (expressed as a proportion of the whole group).

Figure 2.

Subscribership – actual:
This proportion actually want to join.



Subscribership – expected:
If this proportion are expected to join.

Figure 3.

outlined? That account hints at some underlying dynamic, some process whereby demand first must be 'pulled', as it were, up its curve, and then after reaching critical mass demand pushes itself to maturity. Thomas Schelling has suggested an analytic device to inspect such events in some detail.[7]

Choices of critical size

The network externalities carry another message: that a prospective subscriber will actually decide to join only if some minimum number of the other prospects also decide to join. Each person has, as it were, an individual vision of what constitutes critical mass. That personal view may vary between individuals. Yet it seems from the earlier pass through these events that the network can succeed if it is able to gain just one critical size. There are challenges to understanding the process of meshing different individual choices of critical size when that process appears able to produce coalescence for the group as a whole at just one size.

'Watching the group'

It seems likely that individuals base their choice on what they expect the others to decide. Thus, the individual's effort to decide hinges upon 'watching the group' – the other members in the community of actual/potential subscribers – to discern what the group choice may be. That would seem to be the case for each prospective subscriber. The outcome for the group then turns literally upon everybody watching while being watched. Change, and finally critical mass for the group, floats it seems on shifting perceptions of what the group outcome may be.[8]

How can this interworking of separate but interdependent individual choices be portrayed? The minimum critical size required by each individual could be learned from the prospective pool of subscribers. With that data a frequency distribution can be plotted which will tell for each minimum critical size what part of the population would require that particular size. For purposes of illustration, let us suppose a normal distribution (see Figure 2).

Expectations of membership

Those frequencies are then converted to a cumulative distribution. The cumulative diagram (Figure 3) reveals for any given critical size how many prospects require that size or less as a minimum, and therefore what part of the total population would actually want to join the new network if it were expected to be that size. The vital question is whether the *actual* number who would join either falls short of or exceeds the *expected* membership.

If the original frequency distribution is normal, the cumulative picture arises, as is shown in Figure 3. Since the broken line at 45° is a visual cue which indicates where the two variables equal each other, it provides a handy reference by which to answer the vital question. In other words it helps tell at a glance for any size of expected membership whether actual results will exceed (or just meet) and therefore serve to confirm the expectation of membership, or, instead, whether expectations will be dashed by the actual outcome.

For instance, a company's new intracorporate videotex service could

[7]Thomas C. Schelling, *Micromotives and Macrobehavior*, Norton, New York, 1978. See Chapter 3 in particular.

continued on page 261

be at point A with membership expected to reach the size indicated on the horizontal scale, perhaps because the company has installed that many terminals. However, it is clear from the corresponding point on the vertical scale – or more simply, by noting that A is below the 45° line – that too few employees will actually use the service and expectations will not be met. Prognosis for the startup, given only these facts, would not be hopeful.

Creating a shared expectation

Although the outcome of the process described here begins with fixed individual needs for critical size, it turns upon the repeated formation of a changing and shared expectation about network size. (See Appendix for details of how events inspected up close can portend conclusions such as this.)

If this observation about the expectations is combined with the earlier proposition about subsidy, it suggests the purpose that subsidy should serve, the subsidy without which, according to the earlier views, the critical mass will not assemble. The subsidy needs to create a shared expectation that subscribership will be larger than critical mass. If that occurs, enough users will be attracted so that, simply by the fact of their presence, they will then trigger the growth remaining to reach universality and so full network maturity.

As the detail in the Appendix confirms, the underlying process moves in an incremental fashion through repeated disappointment, or satisfaction, of the shared expectation. Beyond such a step-by-step process, however, there is a larger phenomenon.

Consensus formation

As the expectation changes in an incremental fashion, it ratchets up or down one step at a time. Eventually, however, these increments may cumulate and produce a more profound change in the group's expectation or outlook. The new network begins as an innovation and a curiosity, but if in time it does succeed, it is transformed and becomes a necessity, entrenched in the habits of the subscribing group. The incremental steps can lead, in other words, to a larger shift in outlook that considers as necessary what was earlier viewed as novelty.

Thus, what begins as an attempt at a new network (such as videotex) may become a new orthodoxy in communications. Whether this fledgling effort does actually emerge into orthodoxy depends upon whether the number of subscribers grows large enough to confirm, for the entire prospective subscriber pool, that the network has achieved legitimacy.

The shift from novelty to necessity – the larger-scale discontinuity in expectations – creates, in other words, a sense of new value. It elevates an attempt at a network innovation to the status of essential commodity, sought after by the full pool of potential communicators, so that the once fledgling entrant finally emerges as the preferred mode of contact.

Perhaps it is this capacity to establish new economic value that is the real focus of interest in the larger investigation. This creation of new value is certainly what we notice from the less visible, incremental shifts in expectation.

Price also appears to be in the service of value creation. In this case, network price can signal the changes in value which the interdependent

continued from page 260

[8]This characterization can be seen in a different light. Another related line of research has argued that 'history makes a difference'. Paul David's work is one prominent example; see Paul A. David and Julie Ann Bunn, 'The economics of gateway technologies and network evolution: lessons from electricity supply history', Draft paper, High Technology Impact Program, Center for Economic Policy Research, Stanford University, September 1987. If I take some liberties in describing the line of his argument, then I might say that events are interrelated in a complex way and the interrelatedness reaches across time.

In that light, 'watching while being watched' can be seen as one element in the human process that is necessary to assemble an adequate picture of these events that are related in a complex way over time. Or, put differently, our view of complex reality is a social construction, and the 'watching' is a key step in that construction. We individually rely on each other in an integral way for our view of a complex world. In this light, understanding how knowledge emerges from the social process – understanding the 'watching' becomes a useful enterprise.

Perhaps it is also worth noting that the notion of economies of scope appears to reflect one attempt to acknowledge the interrelatedness. Also, reversibility in demand theory, and more generally the *ceteris paribus* style of logic, stand in some contrast to a 'history matters' approach.

demanders are reflecting (changes in value based, of course, upon the changing size of the group). This signalling is concentrated within the group of demanders.

However, price is only one of numerous signals in this process that might be called consensus formation. Consensus formation, it would seem, is the subject to understand. And the signals that are worth studying must be human communication in its significant richness – the communication, that is, which comprises an endless tug of war between potential consensus and potential diversity.[9] Nor do I expect that the price signal is likely to be the principal independent variable to determine outcomes in the human give-and-take of consensus formation.

Recent work on standardization[10] addresses some of these questions, though the game theoretic approach that is typically used tends to emphasize the combative over the consensual aspects (despite efforts to redress that balance) and also stylizes away vast realms of real human communication. In an earlier era, Robin Marris,[11] and before him James Duesenberry,[12] considered the range of questions here. Though Marris specifically rejected a positive-sloped demand curve as not being a useful approach, he and Duesenberry both emphasized what I have termed value creation. At about the same time, J.R. Hicks touched on some of the preliminaries to these questions.[13] Finally, Schelling's model is one in a class of such models; several economists are interested in demand phenomena that are at first threshold limited, and then self-generating.[14]

Minitel: a case in point

If Minitel is to be used as an informal check of these ideas, it must first be examined whether the ideas can be applied to videotex in general. The question must be asked whether videotex is a new telecommunications service that involves externalities.

The evidence from Minitel, as well as from a number of other videotex trials, is that person-to-person communication amounts to a significant portion of the traffic. One example is the usage experienced by five major services on Kiosk, the segment of Minitel which serves the general public. About half of the activity on these five main services has recently been devoted to various forms of messaging (person-to-person contact) which necessarily entail the classic network externalities.

Perhaps the role of person-to-person contact on Minitel is so central because it has emerged from communication that initially began as person-to-machine. Database queries (from person to machine) on the subject of some niche interest (for example sailing) have made it possible, as a serendipitous side effect, to identify the subgroup of subscribers with a common interest in sailing.

Previously these individuals had primarily print sources of information on sailing, such as magazines, which offered no easy means to identify other individuals with a like interest. Once an electronic, person-to-machine 'magazine' made it possible to identify the group however, person-to-person contact could begin to flower.

Applying the analysis

Person-to-machine usage has served to establish an expectation that a critical mass of like-minded discussants will subscribe to a related

[9]Communication, within the group, then becomes our primary focus to understand network startup. That our subject is the technology which supports this communication only spices the fray. To study the interrelation between adoption of a communications technology (ie, network startup) and its own support role in that adoption, my inclination is to start with human events and then ask how the technology may interact in the evolution of outcomes.

[10]Joseph Farrell and Garth Saloner, *Installed Base and Compatibility: Innovation, Product Preannoucements and Predation*, Working Paper Number 411, Department of Economics, MIT, Massachusetts, revised February 1986. Also see Michael L. Katz and Carl Shapiro, 'Network externalities, competition, and compatibility,' *American Economic Review*, Vol 75, 1985, pp 424–440. Also S.V. Berg, *Technological Externalities and a Theory of Technical Compatibility Standards*, mimeo, University of Florida, Florida, revised 7 November 1984. Stanley M. Besen and Leland L. Johnson, *Compatibility Standards, Competition and Innovation in the Broadcasting Industry*, Rand, Santa Monica, November 1986 offers a fairly complete review of the work, including several not listed here.

[11]Robin Marris, *The Economic Theory of 'Managerial' Capitalism*, Free Press of Glencoe, New York, 1964. See Chapter 4 in particular.

[12]James S. Duesenberry, *Income, Saving and the Theory of Consumer Behavior*, Harvard University Press, Cambridge, MA, 1949. See Chapter 6 in particular.

[13]J.R. Hicks, *A Revision of Demand Theory*, Oxford University Press, Oxford, 1956. See particularly sections 3 and 4 in his summary. Of course, a more historical provenance must reach back at least to Pigou.

[14]Just one example is W. Brian Arthur, 'Competing technologies and lock-in by historical small events: the dynamics of allocation under increasing returns,' Technological Innovation Project Working Paper No 4, Center for Economic Policy Research (Publication No 43), Stanford University, 1985.

Since the main subject under discussion here is demand, and the discussion of critical mass in supply is a full-blown topic in its own right, it is not explored here. A comment on diffusion, however, does seem appropriate.

Suffusing the desire, throughout a community, for some new technology – one description for value creation – also describes, of course, the diffusion of an innovation. If the horizontal scale in Figure 3 is renamed 'time', which would be an appropriate alternative, the critical mass process nicely reproduces the S-curve that is typically found in research on diffusion. That seems at least suggestive of the role which externalities (perhaps in a weaker form but nonetheless there) may play in

continued on page 263

person-to-person sub-network. Or even more generally, the mechanism to implant this expectation involves a symbiotic interplay between separate critical mass scenarios for a pair of related videotex services, one person-to-machine the other person-to-person. Critical mass for the person-to-machine half of this pair is largely concerned with its supply side because there are not strong externalities on its demand side.

Minitel and pricing

Minitel is particularly striking because it has followed the contrary pricing path that the earlier analysis for externalities suggested. Service began with a zero pricing policy. As perceptions of value have risen with growth in the subscriber pool, it has been possible to begin charging prices which reflect the increase in value, and in particular to begin charging for terminals, the surrogate for access to Minitel.[15] As of late 1987, about 8% of all terminals in place generated a fee (though only a portion of these represented access).

Subsidy and investment

Clearly, Minitel is committed to a significant subsidy in its effort to surmount what would here be called its point of critical mass. From the American perspective, there is a temptation to conclude that this subsidy has somehow been disproportionate and too large. In France such questions have also stirred political controversy.[16] One point of comparison is the amount of investment that has been expended in the USA to establish videotex there.

Estimates of the expenditures on videotex in the USA range from several hundred million dollars to as much as $1.5 bn – and of course these efforts have not yet borne similar fruit to the French programme. Since the French investment is probably not significantly larger, perhaps the size of their subsidy is after all not so disproportionate.

Free and blanket installation of terminals has proved the most remarkable aspect of the subsidy in the eyes, anyway, of the world telecommunications community that has been watching. But it is worth noting that all the expenses of operation – terminals, but also all other activities – are surely important to an effective subsidy.

For instance, I expect that communications (about the progressive spread of the network and about its widening usage) have played a key role for Minitel – at least that is what the ideas here would suggest. To the extent that those communications were funded by marketing expenses of Minitel, they would appear to be another important element in the subsidy.

Value creation

There is one point that was not drawn out explicitly in the discussion of value creation, but the implication is clear. Part of the challenge in critical mass scenarios is to change general perceptions of a new service from a view that sees novelty to one that sees necessity instead. By using customer premises equipment that were as little different as possible from the ubiquitous voice telephone, and also by giving the system a necessity service (the directory service),[17] Minitel reduced the level of this challenge to a minimum. That of course improved the odds for critical mass. In more straightforward terms, such a tactic might be described as reducing the fear of an unfamiliar activity.

The necessity directory service also played another important role, it

continued from page 262
the spread of many sorts of innovations, even those where interdependence appears to be absent.

But this is not the place to inquire whether the analysis here may apply to investment/subsidy in any generic new product. Instead I will just note that in the diffusion literature Bass models, and the progeny of Bass models, visualize innovators as distinct from imitators. By contrast here, innovation versus imitation is treated more as a gradient, with some mix available in each person.

[15]As Ref 4 noted, it is the price of access to Minitel, not the price of usage, which is relevant here.

[16]Again Ref 3 regarding subsidy and investment is integral to the questions in this controversy.

[17]Of course the full history is that the paper directory was eventually reinstated for Minitel customers. But it is easy to suspect that during the early stage when Minitel was the sole directory, the necessity aspect made itself felt when critical mass was most at stake.

appears. The presence of a necessity confers some guarantee of usage and, therefore, of active subscribership. Thus, incorporating a necessity component in the service design should help to cinch the expectation that there will be the largest possible network membership – hopefully a number greater than critical mass.

Forming the expectation

The core objective, of course, is to establish an expectation that subscribership will exceed critical mass. In fact, Minitel employed several means to implant the expectation. One important means was a manageable campaign whereby Minitel was introduced sequentially in one geographical region after another. Critical mass could be begun in a region where prospects were strongest. As the campaign moved to progressively less fertile ground, early successes could serve as a growing foundation to support the later, more difficult efforts and thus help to build towards a regionally combined, nationwide expectation for critical mass.[18]

Another fundamental means to the required expectation was probably a separate critical mass on the supply side.[19] After initial conflict with the press, Minitel was able to bring together banks, newspapers and the computer industry, along with the PTT, into a combined effort that could orchestrate network, system, services and terminals.

Presumably, the function of this critical mass on the supply side would be to bolster the expectation that connectivity will be available. That expectation might best be strengthened when the prospective subscriber is able to try and use a repertoire of new services. It seems likely that Minitel's suppliers, coordinated in their efforts because of their critical mass, could certainly provide a fair modicum of variety in new services.

How can the interaction between these two supposed critical mass phenomena, the one in supply the other in demand, be captured? One possible approach is that for Minitel the expectation that supply would be available led developments on the demand side. However, by indirectly undergirding the expectation that subscribership demand could exceed critical mass, the development of supply impacted on both sides.

Indeed, Minitel lends itself to an interpretation along the lines that have, so far, been developed here. Perhaps interpretation is suited to a setting where it is easy to imagine a critical mass at stake, where, that is, a PTT takes decisions about a network that can potentially connect an entire society.[20] However, the liberalized and actively competitive environment in the USA is a different case. How critical mass assembles when the action is in pieces instead of as a whole must be examined.

Critical mass in competitive markets

The essential issue is the size of what we might call the critical mass group. The competitive model is already apparent in Minitel's sequenced introduction, region by region. Sequencing means that the larger outcome is disaggregated into smaller, individual critical masses, which, of course, describes competition.

So what are the differences between the competitive environment and the single decision by a PTT? A major difference, it appears, is the separate critical mass on the supply side. In the PTT case, a supply side critical mass with the expectation of available supply that it engenders

[18]With this sequential growth pattern, the ability of the supply side to be appropriately responsive turned upon there being a flexible network architecture. To create the decentralization essential for this flexibility, Minitel pioneered some of the innovations necessary for decentralized operations on such a large scale.

[19]Even though the primary discussion has not emphasized issues in supply, it seems likely that critical mass in supply and its interaction with the parallel phenomenon in demand may be useful or even essential for a satisfactory conceptualization of these events. But, as already noted, that discussion is not included here.

[20]Has Minitel in fact achieved a nationwide critical mass? To begin to answer that question, in any more than an informal way, we must take a closer and more careful empirical look. One of the possible tools? Sample to determine the distribution of individuals' needs for critical size and then construct the analysis described earlier.

can support spread of the main target expectation on the demand side – spread, that is, from one group in a sequenced introduction to the next group.

The competitive situation, without such a supply-side link for its groups on the demand side, might fall back instead on excess capacity – supply, that is, which has been left over from some prior round of critical mass. In the competitive environment, in other words, excess capacity from a prior round, as a surrogate for linkage from the supply side, may serve to support some futher round of critical mass for some later group.

The question then arises of how critical mass assembles in a competitive situation, how the shared expectation forms and creates a new sense of value – is some planner needed to manage the subsidy? Certainly an intracompany planner or manager may decide on an investment that is intended to build confidence in a new telecommunications service that the company is offering. But the spread of a service in the competitive setting, so that different providers' different offerings *interconnect* with or within a company, is surely more serendipitous than planned.

A further question for analysis is the speed of service spread, and how the competitive situation compares with the PTT in reaching whatever will be universality for a given service. Perhaps the PTT with its planned environment can move the situation more quickly to a sense of new value – particularly if it guesses correctly as to service and configuration and if in addition it does not lose a first mover's advantage to some second mover who learns from the PTT's experience.[21] Does this suggest, then, that planning has some natural advantage?

I believe that would be premature, and return instead to the troubling proposition at the core of this analysis – whether it makes any sense to suggest that demand can be upward-sloping.

The process of creating a sense of new value might be decomposed, for purposes of the analysis, by considering the constituent subgroups which together comprise any whole subscriber pool. In the case of videotex these are the sailing *aficionados* (to return to Minitel for an example), the chatline users, the business subscribers, the professionals at home, and so on, each in his or her subgroup.

Upward-sloping demand curve

At the beginning of network startup, one or more of these subgroups try out the new network. Each such initial subgroup must then, it would seem, traverse its own process of creating a new sense for what is the preferred network; must reach, in other words, its own critical mass within its own membership regarding this network innovation. Later, there could arise the prospect for critical mass across several of these subgroups and thence for communication across subgroup lines. Later even, if the new network has the potential to span a society, there might be the prospect for joining across these intermediate groupings and so to reach a universal critical mass.

So far this is no more than a slightly formalized restatement of sequenced introduction, the notion that related the case with a PTT to the scenario with competition. Can this description of events help to address the present question about the tilt of demand curves?

It is useful to look at the upward-sloping demand curve in terms of polar coordinates, and to compare the three such curves that would

[21] Of the work on this endlessly debated question, Farrell and Saloner have a recent intriguing piece that illustrates how planning may have an advantage in speed. See Joseph Farrell and Garth Saloner, *Coordination Through Committees and Markets*, Working Papers in Economics E-87-25, Hoover Institution, Stanford University, June 1987.

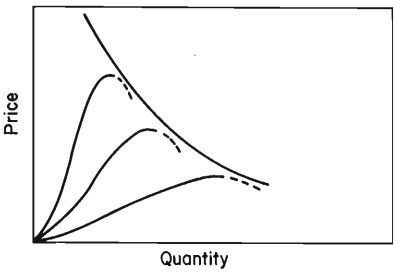


Figure 4. The long run average demand curve.

represent the three levels of critical mass – subgroup, intermediate and universal. Then suppose that the value added by that marginal subscriber who just edges the group up to the critical mass point is, when compared among these three levels, *least* for the universal case. Once, in other words, a telephone system has reached the verge of universal service for a society, the value of critical mass is less than the value of critical mass for the smaller, more compact intermediate stage; and that, in turn, is less than for the individual subgroup. This result, which is at first glance paradoxical, seems likely to arise since the desire to communicate will be strongest within the individual subgroup – for instance, stronger than the desire to communicate with, say, anyone in the society.

In polar coordinates, the upward-sloping curve for the individual subgroup has the largest angle: the highest value for the smallest network size. The effect of adding together the subgroups to cumulate to the intermediate stage, then adding together intermediates for universality, is, in polar terms, progressively to close the angle. And finally, when these three curves are drawn from the same origin, their separate critical mass inflection points perhaps begin to trace out what is suspiciously similar to the downward-sloping demand curve (see Figure 4).

The name I have given to the downward-sloping portion of the curve – long run average demand curve – as it accretes through the combining of progressively more subgroups, of course reflects its conceptual symmetry with the long run average cost curve. Jeff Rohlfs presaged this formulation, in which the dynamic growth phase of a network innovation folds over into demand for a maturing network, when he posited that upward-sloping demand would bend and be followed by a conventional segment.[22]

The Minitel evidence

The price of access, which at present in Minitel is the fees paid for some of the terminals, has become positive with growth in the subscriber pool. France Telecom's plan is that, with time, this charge will eventually fall back again to zero. The pattern of a transient rise in price during startup, followed by a reversal and decline in price with the onset of maturity, is of course the path described here.

However, a response to this could be that if this does have any usefulness, the upward-sloping portion is not a demand curve; it is rather the series of equilibria traced by a shift that produces a set of demand curves.

Revising the view of demand

The question arises, where is the conceptual tool for understanding the emergence of demand for a new, interdependently demanded service, where is a demand curve for the creation of new value in a network? The present conceptual machinery, following the line in the above response, would use the equilibrium points from a series of demand curves.

Indeed it seems likely that our mental apparatus, in its effort to conceptualize, is not so comfortable if deprived of static points, such as equilibria, to fix upon. But understanding of demand did advance when demand was conceived as a shifting curve.[23] Can the view of demand now be further revised to describe the dynamics of value creation in a

[22]See *op cit*, Ref 5. Of course, the argument that the marginal subscriber adds less value for the universal case than for, say, the individual subgroup does not by itself indicate that the *total* value for critical mass at universality is less than the total value for critical mass within the individual subgroup. In fact, a lower value for universality, which after all is composed of subgroups, may at first seem paradoxical.

But if the same marginal analysis holds for each instance of the adding together of subgroups, if adding one subgroup to another results in a lower value being attached to communication across the two, then the apparent paradox dissolves. The marginal result of each accretion of a new subgroup leads to a steady progression down the usual, long run average demand curve. And upon reflection, it is becoming clear that critical mass for universality holds less value than critical mass for, say, a small subgroup.

[23]A summary of this history by Elster and Hylland is succinct but pungent. See Jon Elster and Aanund Hylland, *Foundations of Social Choice Theory*, Cambridge University Press, Cambridge, UK, 1986. See page 1 in particular.

parsimonious way? Can a new conceptual tool be developed – demand curve or otherwise – which shows with simplicity the *change*, across that transient period, when interdependent demand emerges?

To the extent that that may be possible, upward-sloping and downward-sloping demand perchance have been got to lie down together, at least in these pages. What does such a formulation portend?

Since the industrialized economies have begun to liberalize their telecommunications sectors, a persistent question re-occurs: how to locate a boundary that will separate telecommunications activities that are properly market-based from those that are government-regulated.

Liberalization and the regulatory domain

Without much question, the US situation epitomizes this quandary about the regulatory domain. The FCC has now seen fit to convene, over the last 20 years, three major inquiries into the question. Each of these three episodes has proposed a new definition for the boundary line. The first two sets of definitions were each superseded on grounds that they had each proved to be inadequate, and the most recent Computer Inquiry III proposals seem to offer little prospect for being more enduring or stable.

And this turbulence in the quest for a boundary line has also been accompanied, through a separate proceeding, by divestiture. As the major act of liberalization in the USA, divestiture responded to a second boundary question that was related to the first. A division between the regulated and the market-based seems clearly to obtrude as a basic issue.[24]

The new view here – now with a focus on the demand side of events – could suggest a regulatory divide at the point of critical mass, between the growth phase in the first half and the period of maturation in the second half of the process. The first half is, of course, along the upward-sloping segments and the second part along the downward-sloping. Pricing in the first half would be subsidy-based (for example cross-subsidies from long distance to support the universal spread of voice telephony). Pricing during maturation, along the standard demand curve, could be conventionally competitive.

More fundamentally, such a divide is dynamic in its nature – it must move over time to accommodate rises and falls in new technologies, in new demand. That is to say, there is a reason why the search for an 'equilibrated' regulatory divide is repeatedly frustrated. That surely suggests that attention needs to be turned from an equilibrated solution and onto requirements for the flexibility which can support a continuing sort of change.

If the flexible approach that is contemplated here does describe the divide between the regulated and the competitive, then it connotes a view of regulation that is substantially at variance with the prevailing focus, in today's regulatory practice, on quarantining monopoly from competition. This current view is of course the product of a supply-side perspective. The alternative, demand-side view of regulation would be concerned rather with the stage of development of the network.

The demand-side view

Government is not a watchdog to industry. Instead government and industry are co-facilitators who expect technology and its demand to

[24]Of course if we broaden our horizon, we know that the search for a regulatory divide sweeps a range of sectors, only one among them being telecommunications.

continue to churn. Schumpeter, surely, is the doyen of this view.[25] As networks flourish and later wither, the focus is on the respective matching of subsidy and competition to growth and maturity.

Issues of 'picking winners' do arise. While price is one mechanism available, it is not the only one; and when used in a subsidy role, price does not serve the usual competitive/allocative function. Negotiation, among all the actors, is another vital factor. So is competition among organizations, as an ongoing spur, even when that competition is not on a price basis.

The position of the economically powerful, who might prey upon the less powerful, is the central issue. However, control is exercised not by rendering that power impotent but by harnessing it. To be sure, this is a fantasy in today's American economy. Perhaps a few other capitalist economies are a bit closer to such a conception.[26]

The issue of where, and ultimately how, to regulate is one question to which the demand-side conception might address itself. There is another intimately related problem which liberalization has raised. As networks have moved towards maturity and competitive, cost-based pricing, there is an increasing conflict with pricing for universal service.

With a migration to cost-based pricing, higher prices for those subscribers who cause higher costs come into conflict with an objective, held by some in the policy dialogue, that all subscribers should be maintained on the network. Of course, one of the basic observations of our analysis is that the snowball effect during growth to maturity in the second half has built into it a natural tendency for the network to remain at its universality equilibrium. However if the demand curve in maturity is conventionally downward-sloping with the standard effects that implies, price differentials assessed against higher cost subscribers do countervail against any buoyancy that network universality might otherwise enjoy.

Liberalization and universal service pricing

Put most simply, an approach from the demand side of the question raises the possibility for treating the critical mass group as the unit of purchase in the transaction between providers of networks and users of telecommunications services. The relevance of this to universal service pricing works itself out, fairly obviously, as follows.

First, to begin with the contrasting cases; if the largest critical mass group is small relative to the society – if, in effect, universality for the entire society is not really meaningful or important – then the distinctions drawn in this paper lack force. Group consensus formation does not have significant impact on the economic outcomes since, given only small groups, the network analysis approximates the standard microeconomic analysis.[27] And without universality for the society, there is no conflict between competitive pricing and universality. Pricing can be determined in the usual competitive ways, with differentiation that recognizes variety in communications services and related costs, as demanded by many small groups.

On the other hand, when the critical mass group is large and universality a prospect, there is potential for the conflict of competitive pricing versus universality. If, however, network services are purchased in units that are the size required to accommodate the critical mass group, then the conflict may dissipate. That is to say, when the unit that

[25]Schumpeter emphasized, of course, both value creation and value destruction, though he did not see the role for the demand side that has been portrayed here. See J. Schumpeter, *Business Cycles*, McGraw Hill, New York, 1939, p 73.

[26]On a related tangential note; if, with a more metaphorical view, the diagram in Figure 4 is seen as growth demand emerging into mature demand and if it is imagined on a grand scale – at the level of the society – then liberalization itself is represented by the process of sectors emerging from the shelter of more centralized subsidy/support and into the arena of disposition through competition.

[27]So, small groups approximate the standard competitive situation. That, however, should not be confused with substituting the critical mass group (large *or* small) to be the 'q' in Figure 1, in place of the individual subscriber. To make the critical mass group the basis for counting quantity serves again to illustrate how critical mass formation may fold over into mature growth. If whole groups constitute individual 'q', the groups proceed down the usual demand curve ('adding together', of course, to be larger groups at higher 'q'). On the other side, if activities internal to a group are inspected up close (so that 'q' again becomes a count of *individual* subscribers), the transient, upward-sloping critical mass process emerges to detail any single point on the downward-sloping curve.

is purchased will serve the group, the price can be apportioned in typical universal service fashion and so average across the subscribers.

To make such a proposal is to do no more than recognize that, in demand for the network, there can be a counterpart to natural monopoly in supply. Thus, this recognizes a natural monopsony counterbalanced to a natural monopoly.[28]

On a sober note, this proposal does fail to deliver the promise of real competition during the downward-sloping maturity phase. Indivisibility in demand only revisits all the problems of indivisibility in supply and natural monopoly, certainly including the problem that large-scale 'lumpiness' precludes any real prospect for competition.

Price signalling

The signalling function of price in the new network case can now be considered; first in terms of the rise, and then of the maturity of a network. This echoes a refrain on the 'paradox of value' which traces back at least to Adam Smith.

As discussed earlier, price signalling during startup may just be within the group that comprises potential demand (and, at that, not the most important signal in a consensus formation/fragmentation process). Then, during the maturity phase, price returns to its more usual function of mediating between supply and demand.

But if the balance in maturity is really between two sides which each have natural scale economies, does price signalling in maturity come down to the same marginal result which made it possible for neoclassical microeconomics to resolve what might be thought of as Smith's paradox of value? It would appear, instead, that price in maturity serves to reflect relative balance between the two sides at the bargaining table. And, intriguingly, the role which price plays during startup implies that price then takes on its typical commonsense meaning, namely value to the consumer.

Finding the appropriate organization

Discussed above is the question of whether a planned approach might have some advantage over the competitive marketplace, perhaps with respect to the speed of change. Subsequently, however, the flexibility necessary to support that change has been found to be important, and that seems to argue for the marketplace. In considering whether hierarchy or market is the appropriate organization it becomes clear that for this network case both are appropriate. At least, that is what these competing requirements would suggest. Thus, the mixed economy may be the best, not a second best, solution.[29]

One possible such intermediate case, which appears to have nicely combined elements from both opposites, is the programme in Japan which created universal service in a few decades. Japan's financing of the facilities for universal service perhaps illustrates, at least in one small way, the intermediacy of the path taken by that programme – in other words how that path was 'between the ideal forms'.

While there was significant coordination and funding from Japan's centralized authority, individual subscribers to the network also participated in the financing. Subscribers contributed both through their purchases of bonds and through the payment of installation fees. In fact

[28]And thus, with this new view, we see why the monopolist does not control price (as in the standard analysis). Instead, there is a monopsony/monopoly seesaw. So then we see, further, that it is appropriate to speak of a supply curve, even in this situation.

[29]See *op cit*, Ref 21, for some parallel notions that 'the hybrid case' may work best.

such subscriber financing provided over 40% of the investment required.

Conclusion

The prevailing view of economic activity, now so carefully worked out, has surely provided an elegant understanding of an internally consistent and self-generating resource allocation. Could a process of value creation be a demand counterpart to the resource side, but one that injects an inherent dynamic into the interplay between the two – underwriting a continually changing ebb and flow of technology-with-demand, a repeated undulation in value and in the wherewithal to provide that value?

If, along with our neighbours, the fruits of evolving new technology/new demand are to be understood and enjoyed, then they need to be conceptualized.

Considering again the rising demand curve (which perhaps does not arise from the origin and whose shape is unknown), in looking around one finds even common cases that are not adequately explained by a shift in curves, particularly in relation to new technology, although these are not always instances of a pure network externality.[30]

[30]An everyday, seemingly trivial, example comes from the world of marketing and new product introductions. The exemplar case is the product introduction campaign which gives away free to consumers, for example by mail, some new product such as soap. If the product succeeds and demand grows, so does the price. Or, more parochially, as first one, then another new parking garage opened in the vicinity of my office, it finally dawned on me that the pricing history in both cases involved an initial price at a low level; as demand built, so did the price. Or, again, what is otherwise condemned as 'dumping' might in some cases, be seen as penetration pricing designed to build demand.

If the increase of quantity with price is taken seriously in the case of nascent demand, then the implications have to be ferreted out. As one first step, I am indebted to Marcellus Snow for pointing out to me that the income effect naturally reinforces the workings of a positive-sloped conception for growth in demand. Since value increases with demand, the individual is wealthier as demand grows and so is prepared to consume more.

Appendix

Network demand

In examining Figure 3 three conclusions can be drawn. Throughout the region below the inflection point of the curve it is clear that too few of the population actually responded to reach whatever level of membership has been expected. (This outcome depends, of course, upon the data, ie, the normal distribution, with which we began originally.) So those who do respond find that actual network sizes are inadequate to sustain their own, individual requirement as to critical mass. Any fledgling network that operates at sizes in this region will disappoint those hardy souls who do give it a try, and the effort will slide back to oblivion (lacking, that is, exogenous tweaks).

Above the inflection point, on the other hand, the network size always exceeds expectations, even for prospective subscribers with the next higher required critical mass – and so these potential subscribers also find reason to join. That creates the self-powered snowball which generates growth to universality.

The inflection point itself is, of course, the point of the group's critical mass, which any startup effort must surmount. That point of critical mass is in equilibrium, but is unstable so that any slight swing to one side or the other – the loss of a few subscribers for extraneous reasons, or the gain of a few – will unbalance the situation and the process heads off, pell mell, to the destiny towards which it swung. Zero subscribers and universality are both also points that are in equilibrium; and these equilibria, in contrast with the critical mass, are stable.

The process of change

As suggested in the text, changes in the demand for a new network – whether that means a slide back to the starting point or a climb to universality – might be described in terms of each prospective subscriber 'watching, while being watched'. The process of change, the 'watching', can now be investigated in some detail.

To inspect one cycle of change in the pool of subscribers, a single slice is taken, executed at any point along the history of the new network. (Since the focus in this closeup inspection is on change, the only slices that are not considered here are the three equilibria.) The group holds an expectation about what the network size will be, with some people having selected membership, while others still hold back.

There is a subgroup for whom the shared expectation is particularly salient – those whose personal requirements for critical size are near, or equal to, the present level of expected membership. These individuals must confirm whether their needs have been, or might be, met by the existing state of affairs. They do so by comparing actual network subscribership against the critical size which they personally require. And because personal critical sizes for this subgroup approximate the current group expectation – during this particular cycle of change – their check serves as a check of the group expectation against the actual size.

Changes, individually and for the group, work out as follows: when the network is below critical mass, the key

individuals that comprise the subgroup are those current subscribers who find that their critical sizes – and thus by approximation the group expectations – are not met. Alternatively, above critical mass the key individuals are those prospective subscribers who find that their critical sizes, again along with the group expectation, are exceeded. These individuals then desert, or in the second case join, the network. To reflect this change in fortunes, the group expectation must adjust and it ratchets down or up to accord with this most recent shift.

Then, with a new expectation established, key individuals – who are different from the key people in the round just completed – check once again and the process then recommences.

For instance, if the expectation is 35, but the actual only 30, then present subscribers with critical sizes between 30 and 35 desert the network. The group expectation then adjusts to reflect the new reality of 30. But (at least with the curve from our original, normally distributed individual needs for critical size) after those desertions there will, again, not be enough subscribers to support the revised expectations, so it is further revised downwards.

Change for the network, its ebb and flow, turns on a basic interplay between two factors, one constant and the other in flux. On the one hand, the distribution of different individual needs for critical size within the group is a fixed element. On the other hand, the group's shared expectation about network size continues to change (and, as with any expectation, tends to lag the situation it mirrors by at least a period). Outcomes for the network – the seemingly inexorable climb, or demise – grow from the interaction between the fixed, though different, individual critical sizes and the changing, but shared, group expectation.

The role of communications

The speed of the adjustment process – in other words, the speed of change – depends on whether perceptions of the changing actuals are fine- or coarse-grained. Speed of adjustment also depends on how well 'lubricated' communications are within the group. Those communications underpin the test of expectations, the test that compares actuals against the expectation. The communications also underpin what is the decisive step, namely the shift in expectations that marks the end of one adjustment cycle – here the communications must match perceptions of the *change* in actuals against the current, but soon obsolete, expectation.

This approach emphasizes the importance of communications within the group for the speed of the adjustment process. The central character of the communications, the 'watching, while being watched', means that in a real sense outcomes 'float' on the perceptions of, and as a creation of, the group. Since styles and efficacy of communication vary significantly by group or culture, this floating creation of the group may move quickly, or slowly, to establish new economic value – the effect of the larger shift in expectations. Some groups may evolve new values, where others are not so able.