Current Issues in Telecommunications
Regulation: Pricing

Alfred E. Kahn†
William B. Shew††

The American telecommunications industry is in a turbulent transition, whose outcome is still far from clear. Amidst all the controversy over deregulation, the overwhelming majority of transactions in telecommunications markets still take place at regulated prices. In actuality, what the industry has been experiencing is an uneven and often uncoordinated series of movements in the direction of outright deregulation of some markets and less comprehensive regulation of others.

Both the transition and our ability to predict its ultimate outcome have been especially complicated by the tendency to deregulate entry while continuing to regulate the prices of incumbent suppliers. The retention of price controls was certainly politically inevitable, in view of the wide range of social goals and interest groups served by the existing regulatory regime. It was also probably desirable: it has seemed only prudent to continue to control the prices charged by putative monopolists until markets show themselves capable of effective competition.

The supposedly transitional retention of regulation—involving both price ceilings and floors—may, however, thwart achievement of a rational institutional structure for this industry. This is true whether that arrangement is largely competitive and no longer subject to traditional public-utility style regulation (as we believe likely) or retains a large component of regulated public utility monopoly. The retention of price ceilings, because of the perceived need to prevent the exploitation of residual monopoly power, may prevent the closer alignment of prices and costs that is necessary if competition is to function; and the prescription of minimum prices for the incumbent companies, by artificially protecting their competitors, obstructs the discovery of whether true competition is really viable.

† Robert Julius Thorne Professor of Political Economy, Cornell University; Special Consultant, National Economic Research Associates, Inc.
†† Vice President, National Economic Research Associates, Inc.

We acknowledge with gratitude the helpful criticisms of John Haring, William E. Taylor, Ivan Seidenberg, Roger G. Noll, Charles L. Jackson, Bridger M. Mitchell, and Peter Manus, and the research assistance of Laurent Guerard.

Copyright © 1987 by the Yale Journal on Regulation.
As a result, we simply do not know whether large parts of the business may really be natural monopolies. This is certainly not the case so far as the equipment that subscribers attach to the end of the telephone line is concerned—the phone instruments themselves, interior wiring, answering machines, switchboards (PBXs), and computers. In contrast, it is by no means clear to what extent the competition now prevailing in the long-distance business is merely the result of the artificial pricing and other handicaps still imposed on AT&T and its successor companies. The same uncertainty extends to the competitive bypassing of the local telephone companies and the competitive provision of the equivalent of local service by geographically concentrated business users, such as in the shared tenant services provided in "smart buildings."

So the makers of telecommunications policy confront something of a dilemma: most of them are unwilling to deregulate completely until they are confident that the markets are capable of being truly competitive; but they are unlikely to be able to find out unless they deregulate.

The other major reason for the turbulence of the transition and the uncertainty about its ultimate result is that regulation has historically set prices in part on the basis of political considerations—indeed, some observers have argued that is its raison d'etre—rather than in accordance with the purely economic considerations that would govern in a competitive market. When entry is effectively controlled, it is possible to sustain


3. See, e.g., Posner, TAXATION BY REGULATION, 2 BELL J. ECON. & MGMT. SCI. 22 (1971) (arguing that the main historical reason for regulation has been the desire to practice internal subsidization).
Pricing

uneconomic price structures, setting some prices artificially high and insisting that the resulting profits be used to hold other prices below cost. The most important examples of this practice in telephony have been the overpricing of toll calls in order to keep the price of basic service low, and discrimination against business in favor of residential customers.

As regulatory barriers to entry have been relaxed, however, competitors have, naturally, flocked into the exploited markets, thereby threatening to dry up the profits used to hold down basic residential rates. This has in turn generated political opposition to continued deregulation and a search for alternative ways of preserving that contribution.

We do not attempt in this Article to resolve the ultimate question of the proper institutional structure of telecommunications. Our purpose is, rather, to contribute to a resolution of the dilemmas created by partial deregulation, by suggesting solutions to a number of complex and hotly contested issues of regulatory policy that have been raised or enormously exacerbated by these changes. The solutions we suggest would be consistent with either ultimate outcome—an essentially deregulated, competitive industry or thoroughly regulated monopoly, or some combination of the two. The key is economic efficiency: regulatory decisions consistent with that goal will not distort the process of deregulation or of ascertaining whether a deregulated, competitive regime is indeed feasible. Where efficiency conflicts with other social goals, we suggest ways of achieving the latter with minimum sacrifice of the former.

In Part I, we describe the major distortions in the pricing of telecommunications services that continue to be preserved by regulation, identify the principal issues created by the clash between those distortions and partial deregulation, and expose their common source: the pervasive presence of common products and common costs. In Part II, we dispose of a number of fallacious, pseudo-economic justifications that continue to be offered for the present structure of prices. This is the easy part of the task. Many of the disagreements, however, stem from conflicting notions about

4. For an illuminating historical account and analysis of the ways in which the various steps in the process of opening telecommunications to competitive entry constrained the pricing of the incumbent companies, see G. Brock, Telephone Pricing to Promote Universal Service and Economic Freedom (Federal Communications Commission Office of Plans and Policy Working Paper No. 18, 1986).

5. The most obvious of these revolve around the dismemberment of AT&T, under the Modified Final Judgment (MFJ) entered in 1982 and effective January 1, 1984, and the continued regulation and restrictions on the permissible activities of the successor companies by both the District Court, under the MFJ, and the FCC, aimed at preventing unfair competition and exploitation of captive customers. United States v. American Tel. & Tel. Co., 552 F. Supp. 131 (D.D.C. 1982), aff'd, 460 U.S. 1241 (1983). Ultimately, they involve the question of whether, when, and in what ways deregulation may best be brought to a logical conclusion. For a suggestion that the mixture of the two may be the worst of both possible worlds, see Kahn, The Uneasy Marriage of Regulation and Competition, Telematics, Sept. 1984, at 1-2, 7-17.
the costs of the various services; Part III examines these factual issues, which are more difficult to evaluate because they involve assertions about the design of the telecommunications network, some of which are beyond our competence to resolve. In Parts IV and V, however, we present—and to the extent we can, apply—the economic principles pertinent to the resolution of those and other major current issues of telephone pricing.

I. Regulatory Distortions and the Allocation of Common Costs

At the core of almost all the pricing issues in telecommunications is the fact that the products of this industry are a large and increasing diversity of services issuing from common facilities. While telecommunications is by no means unique among the public utilities in this respect, the problem is much more pervasive and central here than in the others. Whereas a kilowatt-hour of electricity is a kilowatt-hour and a cubic foot of gas a cubic foot (this is by no means to ignore the fact that a kilowatt hour or cubic foot supplied at one time of day on a firm basis is not the same service as at another time on an interruptible basis), telephone service is an array of services, increasingly standardized over time, and therefore less and less susceptible to traditional regulatory treatment. Since regulators and economists generally accept the desirability of basing the prices of these services on their respective costs, the issues tend to be framed in terms of the proper apportionment of their common costs among them, a task further complicated by the fact that some of the services have become subject to competition, while others continue to be offered by a single supplier, under close regulation.

Under the historic process for apportioning the costs of the telephone companies between intrastate and interstate regulatory jurisdictions (known as “separations”), an ever-increasing portion of the cost of merely linking subscribers to the communications network has been recovered in the charges for long distance calls, and used principally to hold down the basic monthly charge to residential customers. The states generally followed a similar practice in pricing toll calls within their borders. Since the cost of connecting subscribers was widely believed to be “non-traffic-sensitive” (NTS) and therefore not part of the marginal cost of calling, its transfer to toll violated the most elementary principle of efficient pricing. Still, by loading on to long-distance callers costs that have reached a


7. We discuss recent contradictions of this factual assumption in Part III.

peak of something like eleven billion dollars annually, the transfer made a major contribution9 to the fifty-five percent decline in real terms in the price of basic service between 1940 and 1980.10 And that in turn helped raise the proportion of households subscribing to telephone service from thirty-seven percent in 1940 to the present level of more than ninety-two percent.11

Most state regulators opposed deregulation, recognizing that if the barriers to entering the overpriced markets were removed, competition would erode the monopolistic margins producing that multi-billion dollar contribution to local rates.12 And as deregulation of entry into long-distance communications has nevertheless proceeded, they and Congress have successfully insisted on perpetuation of the contribution, which now takes the form of a charge by the local telephone companies to all long-distance carriers for access to the local networks, which the latter companies require in order to originate and terminate calls.13

The fears of the state regulators have nevertheless been partially realized: between January 1980 and November 1986, the local telephone charges component of the Consumer Price Index rose 31.6% in real terms, while the price of interstate toll service decreased 28.5%.14


9. The eleven billion dollar figure includes an estimate of the contribution from intrastate as well as interstate toll. We refrain from characterizing this transfer of costs as a "subsidy," even though that clearly has been its purpose, because that term is susceptible to a variety of definitions, and it is not clear that the practice we describe here satisfies all of them. See generally N. Cornell & R. Noll, Local Telephone Prices and the Subsidy Question (Oct. 21, 1985) (unpublished manuscript on file with the authors) (examining validity of general assumption that basic service is subsidized). On the uncertainty about whether the transfer of costs meets one widely used test, see infra note 76.

10. The figures in current dollars are from AT&T Economic Analysis Section, Relative Costs of Telephone Service 1940-1980 (1980).

11. The 1940 figure is from U.S. Department of Commerce, Bureau of the Census, Statistical Abstract of the U.S. 495 (90th ed. 1969). The 1986 figure is from Telecomm. Rep., June 2, 1986, at 6. Because of a change in the question posed to households, we are informed, the two ratios are not strictly comparable, and the comparable 1986 figure would probably be substantially higher than 92%.

12. On issues of communications deregulation, the votes of the Executive Committee of the National Association of Regulatory Utility Commissioners (NARUC), of which one of us was a member, typically ran on the order of 90% to 95% in opposition. On the position of NARUC, see Kahn & Zielinski, New Rate Structures in Communications, Pub. Util. Fort., Mar. 25, 1976, at 23.

13. The telephone companies impose a second charge on those carriers that is intended to recover what are classified in the FCC accounts as the traffic-sensitive costs of providing access. Some portion of these charges, as well, appears to be recovering costs (principally of terminating the dedicated subscriber lines in the central office) that are actually NTS. Our discussion in the text, in any event, relates to the recovery of costs of the local exchange network that are believed not to vary with the amount of calling, and are so treated in the FCC accounts.

14. Most of this divergence took place after January 1, 1984, the effective date of the AT&T breakup. Between January 1980 and December 1983, local charges went up 4.4% and interstate long distance charges dropped 6.5%. Between the latter date and November 1986 the respective charges were up 26.1% and down 23.5% (all figures in real terms). U.S. Department of Labor, Bureau
This recent divergence between local and toll rates is apparently explained less by the direct functioning of competitive market forces—since rates continue to be regulated—than by the Federal Communication Commission's (FCC) recognition (and, in varying degree, by the state commissions') of the inevitable direction of those forces. Alarmed by the growing tendency of large users of long-distance services on the one side and interexchange carriers on the other to bypass the facilities of the local telephone companies, in order to evade those inflated access charges, the Commission determined to reverse the long trend of interstate separations. It therefore proposed gradually to transfer those putatively NTS costs back to the subscriber, in the form of a universal flat monthly “interstate access” or “subscriber line charge” rising by one dollar each year over a five- to seven-year period. Not surprisingly, that decision evoked violent
opposition from consumer advocacy groups, state regulators, and Congress, which forced the FCC to compromise. It responded by freezing the percentage of the NTS costs to be recovered from interstate usage and increasing the subscriber access fee only to one dollar and then two dollars (the latter effective June 1, 1986) for residential service, and up to six dollars per line for multi-line business service.\(^8\)

The job of conforming rate structures to the dictates of competition and economic efficiency is therefore still far from completed.\(^9\) An estimated eleven billion dollars per year of revenue requirements labelled non-traffic-sensitive are still being collected from long-distance calling, by incorporation in the access charges to interexchange carriers and in the intrastate toll charges by the state utility commissions.\(^20\) Long-distance rates therefore are still apparently many times incremental costs,\(^21\) and the

18. The six dollars is a maximum. If the cost allocated to the interstate jurisdiction—which varies among telephone companies—comes to less than six dollars, the "interstate access" charge is set at that lower level.

19. We recognize that the typical injunction on regulators to set rates at "just and reasonable" levels does not constitute an undiluted endorsement of economic efficiency as their only proper goal. See, e.g., E. ZAJAK, FAIRNESS OR EFFICIENCY: AN INTRODUCTION TO PUBLIC UTILITY PRICING 59 (1978) (outlining problem of balancing equity and efficiency with public utility pricing as its focus). We proceed, however, on the premise, which we believe to be unexceptionable, that an overriding goal of economic efficiency is in fact the inescapable implication of deregulation—which consists in essence in substituting competition for regulation as the determiner of price and the other dimensions of economic performance. This does not preclude specific governmental interventions in the interest of other social values, but it does suggest—and we accept—that those interventions should be as undisruptive as possible of competition and of economic efficiency.

20. Moreover, the eleven billion dollar estimate does not include the portion of costs labelled TS that are actually NTS. See supra note 13. The reason for the stubborn refusal of the interstate burden—some two-thirds of the eleven billion dollars—to decline in absolute terms, despite the transfer of approximately 3.8 billion dollars to direct charges on subscribers, is, principally, that the frozen percentage allocation to the interstate jurisdiction (a nationwide average of approximately 28%, scheduled to move to a uniform 25% within eight years) is applied to a total that grows from year to year. Fortunately, since the number of minutes of interstate calling has been growing even more rapidly, the roughly constant aggregate amount has translated since 1984 into a declining markup per minute—roughly from 9.4 to 5.9 cents, effective January 1987, for the origination plus the termination of calls (observe that each call is subject to access charges by both the originating and the terminating local telephone company). This sharp reduction has been made possible partly by an exogenous factor—the decline in the cost of capital. Another important contributor, however, has been the response of demand to falling long-distance rates, and this is exactly what economic efficiency required.

The termination charge has fallen only from 4.71 to 4.33 cents, while the origination charge has fallen from 4.71 to 1.55 cents. TELECOMM. REP., Oct. 6, 1986, at 5, 33; TELECOMM. REP., Jan. 8, 1987, at 2-3. The reason the FCC has decided to concentrate most of the reduction on the latter charge has been its perception that the danger of bypass by large users is greater at the origination end. But a mail order firm, for example, would have a strong incentive to bypass the termination charge, for obvious reasons.

21. See, e.g., L. Perl, Impacts of Local Measured Service in South Central Bell's Service Area in Kentucky 24 (May 21, 1985) [hereinafter L. Perl, LMS in Kentucky] (prepared by National Economic Research Associates for South Central Bell Telephone Company) (unpublished manuscript on file with the authors). This study, which is limited to Kentucky, suggests that intrastate rates are more than four times and interstate rates three times their respective incremental costs. Southwest Bell estimated its average incremental costs for intrastate inter-LATA calls in Oklahoma at 5.9 cents per minute in 1984; the corresponding average rate was 34.9 cents—a ratio of almost six to one. See also Development of Intrastate Access Charges: Hearings on Cause 28309 Before the Corporation

197
same is true of local calls where they are measured and charged for separately. In a compromise, a joint board of state regulators and FCC members recommended resuming the FCC's intended transfer of the burden of NTS costs, by further increasing the flat "interstate access" charge on residential subscribers in three steps from its present two dollars to three and a half dollars monthly by April 1989—promptly eliciting indignant protests from consumerists and legislators. And the FCC on July 2, 1986 announced yet another new comprehensive review of the entire access charge question.

Manifestly, while the relaxation of entry restrictions has limited the range of regulators' discretion over prices, it has not eliminated it. The consequence has been to hamper the transition to complete deregulation, and to deprive us so far of a fair test of whether a deregulated telecommunications industry will work.

The consequence has also been to leave most of the central issues of telephone pricing still very hotly contested. All have at their core the question of the proper recovery of costs common to the provision of a variety of telephone services:

- The contention that (a) providing subscribers with mere access to the telephone network, (b) local calling, and (c) long distance calling are joint products, and therefore do not have separate, objectively identifiable costs;

- The question of whether subscriber access—merely putting subscribers in a position to place and receive calls—is a separate service, which should therefore be charged for separately;

- The contention that a subscriber access network suitable for long distance calling is more expensive to construct than one designed for local calling only;

- The question of whether customer access becomes more costly to provide when it is used heavily—whether for local or long-distance calling;

Comm'n of the State of Oklahoma (June 1, 1985) (statement of Alfred E. Kahn at 7-8) [hereinafter Kahn Oklahoma Testimony]. A New England Telephone Company study estimated the marginal cost of intra-LATA toll calls at 2.0 cents per minute on-peak; the charges averaged 22.6 cents per minute. L. Perl, Can Marginal Costs be Measured 12 (1987) (unpublished manuscript on file with the authors).

22. See infra note 140.

23. N.Y. Times, Mar. 13, 1987, at 1, col 2. Of course, imposition of this charge by the Federal regulators does not ensure more efficient pricing: the state regulators are free to offset it by reducing the portion of the basic charge under their control correspondingly and making up the deficiency by increasing other rates—for example, intrastate toll or for local calls. It appears, however, that in general they have not done so.

Pricing

- The claims that the telephone companies are inefficiently designing their systems in a way that inflates the marginal costs of subscriber access;

- The controversies over whether the telephone companies are overinvesting in modern technology—notably digital switching and fiber-optic transmission—in order to be able to provide newer and more versatile services that are of no interest to people who want only plain old telephone service (POTS), and are trying illegitimately to impose some portion of the costs on them;

- The question of whether subscriber access and local calling should be priced separately or offered only bundled, in the form of flat rate local service;

- The fairness of burdening captive POTS customers today with depreciation charges inflated by the increasingly intense competition to which the telephone companies have been subjected in the provision of other services, and with the other carrying costs of sunk investments left stranded by that competition.

Some of these issues involve empirical questions to which we are not able to offer definitive answers. We lack sufficient information to judge, for example, whether the companies’ investment decisions are or have been either prudent, in the technical regulatory sense, or efficient. Nor do we attempt to judge the comparative equities of having the heavy sunk costs associated with past investments borne by stockholders on the one side or captive customers on the other. In these cases all we can do is elucidate the pertinent economic and regulatory principles.

Finally, while we do not attempt in this Article to provide solutions to the institutional issues of telecommunications policy, it seems to us that current efforts to deregulate some commonly provided services totally while continuing to regulate others encounter problems that have no satisfactory solution within the framework of conventional regulation. As we will show, given the pervasiveness and centrality of common costs, this effort tempts regulatory commissions to engage in economically illegitimate cost allocations and second-guessing of company investment decisions in ways that are fundamentally incompatible with both competition and economic efficiency. So long as regulation of POTS for households and small businesses continues to be necessary, the only logical solution, ultimately, is total deregulation of the other services and total separation of their revenues and the costs assigned to them from the rates that continue to require regulatory attention.
II. Six Pricing Fallacies

It is a sign of the effectiveness of a market economy that if there is a demand that is capable of being satisfied on acceptable terms, the supply will be forthcoming. So it has been with the avid demand for purportedly economic rationalizations of the present structure of telephone company prices; they have appeared in a profusion, with a versatility and resourcefulness, that cannot fail to inspire admiration. Typically they involve some combination of the following assertions:

1. The access component of basic service is not itself a service, but simply a means of enabling customers to make local and long-distance calls;

2. In a competitive market, customers are never charged separately for "access"—that is, the mere opportunity to buy particular goods and services;

3. Identifying the separate cost of access is impossible, because access and telephone calls are joint products;

4. Since long-distance callers and carriers benefit from (and depend upon) the lines linking customers to their telephone central offices, they should pay part of those access costs;

5. Failure to charge these beneficiaries would give them a "free ride" on the facilities, and competitive markets do not give free rides;

6. It is unimportant how the costs of access are recovered because telephone pricing is a "zero sum game."

All of these propositions are economically fallacious. Yet they must have a superficial plausibility: they are presented in rate case after rate case, with a regularity that would be hard to explain if they always fell on deaf ears. For this reason only, each requires careful scrutiny.28

A. The First Fallacy: Customer Access is Not a Service

Customers have to be connected with the telephone network if they are to place or receive telephone calls. This is usually accomplished by a physical link between their premises and a switch at a nearby central

25. We make no effort to provide a systematic attribution of these views to particular individuals: it is the ideas themselves whose fallaciousness we seek to expose. We have encountered some or all of them, from various witnesses, in every state regulatory proceeding involving telephone pricing with which we have come into contact.

26. In view of the adversarial, politically charged character of the regulatory process, we disclose that we have offered testimony on behalf of both Bell Operating Companies and AT&T along some of the lines of argument that follow.
Pricing

office, which has the capability of routing calls to their intended destinations. The connection is commonly referred to as access. 27

The argument we examine here is that (a) access is merely a prerequisite to placing and receiving calls, (b) the cost of providing it is therefore properly regarded as part of the cost of telephone calls, and (c) that cost should therefore be recovered partly in the charge for local calls, partly in the price of toll calls.

The assertion that customer access is not a service, but instead merely a prerequisite to the provision of “real” telephone services, is in itself merely semantic. But the inferences drawn from it are fallacious. The defining characteristic of a service is that it is or would be demanded in its own right. By that criterion, access is clearly a service. Even if most customers were not interested in it in order to place calls, many would want it if only to receive calls. And even if they had no specific expectation of placing or receiving calls, many would still be willing to pay for the opportunity to do so, when and if the occasion arose.

The important question is in any event not semantic but economic: whether or not access is labelled a “service,” what is the efficient way of recovering its costs? Specifically, is it desirable to recover the costs of providing subscribers with the opportunity to place or receive calls, whether or not they actually do so, in a flat monthly charge, or in charges for usage? That is the only relevant economic question.

And that question breaks down into two others. First, does subscriber access have a separate identifiable incremental cost associated causally with providing it? The answer is, unquestionably, yes. Connecting a customer to the network uses scarce resources, even if he or she never uses the connection. The customer who subscribes to two access lines imposes a greater cost on the system than the customer who subscribes to one, even if they make the same number of calls, at the same times and places.

Second, does charging for access separately serve a purpose? The answer is that it serves the very important purpose of economic efficiency if buyers are confronted, in each of their purchase decisions, with prices that reflect the respective incremental costs to society of their taking more or less of each available good and service or, to put it another way, what costs society would save if they took less of each.

27. It is also referred to as customer or subscriber access, to distinguish it from “carrier access.” Although, so far as the link between subscriber and central office is concerned, the two are different ways of looking at the same thing—the same network that gives subscribers access, via the switch at their local telephone company office, to long-distance carriers obviously gives the carriers access to them—the distinction becomes significant when it comes to the hotly contested issue of how the telephone company should recover the costs of that network: in “access” charges to the carriers or in the basic monthly charge to subscribers.
It would serve no purpose to charge for access and usage separately if subscribers always purchased them in fixed proportions: it would not make any difference in those circumstances whether the combined incremental costs were distributed between the two services in separate prices or incorporated in a single price. The fact is, however, that subscribers use their access facilities in widely varying amounts and patterns; in those circumstances economic efficiency demands that they be confronted with prices for those several services corresponding as closely as possible to their respective marginal costs.

Using the price of telephone calls to recover access costs that do not in fact vary as more or fewer calls are made therefore induces wasteful choices by customers. It encourages them to order underpriced access lines that they value less than the incremental costs to society of providing the lines, and it discourages them from making overpriced calls whose value to them would have exceeded the incremental cost to society. The same result would follow if an electric utility were to supply its customers with all the appliances they wanted at no charge and recovered the costs in the price of electricity—wasteful overpurchasing of appliances and underconsumption of electricity.

B. The Second Fallacy: Access Would Be Free in a Competitive Market

The contention that access is not a service is often followed by an assertion that competitive markets never charge for the mere opportunity to buy a good or service. "Only a monopoly," one witness has asserted, "can charge a flat fee on a continuing basis whether you use the service or not."
These assertions are plainly wrong.

Country clubs charge a flat membership fee as well as separate fees for the use of facilities to which membership provides access—golf courses, tennis courts, and restaurants. Restaurants frequently impose minimum or cover charges, independently of the prices for items ordered from the menu. Banks and credit card companies impose flat monthly or annual service charges, along with separate fees based on how much a customer uses their facilities. And what are long-term leases of residential or commercial space but contracts under which the lessees agree to pay for access to those facilities fixed rentals that have no relationship to the extent to which they actually use the premises?

On the other hand, most restaurants in this country do not impose cover charges. Department stores and supermarkets do not charge admission fees. Most suburban shopping malls do not charge separately for parking; but parking lots in cities do.

What distinguishes these various cases? The principal factors seem to be whether the supplier incurs a cost in providing "mere" access; the size of those costs relative to the costs and potential revenues associated with actual use of the facilities; and the size of the lost net revenues resulting from reduced trade if the supplier charges separately for access—by keeping customers away entirely—relative to the cost of not doing so, which depends in turn on the extent to which customers might take advantage of a zero access charge to impose losses on the supplier.

To offer shoppers the opportunity to buy its wares, a department store must incur the costs of providing space, heat, light, display facilities, and at least a minimum sales force. It incurs most of these costs regardless of whether the shoppers buy a great deal or nothing at all. Similarly, a country club must construct and maintain golf courses and tennis courts to be able to offer its members the opportunity to play those games; it incurs these costs even if the facilities are rarely or never used. For most department stores, supermarkets and restaurants, however, the costs of providing mere entry are small compared with the expected revenues from actually making sales or providing meals. Their profit will come from a heavy turnover, the variable costs of which will typically be many times the costs of providing the facility in which the sales are made. In these circumstances, imposition of a cover charge or entry fee would, by driving away

1985) (statement of Allen G. Buckalew at 33). Economists have made similar arguments in opposition to including construction work in progress in a utility company's rate base: "You don't see consumers in competitive markets being forced to pay for the costs of facilities that do not yet serve them." For an exposure of the fallacy of this assertion and other related regulatory shibboleths, see In re Illinois Power Co., Illinois Commerce Comm'n, No. 84-0480 (Apr. 25, 1985) (rebuttal testimony of Alfred Kahn).
customers, probably lose them more profits than it would gain them. In the case of country club golf courses, in contrast, the costs are preponderantly independent of usage or turnover. The same is obviously true of providing apartments and offices for rental.

Exactly the same calculations that induce most restaurants not to impose cover charges or minimum bills will often lead richly appointed restaurants, and particularly ones that offer entertainment, to do so. If they did not, they might well be filled up with patrons who order only a cup of coffee and remain at their tables for hours. The costs of providing mere access in the last case are heavily inflated by opportunity cost: since the space is limited, admitting freeloaders without charge results in large losses of net revenues from the customers that they crowd out.

In suburban shopping malls, the costs of providing parking spaces are quite low compared with the sales revenues of their various shops; and, since customers will rarely use the free parking for purposes other than making purchases in the shops, the opportunity costs of non-shoppers crowding out shoppers are low as well. In these circumstances charges for parking would, by keeping customers away, probably result in a greater loss of net revenues by the shops than it would save in the direct and opportunity costs of providing parking. In cities, in contrast, space is scarce and costly, and drivers quickly discover that they can use a bank's downtown parking lot while shopping elsewhere. So the bank, to avoid crowding out of its own customers, must develop a system for charging non-patrons, and excusing patrons of access charges only for a limited time.

Competition forces companies to price the various services they offer at their respective costs. If, then, providing "mere" access involves substantial costs—direct or opportunity costs—firms in truly competitive markets must either charge for it or go out of business. Thus, the assertion that competitive markets would never charge for access is not merely wrong; it is in some ways the opposite of the truth. The greater a firm's monopoly power—its ability to charge above marginal cost for usage—the greater its ability to offer access at less than cost, in the confident expectation of recovering the deficiency through usage charges. 32

32. Just as in the practice of tie-in sales, a firm might in this way more effectively exploit its monopoly power, discriminatorily obtaining larger monopoly profits from customers to whom its services were worth more (as measured by their usage) and smaller ones from light users, to whom they were presumably worth less. See Bowman, Tying Arrangements and the Leverage Problem, 67 YALE L.J. 19 (1957); Burstein, The Economics of Tie-In Sales, 42 REV. ECON. & STAT. 68 (1960).
C. The Third Fallacy: Access and Calls are Joint Products

The joint products argument proceeds generally along the following lines. Since calls cannot be made without access, and/or since the same subscriber plant provides access for both local and long-distance calling, the three services are joint products. It is an elementary economic principle that joint products—wool and mutton being the standard textbook example—do not have separately identifiable costs. It is therefore simply not possible in this case to follow the unexceptionable rule, which few economist witnesses have been so shameless as to reject, of pricing services at their respective costs. The regulator has no choice, therefore, but to allocate the joint costs among the services in some "reasonable" way, such as (1) one-third each to access, local calls, and toll calls, (2) in proportion to the minutes of use of the network for local and toll calls, (3) on the basis of the relative values of local and toll calls, or (4) in proportion to the costs of building hypothetical stand-alone systems for local and long-distance calling.

This argument, either disingenuous or ignorant, is fallacious from start to finish, for the simple reason that these telephone services are not joint products. The foregoing chain of reasoning, which depends entirely upon that erroneous premise, is therefore irrelevant.

Joint products, as economists use that term, can be produced only in fixed or unvarying proportions. The traditional examples—wool and mutton, cotton and cottonseed oil—may not be completely realistic (sheep may be kept alive through a varying number of wool-bearing seasons before they are slaughtered; and the proportions of wool and mutton can probably be varied by breeding). But the concept is clear: products that can be produced only in fixed proportions do not have separately identifiable production costs. The cost of bringing a sheep to market is ascertainable, but how much of that cost is attributable to the wool and how much to the mutton cannot be said. It is impossible to identify a marginal production cost of wool and another for mutton: the only one we can identify is for wool and mutton.33

33. This is not to say that joint products do not have unique economically efficient prices that can be separately specified. On the contrary, the economically efficient solution to pricing them, as Alfred Marshall demonstrated clearly a century ago, involves determining the efficient level of supply by equating the combined prices purchasers will be willing to pay for them to their joint marginal supply cost and selling those quantities for what each will yield in its market. This has the effect of equating the price of each to its separate marginal opportunity cost—that is to say, to the value of the last or least valuable unit to the purchaser next in line excluded by that price. A. MARSHALL, PRINCIPLES OF ECONOMICS 389-90 (9th ed. 1961); A. MARSHALL, INDUSTRY AND TRADE 190-94 (1932). See also Hirshleifer, Peak Loads and Efficient Pricing: Comment, 72 Q. J. ECON. 451 (1958); 1 A. KAHN, supra note 6, at 77-83. The efficient prices of joint products, thus, are determined by their respective demands: they can not and need not be based on separate costs of production.
Telephone access and local and long-distance calls are not joint products, because they are patently not produced in fixed proportions. Subscribers connected to the telephone network may, both individually and in the aggregate, make no calls or many; they may place local and long-distance calls in all conceivable volumes and proportions to one another; and they do. Each service has a separate, identifiable marginal supply cost, which can be determined by examining how the telephone company’s costs increase or decrease as, holding the other services constant, customers order more or fewer access lines or use their access to place more or fewer local or long-distance calls.

D. The Fourth Fallacy: Costs Should be Distributed on the Basis of Benefit

Each of the preceding three fallacious contentions typically leads to a conclusion that the cost of subscriber access should be recovered at least in part through the prices of local and long-distance calls because the callers use those facilities and benefit from their availability. Failure to do so would be objectionable, it is asserted, because that would give the caller a "free ride." Another version of this contention is that the companies providing long-distance service profit from the existence of a customer access network, and therefore ought to pay for it. This version is economically illiterate, because it assumes that toll carriers will not charge customers for the costs they incur; we return therefore to the preceding assertion about the fair distribution of access costs among customers.

Economists are not, as such, experts on what is fair and what is not. A reasonable case could surely be made, however, that there is nothing particularly fair about charging users of a facility more than the costs that their use of it imposes on society, or that restraint on their part would

34. Some economists use the term "joint costs" more loosely, to embrace cases in which the products are produced in common but the proportions can be varied. Whatever the possible confusion in terminology, the significant economic question is whether the products or services in question are or are not produced in fixed proportions.

35. See, e.g., Melody Testimony, supra note 30, at 13. The placers of long-distance calls would of course still be paying the costs of providing them with access even if those costs were merely recovered in their charges for local service alone; but, the contention is, they would not be paying their "fair share"—one reflecting their assertedly disproportionately great benefit—to the extent they made more use of that access for long-distance calling than the average subscriber.

The same free rider argument was frequently offered, years ago, by opponents of peak pricing of utility services: off-peak users, they argued, should be made to bear their "fair share" of the costs of the facilities that served them. Regulatory commissions today, however, widely recognize the validity of the opposing consideration—that it is not only inefficient but could be regarded as at least equally unfair to charge some people more for service and others less than the incremental costs they impose on the rest of us by taking greater service, respectively, off- and on-peak.
save it. By that reasoning, it is unfair—as well as blatantly inefficient—to charge long-distance or local callers much more than the marginal costs of such usage.

Least defensible of all has been the effort of some economists, conceding the economic desirability of basing prices on "cost" (though, for obvious reasons, rarely adding the adjective "marginal"), to justify the allocation of a large portion of the putatively non-traffic-sensitive costs of subscriber plant to long-distance calling—whether on the basis of benefit, value, or usage.

Let us speak plainly about this kind of reasoning. The only costs that have objective reality are ones that describe a causal relationship between the act of purchase and their incurrence. Cost allocations that are not grounded in causality have no basis in objective reality; they have no meaning independent of the prices they are supposed to justify, except in some ritualistic, incantational sense. Allocations of cost on the basis of benefit or some other conception of fairness are tautological, or teleological; they are merely a plausible device for clothing with the appearance of cost justification some preconceived notion of what the proper price should be, rather than meaningfully independent tests of the economic propriety of those prices.

Witnesses who make this sort of argument are not saying: "Usage causes the company to incur some of these costs; therefore the prices charged for usage should incorporate those costs," but, instead, "usage should bear some of these costs; therefore let us allocate some of them to it." The last part of that sentence adds nothing logically to the first; it is merely a superficially plausible way of attempting to justify the predetermined result.

E. The Fifth Fallacy: Competitive Markets Do Not Provide Free Rides

We are aware of no economist witnesses in telecommunications regulatory proceedings who deny the desirability of regulation emulating the results that would in unregulated markets be produced by effective competition. On the contrary, they typically adduce that principle to support their proposals, arguing, as we have already observed, that one never sees, for example, competitive markets charging consumers for "mere access" to facilities or services. Their confident assertion that competitive markets do not give free rides is equally incorrect.

Butchers still give customers bones for their dogs—"free rides" on the carcass; woodworking shops will give away wood scraps and sawdust, and liquor stores their empty cartons—in all cases because (or to the extent
that) the marginal opportunity costs are zero.\textsuperscript{36} Since deregulation, competition has forced the airlines to give off-peak travelers, in seats that would otherwise go out empty, something close to the "free ride" that some economist witnesses decry in telephone cases. The practice has unquestionably contributed to improved efficiency in the use of aircraft and to enhanced economic welfare.\textsuperscript{37}

Toll service is not like bones for one's dog or sawdust for the cat; it does have a marginal cost. But use of a subscriber's access to the local exchange for the purpose of placing or receiving a toll call is, like those others, a free good, so long as its costs are truly non-traffic-sensitive: its marginal cost is zero.

Competition forces prices to marginal costs; to the extent consumers would have been willing to pay more, the difference is a "consumer surplus" that competition has conferred on them. Only monopolists, governments, kidnappers, and blackmailers can and do extract that surplus by regularly charging their "customers" on the basis of benefit or value rather than marginal costs.

The one kind of cost that we can be certain competitive markets will ignore is the kind of allocation that attributes to services costs that do not vary with their supply. Any attempt in a competitive market to hold prices above the costs causally and directly related to supplying particular services will be frustrated by competitive entry. That is exactly what has been happening in long-distance communications, to the extent competition has been permitted to work its way.

F. The Sixth Fallacy: Telephone Pricing is a Zero-Sum Game

Some commenters have asserted, apparently in all seriousness, that telephone pricing is in any event a "zero-sum game" because whatever we take off customers' long distance bills, we simply add to their local charges, in the aggregate. The central conception of economics, on the contrary, is that moving in the direction of efficient pricing is far from a zero-sum game. What these observers fail to grasp is that prices below marginal costs cause a loss in social welfare just as much as prices above

\textsuperscript{36} These are all joint products, and the zero price in these specific instances is precisely the efficient solution described supra at note 33. To the extent the examples are not precisely apposite (many butchers give "free" bones only to regular customers) it is because these may not actually be free goods in the technical sense—that is, they may not be in such abundant supply as to have zero opportunity costs.

\textsuperscript{37} E. BAILEY, D. GRAHAM & D. KAPLAN, DEREGULATING THE AIRLINES ch. 5 (1985). The airline fares differ from the others in that, except for true standby services, they are typically offered only discriminatorily, and thus conduce to "second-best" results rather than "first-best." For a discussion of first- and second-best, see infra Parts IV and V.
Pricing

marginal costs. These two inefficiencies do not offset one another; they are additive.

Various economists have tried to estimate the welfare losses imposed on all of us, collectively, by the inefficiencies flowing from our present set of prices for telephone services. While those estimates are subject to a wide band of error, they run, plausibly, in the billions of dollars annually.\(^8\) The logic is, as we have already observed, that holding the prices of some services far above the incremental cost to society of providing them induces consumers not to make some purchases whose worth to them would have been greater than their cost to society—that is, than it would have cost *them*, since they are *society*. The difference between those two quantities represents a net loss in social welfare. Similarly, holding the prices of other services markedly below marginal cost induces consumers to purchase incremental quantities that are worth less to them than their incremental cost to society; that difference represents an additional social welfare loss.

The loss is real. Eliminating those inefficiencies would improve total welfare in exactly the same way as giving people billions of dollars more income. The size of these losses depends on (a) the size of the various gaps between marginal cost and price, and (b) how responsive the various demands are to price at those levels. For both these reasons, the principal component of the estimated multi-billion dollar total welfare loss from telephone pricing is the benefit lost by systematic overcharging for long

---

\(^8\) See, e.g., Griffin, *The Welfare Implications of Externalities and Price Elasticities for Telecommunications Pricing*, 64 Rev. Econ. & Stat. 59-66 (1982), estimating a welfare loss of about $1.5 billion annually in 1975 dollars—$3.0 billion in 1985 dollars, CPI-adjusted; L. Perl, *Welfare Gains From Cost-Based Telephone Pricing* (June 19, 1986) (unpublished manuscript on file with the authors) (estimating improvement in net social welfare consequent on moving to marginal cost pricing of $5.5 billion in 1984 dollars—$5.7 billion in 1985 dollars, CPI-adjusted). When Perl introduced the (second-best) optimum markups above marginal cost necessary to cover aggregate revenue requirements (see infra Part V), this reduced the benefit to $5.1 billion in 1984 dollars—$5.3 billion in 1985 dollars, CPI-adjusted—still hardly a zero-sum game.

For an estimate showing very large welfare losses resulting from the recovery of NTS costs from long-distance calling, see C. Jackson & J. Rohlfs, *supra* note 16, at 7-11.

Some of the reasons for the differences between the Griffin and Perl results, even after adjusting the former (as we have) for the severe inflation that occurred after 1975, are: (1) Perl took into account the cross-elasticity of demand between access and usage, while Griffin assumed it away; (2) Griffin estimated welfare losses only for toll service, while Perl estimated welfare losses from local usage as well; and (3) Perl’s elasticity of demand for toll service is slightly higher than Griffin’s, while Griffin’s elasticity of demand for local access was higher than Perl’s.

All these estimates apparently fail to take into account the additional loss in welfare that occurs when the inefficiently high price—in this case, of long-distance calling—becomes part of another’s cost of production. The transfer of inputs at a price higher than marginal cost from one stage of production to another results, if there is any market power at that second or any subsequent stage, in a cumulation of interstitial markups that compounds the ultimate inefficiency. See, e.g., 1 A. Kahn, *supra* note 6, at 145 n.44; 2 A. Kahn, *The Economics of Regulation* 258-59 (1970) (treating the pricing of inputs as exception to adequacy of price proportionality to marginal cost as means of achieving ideal output).
distance calling: the gap between price and marginal cost is the most egregious, and this is the major telephone service with the most elastic demand. In contrast, the amount of inefficient consumption encouraged by the apparent undercharging of residential subscribers for access is relatively slight, because most people think of telephone service as a necessity and the demand for subscriership is therefore relatively inelastic.

III. Some Empirical Issues

We turn now to a series of interrelated factual issues, contested with increasing intensity in recent years, revolving around the question of whether the costs of the subscriber access network, as determined under traditional regulatory and telephone company accounting methods, accurately reflect the marginal costs of access. Witnesses in regulatory proceedings have contended that they do not, and that therefore a large portion should continue to be recovered from long-distance and perhaps other services.

In large measure, these arguments involve questions about the design of the network and telephone company accounting that we lack the compe-

39. See supra note 21.
40. The following estimates of price elasticity of demand for telephone services are arithmetic averages of elasticities reported in L. Taylor, Telecommunications Demand: A Survey and Critique chs. 3 & 4 (1980). Although additional studies of demand elasticity have been performed since Taylor's survey, our impression is that they are generally consistent with the pattern he reports.

<table>
<thead>
<tr>
<th>Service</th>
<th>Residential Sector</th>
<th>Business Sector</th>
<th>Residential/Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential connection</td>
<td>-0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Installation charge)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential access</td>
<td>-0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local use</td>
<td>-0.17</td>
<td>-0.11</td>
<td>-0.22</td>
</tr>
<tr>
<td>Toll Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrastate</td>
<td>-0.35</td>
<td>-0.49</td>
<td>-0.67</td>
</tr>
<tr>
<td>Interstate</td>
<td>-0.58</td>
<td>-0.54</td>
<td>-0.74</td>
</tr>
<tr>
<td>Other Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All vertical (i.e., enhanced) services</td>
<td>-0.99</td>
<td>N.A.</td>
<td>-0.15</td>
</tr>
<tr>
<td>Extensions</td>
<td>-0.21</td>
<td>-0.41</td>
<td>N.A.</td>
</tr>
<tr>
<td>International calls</td>
<td>N.A.</td>
<td>N.A.</td>
<td>-0.91</td>
</tr>
</tbody>
</table>

Only U.S. demand elasticities are reported; for international calls, this means calls originating in the U.S. The aggregate residential/business elasticities bear no necessary relationship to the corresponding elasticities reported for the separated customer classes, having been obtained from different studies.

41. See supra notes 11 and 40.
Pricing
tence to assess definitively; we attempt, however, to present the available
evidence bearing on them. In part, they raise critical questions of eco-
nomic principle, which we attempt to resolve in Parts IV and V.

We consider four factual contentions in defense of continuing to recover
a large portion of subscriber access costs from usage.

The first two are contentions that the design of the local subscriber
network depends on expected calling behavior: first, that there are costs of
designing a network suitable for long-distance calling—notably the cost of
improving the quality of the signals—that would not be incurred in de-
signing a system for local calling alone;\textsuperscript{42} and second, that higher levels of
usage require more and longer subscriber loops, either in order to mini-
mize the combined cost of access and calls or in order to keep the
probability of getting a dial tone from falling to an unacceptable level.
Those incremental access system costs are therefore properly regarded as
part of the marginal cost of calling.

Third, the very large expenditures telephone companies have been
making in recent years to upgrade subscriber access networks and central
offices—installing fiber-optic transmission and digital electronic switching
capabilities—are required not for plain old telephone service (POTS) but,
instead, for specialized and sophisticated services such as Custom Calling
(call waiting, automatic call forwarding, teleconferencing) and high-speed
data transmission.\textsuperscript{43} These costs are therefore part of the marginal cost of
the enhanced services, and should not be recovered in the flat charge for
basic service.

Fourth, the marginal cost of subscriber access, and therefore the econ-
omically efficient charge for it, is markedly below the average revenue
requirements associated with providing it, as they have been traditionally
determined under regulation.

As we proceed to examine these assertions in this Part, we will con-
clude that some of them seem to be factually valid, and that others are
apparently not. In Part IV, however, we demonstrate that most of the
concepts of cost that they invoke are irrelevant to "first-best" efficient
pricing—that is to say, to pricing communications services at their mar-
ginal costs. One reason is that some of these arguments hold usage respon-
sible for the kind of access systems in place today and for the sunk costs
associated with them; sunk costs, however, are irrelevant to efficient pric-
ing. Another reason is that in situations where a plant provides a number
of goods or services in common, economic efficiency requires that it be so
designed as to provide them collectively at minimum aggregate cost. The

\textsuperscript{42} See, e.g., Melody Testimony, supra note 30, at 15-16, 42.
\textsuperscript{43} See, e.g., id. at 32-43.

211
fact that such a design may entail a higher marginal cost and therefore a higher first-best price for one of those services than some other system design is irrelevant.

Finally, in Part V we turn to the likelihood that the pricing of telephone services will have to be "second-best" because prices set consistently at marginal costs would evidently produce inadequate total revenues. "Second-best" efficiency in these circumstances, we conclude, requires that company-wide revenue deficiencies be made up primarily in the flat charge for access, not for usage. The same is true of any revenue deficiencies created by a decision to subsidize some customers in order to keep them from dropping service.

A. Designing the Access System for Long-Distance Calling

Telephone company engineers generally deny that the design of the customer access system is in any way affected by the need to improve the quality of signals for long-distance calling, for two reasons. First, when amplification is needed, it makes more sense to install the amplification devices along the interoffice trunks, where the cost of doing so is classified as traffic-sensitive and attributed to interexchange service, rather than in the dedicated wires between the exchange office and individual subscribers.44

Second, of the calling designated as "local," because it is covered by the charge for local service, a very large portion—69% in urban, 46% in suburban, and 34% in rural areas45—is in fact technologically indistinguishable from "long-distance" service: both involve transmission between separate telephone offices or exchanges. Anyone old enough to remember when local telephone numbers were changed from two letters plus five digits to seven digits is likely to recall that the letters referred to the names of various exchange offices within the local calling area. Therefore, even if the need to accommodate calling between exchanges did require a more expensive subscriber access network than calling within an exchange, that would not justify singling out what is now defined as toll service for its present grossly disproportionate contribution to covering the costs of that network.46

44. This assertion would seem to be confirmed by the opinion of an independent authority that the access system we have inherited was in fact designed for the needs of local service, as a result of which costs were imposed on the long-haul network (for example, of echo suppressors) that would not have been required had it been optimally designed from the outset for both services together. Letter from Charles L. Jackson, Principal of Shooshan & Jackson Inc., Washington, D.C., to Alfred Kahn (Oct. 25, 1984).

45. TECHNICAL STAFF & TECHNICAL PUBLICATION DEPARTMENT, AT&T BELL LABORATORIES, ENGINEERING AND OPERATIONS IN THE BELL SYSTEM 125 (2d ed. 1983).

46. The same observation applies to the fact that a telephone exchange designed for calls among
Pricing

We see no need to attempt to resolve the factual issue of the assertedly higher cost of an access network suitable for long-distance calling because, as we will explain more fully in Part IV, it is almost certainly economically irrelevant.

B. The Usage-Sensitivity of Subscriber Access Costs

The contention that higher levels of usage increase the cost of the access network clearly goes to the heart of the empirical controversies, because it suggests that part of the revenue requirements of the access network are really marginal costs of usage. It has several major components:

1. The historic shift from multi-party to nearly universal single-party service, requiring a dedicated pair of wires to each subscriber, was the consequence, at least in part, of increased average levels of calling, which made multi-party service decreasingly attractive. And this in turn has made multi-party service effectively unavailable in many areas: telephone company engineers tell us that with so few subscribers still taking that service, it would ordinarily cost more to assemble the necessary group, link them together and maintain their lines (for example, to check the source of complaints about service) than to give each customer a dedicated line.

Entirely apart from its factual validity, this argument, too, is probably irrelevant to the efficient pricing of telephone services. The marginal costs of linking subscribers to the exchange are what they are today, not what they would have been had history taken a different course. This does not mean we must approve the particular course history took. So long as

ten thousand subscribers or fewer can make do with switches capable of handling only four-digit numbers, whereas exchanges suitable for handling calls involving a larger number of subscribers, whether within a single exchange or between exchanges, require switches capable of handling seven or ten digit numbers. This means that the marginal traffic-sensitive costs of calling do vary slightly for this reason, depending upon their character. It has no bearing, however, on the contention that the non-traffic-sensitive costs of subscriber plant will be higher in a system designed for local than for long-distance calling.

47. But see supra note 44.


49. The average number of calls placed daily per line increased by 50% during the last 30 years or so, from 6.01 in 1950 to 9.10 in 1981. Bell System, Statistical Manual, 1950-81, at 807 (1982). (The ratios exclude from the denominator residence extension phones, since it is use per line rather than per phone that determines the danger of blockage or interference on multi-party systems.) It seems doubtful that this increase in traffic could have played as significant a role in decreasing the demand for multi-party service as rising personal incomes and the increasingly widespread desire for the greater convenience of single-party service. The Rural Electrification Administration has long had the policy of making its low-cost loans to independent telephone companies contingent upon their shifting over entirely to single-party service.

50. For our reservations in this particular case, see infra note 101.
that course is unlikely to be reversed, however, marginal costs must be calculated and efficient prices set on the basis of the present and future consequences of buyers taking more or less of the services now available, not of services no longer available, as we will explain more fully in Part IV.

2. The less people actually use their phones, the greater the opportunity the telephone companies have to install concentrators, with a dedicated pair of wires extending from each subscriber not to the central office but only to a concentrator, from which—just as in the case of a switchboard—only a smaller number of lines or trunks would continue to the central office. Conversely, the more subscribers use their lines, the greater the likelihood, with any given degree of concentration, of one of them being unable to get an outside line to the central office upon demand, and therefore the larger the number of lines or trunks required.

Telephone company engineers say that it is rarely economic to install concentrators. Instead, they say, recent advances in electronic technology have provided the economically superior alternative of multiplexing; this achieves the desired reduction in the cost of providing customer access while still giving individual subscribers their own dedicated electronic channels, which they are free to use continuously, twenty-four hours a day. Critics respond that the telephone companies do not consistently exploit the opportunity afforded by concentrators to lower the cost of access.

The issue deserves closer scrutiny, because it is clearly relevant to efficient service. It raises the question of whether the telephone companies’ traditional practice of giving each subscriber a fully dedicated line is economically efficient, and whether, therefore, the charge for access should reflect the costs of that practice.

3. In a subtle variation of the preceding contention, some economists have contended that costs of the subscriber loop are indeed “usage-sensitive” because subscribers who place large numbers of calls will often order a second telephone line. Similarly, customers with switchboards will order fewer or more trunks to connect them to the telephone exchange, depending upon the volume of usage of all the telephones that reach the network through the switchboard. This fact does not, however, make the

51. The device most frequently referred to is “Subscriber Loop Carrier 96,” or SLC 96, which in effect attaches 96 customers to the central office at some point between their premises and the central exchange via a much smaller number of cable pairs. These are, however, capable of carrying 96 simultaneous conversations.

52. N. Cornell & R. Noll, supra note 9, at 20. The Cornell and Noll paper is concerned only with the question of whether local service may properly be regarded as “subsidized,” not with efficient pricing as such. Their contention that the cost of subscriber plant is not wholly NTS, however, bears inevitably—and equally—on both questions.
cost of *each* line or trunk usage-sensitive in the sense that it is higher or lower depending on how much it is used. And that is the relevant fact, since the issue is how to price *each* line or trunk. Advocates of separate subscriber access charges fully covering incremental costs (or more) would of course wish to see them levied on a per-line basis, since the more lines subscribers take the higher the cost they impose on the phone company. In that event, heavy-using customers who order more lines would indeed bear the additional costs that their heavier usage imposes on the system. But they would not and should not do so in the form of charges *per call* or *per minute* of calling, so long as the cost to the supplying company is the same whether they use their lines or not.

Another observer has found a slightly different route to the same logical pitfall. It is generally conceded, he observes, that the amount of transmission line capacity a telephone company has to install between one central office and another—in effect from one switch to another—is usage-sensitive, since it depends on the volume of traffic. The same is obviously true of the number of trunks or lines required between a customer’s own PBX and the telephone office—again, from one switch to another. The only difference between the two situations is that the telephone company owns both switches in the first case, whereas in the second the subscriber owns one of them and the transmission lines are therefore part of the subscriber access network. But the two situations are technologically indistinguishable. If the costs of the first transmission system between the two switches are usage-sensitive, and therefore properly regarded as part of the marginal costs of calling, the costs of the second must be also: how could it be otherwise?

The facts are indisputable; the inference is not. This line of argument simply confuses the marginal cost of increased usage to the *subscriber*, in terms of the *number* of access lines he will require—like the marginal costs of a greater appetite, in terms of the number of frankfurters one buys—and the marginal cost to the *supplier* entailed in more or less usage of *each* access line. No one denies that the provision of additional access lines imposes incremental costs on the telephone company, just as supplying additional frankfurters imposes costs on the meat packer. For this reason, economic efficiency requires that each line (and frankfurter) be priced at the marginal cost of providing it. But it does not at all follow that the marginal costs of each access line should be incorporated in part in the

---

53. D. Chessler, Notes from “Structure of the Communications Marketplace” (Apr. 9, 1986) (unpublished notes on a conference at Dartmouth College) (on file with the authors). See also NATIONAL REGULATORY RESEARCH INSTITUTE, A STUDY OF TELEPHONE ACCESS CHARGES: AN EMPIRICAL ANALYSIS OF BELL COMPANIES IN FIVE REGIONS 166-67 (1983).
charges for using that line, any more than the costs of the frankfurter be recovered in greater or lesser measure depending on what the purchaser does with it. The number of lines a subscriber orders could be and obviously is sensitive to the total amount of calling he does, while the cost per line to the provider might nevertheless be totally insensitive to the amount of calling the purchaser does on that line.

4. The availability of concentrators is an example of a more general phenomenon: the provision of local loops to subscribers and the incurrence of the admittedly traffic-sensitive costs of switches are substitutable. The telephone company has the alternative of putting more small switches or central offices—in effect, concentrators—out among its subscribers, thereby reducing the average length of access lines to them, or installing fewer, larger switches or offices, with correspondingly longer access lines, on average. In consequence, the particular mixture or proportions of TS and NTS costs embodied in either the existing telephone company plant or newly designed systems reflect technology choices that may or may not have been guided exclusively by considerations of efficiency.

The pertinent question, however, is whether the proportions in which these substitutable costs are actually being incurred exaggerates what the marginal costs of subscriber access would be in an efficiently designed system and underestimates the marginal costs of usage. We can think of one reason why they might: the telephone companies can transfer a larger proportion of the costs of the access network than of their switching facilities over to the interstate jurisdiction, via their NTS-cost-recovering access charges to long-distance carriers. This could distort the way in which they balance the two costs against one another when they design the network. Telephone company engineers say they seek to minimize total cost, without regard to who pays for what. In any event, if such a distortion were present, it would itself be attributable to the present system of overcharging long-distance calling and would hardly constitute a defense of that system.

Nina Cornell and Roger Noll suggest that telephone company engineers are not making system planning decisions that minimize total cost. Specifically, they assert that rapidly evolving electronic technology might be expected to have shown up in increasing recourse to remote switches and the multiplexing devices to which we have already referred, on the one side, and correspondingly shorter subscriber access loops, on the other.

54. See N. Cornell & R. Noll, supra note 9, at 62-63.
55. The allocation factor for interstate purposes has been capped at a nationwide average of about 28%, and is scheduled to decline to 25%. See Fowler, Halprin & Schlichting, supra note 8, at 177.
56. Supra note 9.
57. For a graphic exposition, see P. Huber, The Geodesic Network, 1987, Report on
And yet it appears that the telephone company switches are growing larger in capacity and loop lengths increasing. This strongly suggests, they say, that current practices inefficiently inflate the marginal costs of access.\footnote{See N. Cornell & R. Noll, supra note 9, at 63-64.} In rebuttal, William Taylor and Paul Brandon list a number of other considerations that could well explain why, consistently with efficient system design, switching centers have not become more numerous and access lines shorter.\footnote{Among these are the costs of the land and buildings required to house the additional switching centers; economies of scale in components of electronic switching systems; the changes in the interoffice trunking that would be necessitated by more dispersed switches; and the costs of maintaining the additional switches, which, being highly labor-intensive, have increased more than the capital costs. W. Taylor & P. Brandon, Local Telephone Pricing and the Subsidy Question—A Critical Appraisal 19-21 (1986) (unpublished manuscript available from Bell Communications Research).}

We have not seen a clear exposition of the proposition that additional usage tilts the balance of system design on the side of few exchanges and longer access lines, and therefore higher marginal costs of providing subscribers with access. The case for the causal connection, it appears, would run along the following lines: Splitting an exchange into smaller, more numerous units permits reduction in line costs, but at the expense of higher switching and interexchange trunking costs. Splitting reduces the cost of access but increases the marginal cost of calls. It is therefore less likely to be economic the greater the volume of calling. Heavier usage both increases the switching cost penalty and reduces the savings in line costs—the latter because the greater the volume of calling the larger the capacity of the trunks that have to be installed to connect the newly-created separate exchanges. In a nutshell, heavy usage makes economic a technology in which access costs are relatively high and call costs relatively low.

There have been several efforts to test empirically the proposition that the costs of subscriber access are in fact positively correlated with usage and therefore improperly labelled non-traffic-sensitive. For example, Margaret M. Dalton has attempted to explain statistically the costs per subscriber loop classified as NTS by the Uniform System of Accounts, in terms of the number of subscriber loops, the size and density of exchanges, the percentage of the distribution system that is underground, the average minutes of use per loop, and the percentage of call minutes accounted for by toll calls.\footnote{M. Dalton, A Critique of the Classification of NTS Costs (rev. ed. Jan. 1987) (unpublished manuscript on file with the authors).} The resulting correlations suggest that NTS costs are systematically higher in areas characterized by relatively heavy calling and a relatively high proportion of toll calls. Professor Dalton interprets these
results as evidence that some of the costs classified by the Uniform System of Accounts as non-traffic-sensitive are in fact traffic-sensitive and "consequently . . . prices based on these costs will violate the standard of cost causation."  

This conclusion raises two questions. First, are such statistical inferences "correct"? Second, assuming that they are, what is their relevance to pricing policy?  

It is impossible to say whether the correlations Dalton found between calling and NTS costs are real or spurious; pointing out the deficiencies of her study is easier than assessing their effect on the results. The study did not consider, for example, such important determinants of access costs as average loop length (for which the density of exchanges is an only imperfect proxy) and the share of access lines accounted for by business customers (the costs of which, several studies have found, are systematically lower than those of residential lines, perhaps because they tend to be shorter).

More troubling, the costs Dalton analyzed were accounting costs, as they appear on the companies' books. These are heavily affected by when the subscriber plant was installed: because of inflation, average book costs will be lower for old than for recently acquired plant. There is evidence that this failure to take the varying vintages of plant into account may have strongly influenced her results. A study performed by one of us obtained results similar to hers when differences in the average age of assets were ignored. However, when the change in recorded access costs over a four-year period was related to the corresponding change in local and toll calls and other factors, the correlation between call volumes and NTS costs all but vanished. Since this formulation mitigated considerably the problem of not knowing the years in which investments were made and hence the average age of the assets being compared, it should not be surprising that it also proved to have a much higher explanatory power than the Dalton-like specifications estimated from the same data set. And it suggested that the accounting measure of NTS costs is in fact independent of usage.

Finally, even if true current or marginal costs were indeed correlated with call volumes, it would not follow that economically efficient pricing would recover some portion of these costs in the charges for calling. As we demonstrate more fully in Part IV, the marginal cost of providing subscribers access to the network of an efficiently designed system should be
Pricing

recovered in the price charged for access—even though those costs might be lower in some other hypothetical system designed for less loquacious subscribers.83

C. The Cost of Upgrading for New Services

Telephone company witnesses in the proceedings in which we have been involved have uniformly denied that their estimates of incremental subscriber access costs incorporate any incremental costs of providing supplemental, enhanced services.84 The provision of the various Custom Calling services has no effect, they say, on the design of the local loops, but is taken care of in the electronic switches in their central office or in installations on the customer's premises. Their decisions to install fiber-optics in the subscriber access network (in the high-capacity feeder cable that connects central offices with the distribution lines fanning out to subscribers' premises), which is indeed preferable for high speed transmission of data and video signals, are driven, they assert, by its lower capital and maintenance costs, which make it more economical than the traditional copper cable even for POTS. And any special enhancements of the access system or installations required by the specific needs of specific customers are assigned to those customers.

There seems little reason to doubt that the installation of fiber-optic feeder cable is a more economical method of providing for growth of POTS than copper cable. There is active controversy, however, over telephone company assertions that their very large modernization programs, involving mainly the replacement of electromechanical switching with electronic switching systems (ESS), are similarly justified.85 Since the

63. As that discussion will disclose, the applicable principle in the event the system is not efficiently designed is more complicated. See infra text accompanying notes 91 and 92.

64. For many years the Bell companies performed so-called embedded direct analyses (EDAs), which attempted to distribute their revenue requirements (with a very large component of book or embedded costs) among the broad categories of services. In those studies, their witnesses concede, all the costs of subscriber loop, including such enhancements, were indeed included and attributed to the access category.

65. In an extensive, intensely critical report, the Public Staff of the California Public Utilities Commission presented a slashing criticism of virtually all aspects of the Pacific Bell Company's procedures and practices and of the economic merits of its very costly plant modernization programs. PUBLIC UTILITIES COMMISSION REPORT ON PLANT MODERNIZATION OF PACIFIC BELL (1985). The Company had invested $1.1 billion in switch modernization alone over the preceding decade. In re Application of Pacific Bell Telephone Co., California Pub. Util. Comm'n No. 85-01-034 (April 21, 1986) (statement of Lewis Perl). While the major portion of the Company's modernization expenditures have been concentrated on the exchange itself, some of the Staff's criticisms applied to modernization of outside plant as well.

Over and above accusations that the Company had refused to supply necessary information and documentation and failed to monitor the effects of these programs, the essence of the Staff complaint is that the Company's practices and investment criteria have created a systematic bias toward imprudently replacing existing plant at any cost and irrespective of the effects on ratepayers. (In one specific
The economics of the shift is, properly, influenced additionally by the ability of ESS to provide new services as well, the controversy over the economic justifiability of these investments extends to the estimates of the net revenues those services will generate. If, rather than merely minimizing the cost of providing POTS, this expectation tips the balance in favor of ESS, and the companies' expectations about the profitability of those services prove excessively optimistic, the incremental costs of subscriber access may be illegitimately (and therefore inefficiently) inflated.

The importance of this issue and the intensity of the controversies over it are enhanced by the fact that a very large part of the costs of digital switches are indeed non-traffic-sensitive. This is so because the engineering of these systems has the effect of transferring many of the functions performed in the central processing units of the older analog electronic switches, the costs of which do vary with usage, to the equipment at the termination point of the subscribers' lines, which is really part of subscriber plant. In consequence, the substitution of digital for analog electronics or for the old electromechanical switches, even if justified entirely...
in terms of minimizing overall costs, might nevertheless result in increasing the apparent marginal costs of subscriber access.

The pertinent policy question, then (to which we turn in Part IV), would be whether the economically efficient flat subscriber access charge should in these circumstances be increased correspondingly. As our preceding discussion has already suggested, the answer to that question will depend in the first instance on whether the investments in question have been economically efficient, taking into account both the anticipated net revenues from the additional services they make it possible to offer and minimization of total system cost. It is not possible for non-engineers like us to answer that complex factual question definitively. We will try in Part IV to elucidate the pertinent principles, and will also comment briefly on the totally unsatisfactory character of our present institutional arrangements for seeking answers, poignantly illustrated by the regulatory imbroglio in California over the legitimacy of the telephone companies' modernization expenditures and the legitimacy of charging them to POTS customers.68

D. Marginal vs. Embedded Costs of Subscriber Access

The contention that the marginal costs of subscriber access are lower than the average revenue requirements of non-traffic-sensitive plant, as traditionally measured, has several major components:

First, telephone companies build capacity in lumps, so that typically there will be excess capacity in both their high-capacity feeder cable and distribution lines. This means that short-run marginal costs of subscriber access—the cost of adding another customer to the system—are ordinarily below long-run or average costs.69

Second, the companies systematically overinvest in capacity, so that the proper marginal cost of subscriber plant is typically lower than the long-run marginal and average costs they actually incur.70

Third, many telephone companies have large surpluses of Centrex lines, either installed in anticipation of demand that never materialized or left stranded when the service was replaced by customers installing their own PBXs. As a result, once again, both short- and long-run marginal costs may fall far short of average revenue requirements.71

Fourth, a large part of the regulatorily-determined revenue

68. See supra note 65.
69. See G. Brock, supra note 4, at 27-29; Wilson, supra note 48, at 21-22.
70. N. Cornell & R. Noll, supra note 9, at 74.
71. Id. at 41. Obviously, to the extent the capacity can be or has been used to serve other customers, this condition does not prevail.
requirements associated with subscriber plant has consisted of the carrying charges on the ("non-recurring") costs associated merely with the repositioning of existing phones rather than net new installations. Instead of being charged to the responsible parties, these costs were until recently capitalized and charged to ratepayers at large.

Finally, the capital charges component of revenue requirements, which are extremely important because the telephone business is so capital intensive, are determined, under regulation, on the basis of book (depreciated original) cost. Because of inadequate depreciation rates in the past, the book value of telephone company investments far exceeds current market value; the estimate of this excess most widely cited is on the order of twenty-five billion dollars.

Although we are not in a position in most cases to assess their quantitative importance, each of these assertions appears to be factually correct, with the possible exception of the second, on which we are unable to make a judgment. The critical question about each, as about most of the other factual assertions we have surveyed in this section, is what significance it has for the economically efficient pricing of telephone services. To answer that question it becomes necessary to expound the applicable regulatory and economic principles—to which task we now turn.

72. Id. at 26-27.
73. It was only in 1981 that the FCC reversed that practice and required the direct expensing and charging of these costs to the responsible customers. Uniform System of Accounts, 85 F.C.C. 2d 818 (1981).
74. Letter from Irwin D. Fries, Chairman of the U.S. Telephone Association, ad hoc Committee on Capital Recovery, to Depreciation Rates Section, Common Carrier Bureau, FCC (Dec. 10, 1986) (citing $26 billion as "the most appropriate definition") (copy on file with the authors). For a New York Public Service Commission staff member's disagreement with the $26 billion figure, see Telecomm. Rep., Oct. 27, 1986, at 9.
75. That does not mean they are all equally relevant. For example, while there seems to be no doubt about the overvaluation of telephone company plant on the books and the consequence that on average marginal costs fall short of average revenue requirements, it is less clear that that is true of the subscriber access system. In view of the large component of (current) labor and materials in the costs of installing and maintaining that network, it is more uncertain here than in the other parts of the business which way the balance of the two offsetting determinants of this relationship tilts—the effect of technological progress outrunning historic depreciation rates, on the one hand, and the cumulative effect of historic inflation on current costs, on the other. W. Taylor & P. Brandon, supra note 59, at 18.
76. This is not to deny the importance of the factual questions. The Bell Operating Companies frequently contrast the current average bill for local service of about $12 per month with embedded costs (i.e., the revenue requirements as traditionally determined) of providing subscriber access of $26 per month. See J. Lande & P. Wynn, Primer and Sourcebook on Telephone Price Indexes and Rate Levels 56 (1987) (published by Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission) (providing 1986 figures for average bill for local service); Congressional Budget Office, The Changing Telephone Industry: Access Charges, Universal Service, and Local Rates 12-13 (1984) (providing 1981 figures for embedded costs of furnishing subscriber access). Some more recent studies of the current or long-run marginal costs of subscriber access alone (referred to by some companies as dial tone line cost studies) have produced lower results. Southwestern Bell has produced an estimate of $17 for Oklahoma (plus an additional
IV. The Applicable Principles: First-Best

Only someone who has been exposed to the flood of adversarial testimony on these subjects during the last several years can appreciate how difficult and how necessary it is to try to sort out the economics from the politics and the relevant arguments from the irrelevant. Clarifying the pertinent underlying principles alone should resolve many of the policy issues. In the remaining cases, it should at least help pose the pertinent factual questions; we will never get the policy answers right if we do not ask the right questions.77

A. Economically Relevant and Irrelevant Costs

The basic principle of "first-best" economic efficiency is that all goods and services should be priced at the marginal costs of providing them. (Having stated the principle, it becomes necessary at once to attach at least two qualifications to which we will return—first, "to the extent it is economically feasible to do so," and, second, "except insofar as transactions have external consequences not reflected in the private costs and benefits that motivate them.")

Under this principle, the factual issues that were the subject of Part III are, in large measure, arguments about economically irrelevant measures of cost. This is almost certainly true of the contentions that the present access system developed, historically, in order to accommodate long-distance calling. It is true also of the claim that increasing usage has been responsible for the prevalence of single-party service and the consequent limited availability of multi-party service. Even if those assertions were

$7.39 for average local usage). Kahn Oklahoma Testimony, supra note 21, at 38. Cornell and Noll present extremely telling criticisms of a C&P Telephone Company study that produced estimates of $41 and $23 monthly incremental costs for Maryland and the District of Columbia, respectively, and contrast it with estimates by GTE of Wisconsin of $7 to $14 for various groups of its subscribers, and $3 to $14 for Illinois subscribers by Illinois Bell. They also present analyses of Rural Electrification Administration costs suggesting that incremental costs are already being covered by prevailing charges. N. Cornell & R. Noll, supra note 9, at 64. A recent study by the New England Telephone Co. has produced an estimate of $12.50. New England Telephone Company, Incremental Cost Study for the State of Massachusetts 3 (filed with the Massachusetts Department of Public Utilities on Apr. 18, 1986) (unpublished report on file with the authors). On the other hand, Jackson and Rohlfs continue to regard $15 to $30 as the plausible range of marginal costs of subscriber access. C. Jackson & J. Rohlfs, supra note 16, at 4.

Manifestly, as Cornell and Noll point out, some of the later estimates cast doubt on the extent to which "basic service" (which, however, typically includes unlimited local calling as well) is today still "subsidized" in the sense of being priced below marginal cost—a question to which we return below.77 This is not to suggest, naively, that effective regulation consists simply in separating truth from falsehood or invariably giving priority to the dictates of economic efficiency over other considerations. Policy is not likely to be intelligently made, however, if economists present regulators with purportedly economic arguments that are in fact simple fallacies, or with recommendations purportedly justified on non-economic grounds without considering whether the recommended courses of action would achieve those noneconomic goals at minimum economic cost.
factually correct, they would not provide an economic justification for levying the carrying costs of the access network on usage today. These incremental system design costs would be part of the marginal costs of usage (or long-distance usage) and properly reflected in its price only to the extent that doing so would result in society saving those costs in its future expenditures on providing access.

Economically efficient pricing looks not to the past—not to how we got where we are—but to the future; efficiency requires that prices tell customers what incremental resources society will use if they take more of the good or service in question, what resources society will save if they consume less of it. If, as seems to be the case, charging usage the higher access costs it has assertedly caused would not reverse the historical process that produced the present access network, then the incremental costs of neither single-party (as compared with multi-party) access nor of designing the system for long-distance calling are part of either the short- or long-term marginal costs of calling. In that event, these costs should not be imposed on usage, whatever its past responsibility for the present configuration of the system.

78. Wilson is correct in pointing out that where multi-party service is still available, use of the shared access line by one subscriber imposes congestion costs on the other parties. Wilson, supra note 48, at 20. Those marginal congestion costs should in principle be reflected in the charges for calling. The relevance of that observation to pricing policy is unclear, however, in view of Wilson's opposition to local measured service and the apparent infeasibility of measured charging for local calls on multi-party lines.

The phenomenon of usage imposing congestion costs on other subscribers is a universal one, in no way confined to multi-party service. Economic efficiency clearly calls for reflecting those marginal congestion costs—which are obviously traffic sensitive—in charges for calling, as long as the efficiency benefits of doing so exceed the costs. We discuss that concluding qualification in Part IV-C below.

79. According to Gerald R. Faulhaber, it was Bell's acquisition of control over the process making possible cheap, reliable, and audible long distance service that enabled it by 1913 to regain the industry dominance it had quickly lost upon expiration of the original Bell patents twenty years earlier. G. Faulhaber, Telecommunications in Turmoil, Technology and Public Policy 20 (Sept. 4, 1986) (unpublished manuscript on file with the authors).

80. It may seem anomalous to contend that the level of marginal costs depends on whether incorporating certain costs in price would or would not reverse the historical process: we do not ordinarily think of the marginal cost of a particular service—in this case, calling—being dependent on the elasticity of its demand. If the demand for calling were so elastic that charging it with the incremental design costs that some commentators claim it has forced on the access network would result in a less expensively designed network, and in that event callers should clearly be confronted with prices incorporating those incremental costs, why should they not even if their demand is inelastic?

The answer lies in the straightforward economic principle that marginal cost is not a unique number but a functional relation between output and cost, whose value will vary from one part of an industry's supply function to another. The value that is relevant for efficient pricing will be determined by the level and elasticity of demand at various prices. If the demand for the ability to place and receive long distance calls is so strong and unresponsive to price within the relevant range as inalterably to require a network universally capable of satisfying it, then the costs actually imposed on society by customers subscribing to more rather than fewer lines and making more or fewer calls are what they are in that situation—not what they would be if those services were supplied by some hypothetical network inconsistent with subscribers' demands. In those circumstances, it would in fact encourage oversubscription to lines and underutilization of them if subscription were charged less and
Pricing

The assertions about how the present system evolved are really arguments about sunk costs—fixed costs that have already been irretrievably incurred. There is no truism in economics more elementary than that sunk costs are to be ignored in deciding how best to use the resources that are available to us today and that will be available in the future.

This truism does not mean that first-best efficiency ignores fixed costs or makes no provision for their recovery. Nor does it mean that, as some witnesses in state regulatory proceedings have contended, fixed and non-traffic-sensitive costs are one and the same, or that advocates of recovering the latter in a flat subscriber access fee are therefore simply proposing that all fixed or capital costs be recovered in the flat monthly charge for POTS, and only variable costs in usage charges. Those contentions confuse two distinct dichotomies—one between fixed and variable costs, the other between usage-sensitive and non-usage-sensitive costs.

Short-run marginal or incremental costs do not explicitly include the costs; they include variable costs only. Long-run incremental costs, in contrast, do explicitly include the costs associated with adding capacity. As a matter of pure economic principle, prices should ideally be equated to the former of these, which, properly defined, include opportunity costs (congestion, for example) as well as production costs. Although pricing at short-run marginal cost (SRMC) seems to ignore fixed costs, it is not incompatible with their recovery. On the contrary, SRMC will be above or below average variable costs, depending upon the relationship of production to capacity; so pricing at SRMC will make a contribution to the coverage of capital costs—a contribution that will at times fall short of, and at other times exceed average recovery. Therefore, recommending that calling more than the costs society would save, respectively, if customers took less of either of the two services.

The same principle would apply to the question of whether the subscriber access charge properly includes the additional NTS cost attributable to the shift from electromechanical and analog to digital switching. See supra note 65 and accompanying text. If the latter is the most efficient technology and if it involves higher marginal costs of access than a less efficient technology, then those higher costs are what the access charge should reflect. On the other hand, if the enhancement of marginal subscriber access costs is causally attributable to (a) introduction of other services or (b) reduction of marginal costs of usage, and if levying those incremental NTS costs on those other beneficiary services would render digital switching uneconomic, then presumably the switch to digital would in fact be economically unjustified and the marginal cost of (economically efficient) access overstated.

81. "By definition incremental cost ignores all embedded or fixed (NTS) costs." In re Application of Southwestern Bell Telephone Co., Oklahoma Corp. Comm'n Cause No. 29,321, at 18 (1985) (statement of Thomas H. Fish) (emphasis added).

82. Cf. Wilson, supra note 48, at 20 ("To argue . . . that once a loop is installed its costs are sunk, and therefore marginal cost rules dictate that the price for access use should be zero, is similar to arguing that once a nuclear power plant is built its costs are sunk and nuclear power should be sold at . . . a price equal to only the variable running cost of the plant.").

83. See I. A. Kahn, supra note 6, at 70-75; Andersson & Bohman, Short- and Long-Run Marginal Cost Pricing, 7 Energy Econ. 279 (1985).
local and long-distance calls be priced at SRMC in no sense involves assigning the recovery of fixed costs exclusively to the subscriber access price.

Moreover, it will ordinarily be more practicable to base utility rates not on short-run, but instead on long-run incremental costs (LRIC). To the extent that in the long run additional consumption or use of a particular service will involve additional investment, the capital costs associated with that investment will be explicitly included in LRIC, and therefore in price.

For example, as usage of the telephone system increases at busy hours, the consequence is additional congestion on the lines. This congestion is a short-run marginal (traffic-sensitive) cost. In order to maintain previous standards of quality, telephone companies must expand capacity. The capacity costs they incur in this way are traffic-sensitive, because they are marginally attributable to usage, and may be regarded as the long-run marginal cost equivalent of the congestion costs that they mitigate. In equilibrium, the two will be equal: companies will add capacity (i.e., incur incremental capital costs) only up to the point where the costs of doing so are equated to the congestion costs that they relieve. It is clear, then, that both usage-sensitive and non-usage-sensitive costs can be fixed or variable; conversely, both fixed and variable costs can be usage-sensitive or non-usage-sensitive.8

The proper treatment of sunk costs and of the difference between long-run and short-run marginal cost is raised also by the question of whether the basic charge for telephone service should cover the costs of spare access capacity. The threshold question, under first-best pricing principles, is whether the amount of capacity the telephone companies install at any one time is justified in terms of cost minimization. It is very expensive to go back to the distribution cable, which runs between the heavy capacity feeder cable and the subscribers, in order to install additional capacity when it is actually needed. The companies, therefore, typically install extra wire pairs, whose incremental costs are very low, when they make the initial installations, in order to accommodate future demand. In the case of the much larger feeder cable, in contrast, it is feasible to put additional bundles of wire pairs into the conduit at intervals over time, and it would be also very expensive to build capacity far in advance. Here, therefore,

8. If one were to envision, therefore, a two-by-two matrix, with the first distinction running across the top (one column representing usage-sensitive and the other non-usage-sensitive costs), and the second down the left side (one line representing fixed and the other variable costs), one can readily determine that all four boxes will be occupied. To complete the illustration it is necessary only to observe that NTS costs will embrace both fixed carrying charges on the investment in subscriber plant and variable costs of maintenance and service, to the extent these costs too are independent of usage.
companies will normally install enough cable to handle only a few years' demand growth. These calculations necessarily give substantial weight to the fact that the cost of physically installing the added capacity, which is largely independent of the amount of capacity added, is a large part of the total cost; this consideration makes it efficient to add capacity in large lumps, and inevitably results in the presence of considerable spare capacity at any given time.

To the extent, then, that the minimum incremental cost of providing subscriber lines includes the cost of carrying spare capacity for growth, economic efficiency clearly requires that all of that cost be reflected in the access charge.\(^8\) If, however, installation of capacity has exceeded the point of cost minimization, as critics of the companies assert,\(^6\) the question becomes one not of first-best pricing but whether those sunk costs should be recovered at all, and, if so, how to recover them with the smallest loss of efficiency.

Regulators in the great majority of states that follow original cost or prudent investment principles cannot simply ignore sunk costs. They cannot avoid considering, for example, whether the telephone companies may have designed a system of higher quality, at higher investment cost, than was justified.\(^7\) If they find those investments were imprudently undertaken, they are entitled—indeed obligated—to disallow recovery of the associated carrying charges. By the same token, to the extent those expenditures were prudently made, regulatory commissions have an obligation, in original cost jurisdictions, to permit the companies a reasonable opportunity to recover their sunk costs.

These regulatory principles apply with especial force to the assertion that the marginal costs of access are likely to be markedly below average revenue requirements, as traditionally determined, because a very large component of the revenue requirements is comprised of the carrying charges on the capitalized cost of telephone installations, turnovers, and moves. This fact hardly constitutes an argument against recovery of those costs, considering that the regulatory commissions themselves have histori-

---

85. We consider infra at text accompanying note 118 whether the charge should vary over time, reflecting the changing relationship of SRMC to LRIC.

86. See N. Cornell & R. Noll, supra note 9, at 74.

87. That is, the telephone companies may have introduced incrementally higher levels of reliability or capacity, the cost of which exceeded the benefits of doing so. It is a commonplace of the regulatory literature that such a policy will be in the interest of a company regulated on the basis of traditional rate base/rate of return methods whenever it appears that the return on incremental investments will exceed the incremental cost of capital. Averch & Johnson, Behavior of the Firm Under Regulatory Constraint, 52 AM. ECON. REV. 1052 (1962).
cally not only sanctioned but also insisted on capitalizing them rather than recovering them directly from the responsible parties.\textsuperscript{88}

In any event, however, the question of what proportion of these sunk costs telephone companies should be permitted to recover is \textit{totally irrelevant} to the question of how costs that are determined to have been prudently incurred should be recovered from ratepayers under first-best principles. Economic efficiency requires that they be ignored, in the first instance, and that rates be set at the marginal costs of an efficiently designed system. To the extent, then, that such rates fail to produce sufficient total revenues, we move to the world of second-best, which is the subject of our Part V.

The foregoing enunciation of first-best principles applies equally to the assertions that heavy usage discourages the substitution of remote switches for long access lines,\textsuperscript{89} and that the telephone companies are attempting illegitimately to load on to POTS customers a portion of the costs of upgrading their non-traffic-sensitive access networks in order to offer various sophisticated services. The first question is \textit{whether those investments are economically efficient}, minimizing the combined costs of access, calling, and the newer services; and in this assessment it is necessary to take into account the sufficiency of the incremental net revenues flowing from the services the investments make it possible to offer. If the expenditures are efficient—that is, if they conduce to the efficient design of the entire system—then the marginal costs of the several services at which their prices should be set are their marginal costs under that system. Specifically, if the efficient system entails a higher proportion of NTS subscriber plant costs than some other design, the economically first-best flat rates to POTS customers will reflect those higher costs: The marginal costs of access \textit{are what they are} in the system that is optimally designed to satisfy all the demands it serves.\textsuperscript{90}

What if, however, the current investment decisions of the telephone companies affecting the marginal cost of subscriber access are indeed \textit{not}

\textsuperscript{88} On the change in this practice in 1981, see \textit{supra} note 73.

\textsuperscript{89} \textit{See supra} text accompanying notes 54-59.

\textsuperscript{90} It is theoretically possible that following this principle would result in a price of access for POTS above the costs of providing it alone—i.e., above its so-called stand-alone-costs—and therefore perhaps above the levels that would prevail if competition were feasible. This is the hypothetical situation developed in the literature recently on the sustainability of a monopoly based on economies of scope in the face of competitive entry, which has demonstrated that entry might actually be incompatible with economic efficiency. \textit{See} W. Sharkey, \textit{The Theory of Natural Monopoly} ch. 5 (1982). Whether we are actually presented with such a dilemma here—whether, that is, the efficient price of access would exceed the average total cost of supplying it alone—is a factual question that we are not in a position to answer. Whatever the answer, however, it would not alter the definition of the efficient solution.
efficient? What if they inefficiently inflate those costs? Here the prescription for economic efficiency becomes more complicated. Conceivably, the companies should be made to absorb the costs of the mistakes—whether because the investments were demonstrably imprudent or under some other regulatory principle. But such a rule would result in inefficient purchase decisions by those companies’ customers if prices—in this case the charge for subscriber access—did not reflect the full marginal costs that the companies, and therefore society, actually incur, even if they are the marginal costs of an inefficiently designed system. The first-best resolution of this dilemma would be to set prices on that latter basis, and to apply the penalty to the totality of their costs that the companies are allowed to recover. Reconciliation of the total revenues that would flow from consistent marginal cost pricing and the aggregate revenue constraint is the task of second-best, which we discuss below in Part V.

B. Collective Consumption Decisions

The proposition we have just enunciated may seem unjust. Why should the humble POTS customer be required to pay a price set at marginal costs that have been elevated by the demand for such exotic services as actually using the telephone to place or receive calls, local or long-distance, or to make possible high speed data transmission? Why not reflect the asserted differences in the costs of designing access systems suitable respectively for local calling, long-distance calling, and more sophisticated services in correspondingly differing flat monthly charges to customers depending on the kind of service to which they wish to subscribe?

Some economists have recommended that regulatory commissions estimate the costs of designing these several hypothetical systems, and propose that access costs be distributed among the respective classes of customers in proportion to those separate costs—the same kind of alternative justifiable expenditures or relative cost method that the Tennessee Valley Authority used to apportion the costs of its multi-purpose river development system among the several services it provided. If such differentiated

91. See supra notes 38, 65, and 70.
92. It is possible that the dilemma would disappear as the companies, suitably penalized, mended their ways, so that the discrepancy between the marginal costs they actually incur and marginal costs under an efficiently designed system would disappear. To the extent, however, that the system design is irreversible (or at least not quickly reversible), the efficient solution would be the one we have described.
93. The same logic would dictate offering a flat monthly access charge equal to what the subscriber loop would cost in a hypothetical system designed to minimize system costs if there were no calling at all. This result is a true reductio ad absurdum: presumably that cost would be zero—and so all the costs of the actual access system would have to be levied on usage—since there is no need for an access system at all if there is to be no calling.
rates were confined to the flat monthly charge, they might do little economic damage: because of the inelasticity of the demand for access, prices both below and above marginal cost would not significantly affect consumption. Moreover—setting aside the very large question of whether its factual premises are sound—the recommended departure from first-best pricing might be attractive on equity grounds. The fact that it has been proposed by economists should not, however, deceive anyone into thinking that it provides a measure of economic cost.

Moreover, our speculation that it would probably do little economic damage is subject to two major qualifications. First, the lower charge for POTS must not be financed by inefficiently raising the charges for usage: that does do a great deal of damage. Second, it must be possible to make such differentiations among customers at a cost low enough for it to be possible to offer the lower-priced service without unfairly and inefficiently burdening the others. It appears unlikely, however, that any such differentiations are either economically or politically feasible.

Competitive markets have the virtue of offering consumers a variety of price and quality options, but that spectrum of offerings is not unlimited. It is not economically feasible to provide all conceivable packages. For example, there may be some automobile buyers who would prefer to buy cars without bumpers or fenders, at a correspondingly reduced price; but in view of the economies of producing standardized models, it probably would actually be more costly to satisfy their idiosyncratic desires than to supply them with the models preferred by the great majority of customers. In that event, they have no legitimate complaint about not having

95. The factual premise appears not to be sound in the case of long-distance calling. Design of the subscriber access network for that purpose is evidently no more costly than for "local service." In these circumstances, the only possible purpose of this exercise (which would presumably require blocking subscribers to "local service" from the long-distance network) would be artificially to fashion an inferior service to be offered on an explicitly subsidized basis.

96. As one of the authors has previously written about the Federal Power Commission:

The FPC was later prevailed on to accept a similar method, the 'relative cost method,' for allocating the costs of production on joint-product leases between natural gas, on the one hand, and oil and various natural gas liquids, on the other, in order to ascertain a 'just and reasonable' field price for the natural gas. The method involved distributing the joint costs of producing those same quantities in proportion to the actual costs of producing that same number of barrels of crude oil on leases in which oil was produced in the absence or virtual absence of natural gas, on the one hand, and that number of cubic feet of natural gas from virtually dry gas leases on the other . . . . In advocating this method, at a later stage only for pricing gas that had already been discovered and committed to pipeline purchasers, the present writer emphasized that while it might be deemed to provide a just or a fair distribution of the joint costs, it did not provide an economic measure of the separate costs.

1 A. Kahn, supra note 6, at 151.

97. It would apparently be costly to block subscribers to local service only from the long-distance network; there might be very little demand for the service if the costs of blocking were incorporated in the charge for it; and there might be objections on grounds of both efficiency and equity to imposing those costs instead on the customers subscribing to the long-distance option.
available to them, at a lower price, a stripped-down version that would have to be custom-made.

The same constraints prevail in the provision of public utility services. There is a great deal of internal subsidization among the great majority of customers who subscribe to flat rate local telephone service; light users, or subscribers who do all their calling off-peak, are forced to pay a uniform charge based on average costs that are inflated by other subscribers who do a great deal of talking during periods of peak demand on the system. To the extent that the costs of metering to separate those two categories of users are prohibitive, the light users have no legitimate complaint. The same is true of purchasers of electric service who confine their consumption to off-peak periods, yet who for the same reason cannot be offered a lower price corresponding to the lower than average costs that their consumption habits impose on the system.

Similarly, it may well be infeasible to supply the minority of subscribers who have no desire to be connected to the interexchange network, either to place or to receive long-distance calls, with the limited service they want except at costs that they themselves would regard as excessive. The system may therefore have to be designed with the facilities and quality of service that maximize the net benefits to all subscribers collectively. The same reasoning may well justify such telephone company practices as providing subscribers with individual, dedicated access lines, regardless of how much or how little they use them, and the recent conversions to digital electronic switching.

The kind of telephone network that we have, in short, inevitably represents a collective consumption decision. Because it would probably have been impractical for telephone companies to offer two or more systems, of varying capability, it became necessary to decide, in effect collectively, which quality offered the largest differential between benefits and costs to all subscribers together. The choice of a system suitable for long-distance calling was almost certainly the correct one. It is worth recalling, in this connection, that a large proportion of what typically goes under the rubric

---

98. This, it seems to us, is the likely answer to the complaint by Wilson about the injustice of older people who seldom use their phones being compelled to pay the access costs of a system designed for families with teenagers. Wilson, supra note 48, at 19. It might well cost the phone company much more than it could possibly save if it were to try to match the kind of subscriber plant it extends in different places to the varying and changing telephonic garrulousness of its transient and aging population of subscribers.

99. We emphasize "in effect" because the "collective decision" to make certain options available and others unavailable except at prohibitive cost will in most circumstances not require any single, centralized decision. Instead, as in the case of the standardized automobile equipped with fenders, it will emerge as the consequence of the sum total of individual purchase decisions in the market.

100. In fact, this choice was apparently made, at least in part, by the free choice of consumers between telephone companies offering and not offering the option. See supra note 79.
of "local calling," because it is embraced within the local charging system, is itself long-distance, in the sense that it is interexchange rather than intraexchange.\footnote{101}

Discussion of whether people who want only "basic service" should have to pay for an access system adapted also for long-distance has been confused by uncritical assumptions about what constitutes "basic service," which we presumably want to make available to as many people as possible at the lowest possible price. Consumer advocates tend to equate it with "local service." This equating of "basic service" and "local service" is questionable, at best. In an important sense, the service that is basic is attachment to the entire network, which conveys the ability to place and receive calls everywhere and anywhere. It is no wonder that we think of local calling as a necessity and long-distance as a luxury today, when most subscribers pay nothing for individual local calls and pay for long-distance calling by the minute and the mile at rates many times incremental cost. It might well be argued, indeed, that the ability to telephone one's neighbor is in a real sense less basic than to call someone on another exchange: one can always let one's feet do the talking in the former case. A rural subscriber's conception of "basic service" will for this reason be quite different from that of an urban dweller. And the popular notion that the percentage of income that households spend on long-distance calling goes up as their income rises is simply incorrect, even at present inflated toll charges.\footnote{102}

C. The Problem of Aggregation

A service—more precisely, what the economy prices as a single service—may be more costly to supply to some customers than to others. For example, the cost of providing access to the telephone network within a city can vary substantially, depending (among other things) on the distance separating a customer from the closest central office. Just as it can be inefficient to tailor services to idiosyncratic preferences, so it can also be inefficient to tailor the price of some standard service to reflect all of the variations in the circumstances of individual customers that may affect the

101. See supra text accompanying notes 45 and 46. This is not to say that we think all the collective decisions that have been made about the design of the system have been correct. One of us has expressed skepticism about the wisdom of our social decision to move from a standard of multi-party to single-party telephone service. Kahn, supra note 8, at 151 n.67. Had we priced each to consumers at its true economic cost, there might still be enough people subscribing to the former, lower-cost service to keep that option more widely available than it now is, as a kind of lifeline, and it might be unnecessary for us to look for ways of subsidizing service to low-income families in order to keep them on the system. Instead, we subsidized the transformation by underpricing the more luxurious option, and as a result the multi-party option has become increasingly available.

102. See infra note 155.
cost of supplying it to them. The proposition that first-best principles call for pricing each good or service at its marginal cost ignores the cost of administering multiple separately-priced offerings. Realistically, first-best efficiency necessarily entails some aggregation or bundling.\textsuperscript{103}

The pricing of telephone services involves two types of aggregation. One is "rate averaging," the practice of charging the same price for a group of services despite differences in the cost of supplying them. For example, toll rates are typically uniform on a per mile basis, despite apparent economies of density, which produces lower unit costs on heavily used than on lightly used routes. Similarly, telephone subscribers typically pay a basic monthly charge that disregards differences in the costs of providing them access, which depend on their distances from the nearest central office and on whether the cable must be placed above or below the ground.\textsuperscript{104} The other kind of aggregation is "bundling"—charging a joint price for a combination of two or more services, rather than pricing each separately; the familiar example we have already encountered is flat-rate local service, whose single price covers both access and unlimited local calling.

These two types of aggregation have two features in common: both produce inefficient customer choices by providing misleading price signals; but both save on metering and billing costs. For example, the uniform price for all toll calls in a broad distance interval or time period that is correct \textit{on average} will inefficiently discourage calls in the group that are less costly to provide (for example, calls on high-density routes during the less busy part of the rate period), while inefficiently encouraging more costly calls (for example, calls made on less-dense routes during the busy hours). Similarly, flat-rate pricing of local service fails to deter subscribers from making calls whose value is less than their marginal cost, and (if the flat rate exceeds the cost of access) inefficiently discourages light users of local calling from subscribing at all.

Disaggregated or unbundled pricing would eliminate these distortions, but would entail additional costs of measurement, billing, and inconvenience to customers. It therefore produces an improvement in economic efficiency only where the gain from improved price signals exceeds the increase in those other costs. The case for price disaggregation is therefore strongest when (a) aggregation conceals large cost differences among the services that are bundled or uniformly priced; (b) demands are price-elastic, so that customers would markedly alter their usage in response to

\textsuperscript{103} For a discussion of the relevant principles and their application in various contexts, see A. Kahn, supra note 6, at 83-86, 103-109.

\textsuperscript{104} There are exceptions. In some rural areas a "line extension charge" is added to the normal access charge when the customer's premises are outside the primary service area, or when the terrain makes access particularly costly (e.g., connecting a house perched on a mountain top).
more finely differentiated prices; and (c) the additional administrative costs are small.

In the following discussion we apply these criteria to assess two major aggregation issues confronting telephone companies and their regulators: (1) whether to offer local measured service on either an optional or mandatory basis; and (2) whether and how much to de-average various rates both geographically and temporally—that is, on the basis of geographic differences in marginal costs and variations in marginal costs over time.

1. Local Measured Service (LMS)

Most calling, by far, is local. A number of conclusions from the preceding analysis strongly suggest the desirability of charging for local service on a measured basis, with a flat charge for access and a separate, marginal cost-based charge for each local call:

- The fact that subscriber access and local calling are, in economic terms, separate services, with separate marginal costs;
- Our view that access is more “basic” than unlimited local calling;
- The likelihood that the cost of providing an individual with access is insensitive to usage, which dictates that it be priced separately on a flat or a lump sum basis; and
- The unfairness, inherent in an exclusively flat rate option, of subscribers who make few calls (especially on-peak) in effect subsidizing heavy users.

We will, in later discussions, recognize the following additional attractions of LMS:

- The possible desirability of recognizing differences in marginal subscriber access cost from one group of customers to another;


105. In 1983, 84% of all call minutes on the Bell System’s switched network were local, and 16% were toll. P. HUBER, supra note 57, at 2.3; cf. the somewhat different estimate for 1984 at note 143 infra. One of the most curious aspects of the testimony of consumer advocates who contend that usage increases access costs is that they have at the same time expressed hostility to the introduction of LMS, presumably even on an optional basis. This hostility doubtless partly reflects the comparative unfamiliarity of measured service to many subscribers, and uneasiness about their ability to restrain either themselves or their families from running up large bills. These concerns could easily be accommodated by making the offer of measured service optional only.

The ultimate reason for the posture of consumer advocates is probably the suspicion that the telephone companies are proposing the introduction of LMS in the hope that it will make it easier for them eventually to raise basic subscriber rates for both access and local usage. According to this scenario, they hope to begin with the charge for the more luxurious flat rate service, while holding down the access portion of the LMS rate. Then, once they have succeeded by this strategy in inducing large numbers of subscribers to move over to the measured option, they will raise the flat monthly portion of that rate as well.
Pricing

* The possible desirability, whether on ground of externalities or because of social and political considerations, of offering access alone at a subsidized price, preferably selectively, in order to keep attached to the telephone system people who might otherwise drop off; and

* The possible need for pricing access and local calls at varying markups above their respective marginal costs in order to provide the telephone companies with adequate revenues.

A virtue of competition is that it unbundles services and offers customers a variety of price and quality options, reflecting their respective costs to society. The analogy in the telephone case would be to give people who are willing to exercise restraint in the costs they impose on the system—whether by restricting the volume of their calling or concentrating it in off-peak periods—the opportunity to keep down their bills correspondingly. Offering this option would also drastically reduce the cross-subsidizing of heavy users at peak hours by light users.

This listing of the attractions of LMS merely returns us, however, to the other side of the equation—the costs of such a disaggregation of services—and the three factors that determine the balance of costs and benefits: the extent of variation in costs concealed in the aggregation, the elasticities of demand, and the costs of administration. And these, it appears, combine to make the narrow economic case for LMS an extremely modest one.

The reasons that LMS is not more attractive have to do with the relatively small and probably declining proportion of telephone company costs that vary with usage and are therefore susceptible to being saved by charging for each call (which is another way of saying that the variation in the costs of serving low and high-usage customers is slight), and, second, the relatively low elasticity of demand for local calling. Some critics have contended, therefore, that the savings LMS is likely to induce are too small to justify the costs of administering it. They also suggest that a truly cost-based measured service option would not be very attractive, because the cost-justified differential between the charges for access alone and for flat rate service would be relatively small: the major costs of providing “local service” are the NTS costs of the access network itself.

Moreover, the costs that LMS would enable the telephone companies and subscribers to save, on the one side, and of administering LMS pricing, on the other, are not truly commensurable; a dollar of saving would

not justify incurring a dollar of cost. The reason is that metering and billing costs are a dead weight social loss; their only economic value is the savings they make possible. In contrast, any savings in telephone company costs and customers’ bills are not an undiluted social gain: in order to measure the net benefit, it is necessary to deduct from them the value to consumers of the calls they forego under the influence of price, or the inconvenience of shifting them from peak to off-peak.¹⁰⁷

The foregoing assessment of the three determining factors we have identified may be unduly pessimistic. The costs of measurement and billing are far lower on modern electronic than on older types of switches. Moreover, the changeover by the Bell operating companies (BOCs) to ESS has been greatly accelerated by the requirement of the Modified Final Judgment (MFJ) in the AT&T antitrust suit that they equip themselves to offer all interexchange carriers equal quality access to their local networks.¹⁰⁸

Likewise, the elasticity of demand for local calling in the aggregate is less pertinent than the elasticity of substitution between calling on-peak and off-peak. The latter will be higher than the former: people who may find it difficult to dispense with local calls entirely may nevertheless be induced by marginal cost-based price differentials to shift the timing of some to off-peak.¹⁰⁹

As for the small possibility of savings because of the relatively low proportion of total telephone company costs, on average, that are sensitive to usage: first, the ratio is higher than average for calls on-peak, when additional calls necessitate additional switching and trunk capacity; and, second, it will be higher also for heavy-calling than for light-calling customers. A cost-justified differential between the charges for measured and flat rate service that would be unattractive if it reflected only the traffic-sensitive costs imposed on the system by the average call and average customer will inevitably be more attractive as it comes to reflect the costs imposed by the heavier users on-peak, who are most apt to stay with the flat rate option.


¹⁰⁸. See United States v. American Tel. & Tel. Co., 552 F. Supp. 131, 227 (D.D.C. 1982), aff’d, 460 U.S. 1241 (1983). The MFJ exempts the BOCs from this requirement where they can demonstrate that it would cause them extreme hardship. The Department of Justice reported to the District Court on October 31, 1986 that 70% of the Bell company lines had been so equipped, and that it regarded them as in substantial compliance with the requirement. BOCs Rate Well on Equal Access, TELEPHONY, Nov. 10, 1986, at 17.

¹⁰⁹. Selwyn, supra note 106, recognizes the relevance of the elasticity of substitution, but contends it is likely to be inadequate to generate benefits of LMS sufficient to offset its costs.
Once an LMS option is introduced, that cost differential should increase over time. The reason is that a newly-introduced option at rate differentials reflecting the cost differentials in, say, the first year of its availability will appeal only to the very low-volume callers on-peak. As they start dropping flat rate service, the remaining subscribers will be the ones for whom the variable costs of usage constitute an above-average percentage of the total cost. The process of attrition of the flat rate service group should be progressive; as the differential between the two rates is adjusted to the now-higher average usage by the remaining subscribers to that service, more relatively low-volume users should transfer to LMS, further increasing the ratio of traffic-sensitive to non-traffic-sensitive costs of those remaining on flat-rate, and therefore further widening the cost-justified difference between the two charges.

In the final analysis, however, the efficiency of LMS is an empirical question: do the gains exceed the costs? The few attempts to answer it have produced divergent—but, probably more significantly, quite modest—results. The recent conclusions reached by Rolla Park and Bridger Mitchell are particularly arresting. After presenting the results of complex estimations that show LMS offering very modest welfare improvements over flat rate pricing, they nevertheless conclude that it is unlikely in practice to improve efficiency in most circumstances. The reason for their negative conclusion is the administrative infeasibility of pricing local calls in a manner even remotely approximating the ideal. Economically ideal prices, they point out, would vary from one telephone exchange to another, from one moment to the next, depending on the instantaneous relationship between demand and capacity in each; such a system, they find, would produce significant welfare gains. The only administratively feasible pricing system, however, would, they assume, involve uniform prices, differentiating at most three periods—perhaps peak, shoulder peak, and off-peak—within each of which the relationship between demand and

110. Bridger Mitchell's 1978 study produced estimates of welfare changes ranging from -1.6% to 6.0% consequent on switching residential customers from flat-rate to LMS under a variety of cost assumptions. The percentages are the estimated changes in the sum total of consumer and producer surplus from local telephone service—a total representing the difference between the value of that service to consumers and its cost to producers. Mitchell, Optimal Pricing of Local Telephone Service, 68 AM. ECON. REV. 517 (1978). For an explanation of the concept, see supra text accompanying notes 38 and 39.

Lewis Perl in 1985 estimated that a similar switch for both residential and business customers in Kentucky, with local rates set so as to maximize welfare, would produce a welfare improvement of approximately 0.11%; this included the benefit of a slight increase in subscription to telephone service resulting from the lower price for access alone entailed in the offer of LMS. The modesty of the improvement is perhaps more clearly suggested by the fact that Perl estimated it at 1.0 million dollars per year (for Kentucky), whereas he estimated that the welfare gain for Kentucky from reducing toll rates to marginal cost would be 44.6 million dollars per year. L. Perl, LMS in Kentucky, supra note 21.
capacity would in fact vary widely. Those prices, they judge, would cause so much uneconomic discouragement of calling during those times falling within the designated peak period when—and at those exchanges where—in fact marginal opportunity or congestion costs are very small, and uneconomic encouragement of calling at “off-peak” times when in fact marginal costs are high, as on balance—taking into account also the costs of administering such a system—to do more harm than good.111

None of these studies provides conclusive evidence of the efficiency of local measured service.112 One problem is the sketchiness of our knowledge of cost and demand conditions. Estimates of the marginal cost of local calls by time of day have varied widely. There is even greater uncertainty on the demand side. As we have already observed, the price elasticity of demand for local calls is low, perhaps in the neighborhood of -0.2.113 This implies that any efficiency gain from LMS depends heavily on customers' willingness to substitute off-peak for peak-period local calls, about which even less is known. More empirical attention must be devoted to local call price elasticities, the cost of local calls, and the cost and feasibility of administering more complex pricing systems. In the interim, companies and commissions have to decide how aggressively they should pursue the offer of LMS.

Even if we can at this time be only agnostic about the efficiency advantages of LMS, in our view its other numerous benefits114 tip the balance in favor of making it widely available in exchanges equipped with elec-

111. R. PARK & B. MITCHELL, OPTIMAL PEAK-LOAD PRICING FOR LOCAL TELEPHONE CALLS 23-31 (1986) (published by the Rand Corporation). In addition, they cite the fact that the costs of incremental capacity have declined with the introduction of digital electronics and fiber-optic transmission—in effect alluding to the first determining factor we have identified: the lower the marginal costs of usage, the smaller the differences in the costs of providing different customers with the flat rate bundle.

112. The early Mitchell study assumed that local calls would be priced uniformly, Mitchell, supra note 109, at 518, yet the major gain from measured service is likely to be from differentiating the price of local calls according to whether they are made in the peak period, when costs are much higher. The Park & Mitchell study, supra note 111, ignores the second-best gains in other service markets permitted by LMS when strict marginal cost pricing would not provide the company with adequate revenues and regulators insist on inefficiently low basic service charges. Also, it was limited to a small exchange served by a single office. Yet many “local” calls are indistinguishable from long-distance calls. See supra note 45 and accompanying text. The demand for these longer-distance “local” calls tends to be more elastic, and the calls themselves more costly to provide; both of these factors would cause the Park & Mitchell analysis to underestimate the benefits of LMS in a more typical setting. For discussions of this and other considerations that may have led Park and Mitchell to an unwarrantedly pessimistic assessment of LMS, see M. Koschat, D. Lehman & E. Sieff, A Preliminary Analysis of the Rand LMS Study (Nov. 28, 1986) (unpublished manuscript available from Bell Communications Research); W. Shew, Pricing Local Calls: How Much Imperfection is Perfect? (speech given at Telecommunications in a Competitive Environment, sponsored by National Economic Research Associates, Phoenix, Arizona, March 6, 1987) (unpublished manuscript on file with the author).

113. See supra note 40.

114. See supra text accompanying note 105.
tronic switches. An appealing idea may be to offer customers a choice between flat- and measured-rate service, with the measured-rate prices reflecting the additional administrative costs. This shifts the responsibility for determining whether measured service is worthwhile to the customer, who may be in a better position to make the assessment.\footnote{115} It has the further advantage of preserving the option of flat-rate service for those customers who, quite apart from price considerations, regard it as superior, perhaps because it eliminates all uncertainty about the size of the local service portion of their bills, perhaps because of the attractiveness, however illusory, of "free" local calls.\footnote{116}

2. \textit{Disaggregation by De-averaging: Variations in Marginal Access Costs}

In principle, every single transaction or taking of service by one customer at one moment in time will have its unique marginal cost. To call attention to that fact is to recognize the inevitability of grouping customers, transactions and time periods, and pricing these groups of transactions at average marginal costs. The illustration of this phenomenon that we pursue here is the ubiquitous presence in communications of spare subscriber access capacity.\footnote{117}

There are two polar occasions for this phenomenon. The first is when the spare capacity is in fact efficient, because its incremental benefits exceed its incremental costs—that is to say, because it ensures the optimal quality and reliability of service at minimum cost. In this event, the mar-

\footnote{115. If customers are to make an efficient choice, the measured rate prices must reflect the additional administrative costs of providing that option. But some portion of those costs might be fixed, varying with neither the number of subscribers to local measured service nor the number of local calls they make. The presence of these fixed costs means that, even if measured services are considered an option, the regulator must still form some \textit{a priori} estimate of whether it would be worthwhile, and make some sort of "collective decision" about whether it should be offered.}

\footnote{116. More people opt for flat-rate service than can be readily explained in monetary terms; some flat-rate customers persistently make so few calls that their telephone bills would be lower under measured-rate service. One possible reason, in addition to the ones we have mentioned, may be simple ignorance.}

\footnote{117. The principles we discuss here are equally applicable to other kinds of spare capacity as well. We concentrate on the subscriber access plant because that is the locus of the most intensely contested issue. It has been in opposition to proposed increases in the basic residential access charge that the asserted difference between marginal and average costs has been most widely cited.}
ginal costs of the spare capacity should unequivocally be recovered in price.

This case embraces the consideration we have already mentioned, that the minimum-cost expansion path will involve introducing additional capacity in lumps. In this event, short-run marginal costs will vary from below to above long-run cost, from the time when the new block of capacity is installed to the time when the next one comes into service.

Whether it will be feasible and desirable to recognize those fluctuations in prices will depend heavily on their amplitude and duration. If the cycles are short and, in addition, vary from one part of the system to the other, pricing on the basis of long-run incremental costs is likely to represent the most efficient feasible course: The superior efficiency of pricing at short-run marginal costs will have been outweighed by the costs (including inconvenience and confusion) of disaggregation. If, on the other hand, the cycles are likely to be protracted, it might be extremely inefficient not to reflect them in price. Consider, for example, the practice of telephone companies installing extra wire pairs in the distribution cable terminating at subscribers' premises, precisely because the marginal cost of installing extra wire pairs is very low, in order to accommodate future demand. The general practice of pricing the second line at the same rate as the first seems a likely source of considerable inefficiency, because on the one hand its marginal cost is very low, and on the other, the demand for a second line is probably much more elastic than for the first.

The other occasion for the presence of spare capacity will be where capacity is genuinely excessive. This could be because of mistakes—for example, overestimation of demand—or because of the distorted incentive of a utility company to "gold plate" its facilities when it is regulated on a rate base/rate of return basis. Whether or not the companies should be entitled to recover those costs, they manifestly would not be reflected in the first-best efficient prices of the affected services.

The importance of the phenomenon of spare access capacity will depend, among other things, on the extent to which stranded capacity can be placed at the disposal of other customers. This fungibility is likely to vary substantially from one time and place to another. Conceivably, one of us has suggested elsewhere, phone lines abandoned in center cities, with high vacancy rates, may have no such alternative possible employments;

118. See supra text accompanying note 85.
119. We are indebted to Roger Noll for this suggestion.
120. See supra note 87.
121. Kahn, supra note 8, at 147; see also Kahn, The Next Steps in Telecommunications Regulation and Research, PUB. UTIL. FORT., July 19, 1984, at 16-17 (raising this and other problems of aggregation).
their marginal opportunity costs and first-best prices would therefore be low. The same might be true of access lines planned for large amounts of Centrex service whose customers have been lost because they have installed their own PBXs.

In new residential or business centers, in contrast, marginal costs are likely to be much higher, reflecting the lesser density of subscribers, the greater average length of distribution lines, and the effects of inflation over the last decade. The same might be true in cities where existing conduits are already congested, and wherever new cable must be placed underground.

In principle, these differences in access costs suggest the possible desirability of disaggregating charges. For example, it would apparently be much more efficient to impose higher flat monthly charges for a subscriber's first telephone line than for the second line. And regulators might well permit telephone companies to charge a lower price for service in areas expected to have excess capacity for a long time.

None of these considerations constitutes a case for imposing non-traffic-sensitive costs on usage: they argue only for varying the flat access charges among different categories of customers.

D. Externalities

The simplest and most familiar case for pricing residential access or basic service well below marginal cost is that subscription to telephone service yields benefits to others. When an individual connects to the network, it increases the value of the service to others, because it increases the number of people they can reach by phone. In deciding whether to take phone service, however, a consumer tends to ignore the benefit his or her subscription confers on others. So, from a social perspective, if service is priced at marginal cost too few people will subscribe.

The theoretical basis for subsidy in this situation is straightforward. Economic efficiency requires that the marginal value of each service be equated to its marginal cost. In the absence of externalities, price must therefore equal marginal cost: in deciding whether to subscribe, the customer will compare the price with the value to himself. But when (for example) telephone subscription confers benefits on others, its social marginal benefit is the sum of (a) the value of the service to the marginal subscriber and (b) the value to other subscribers of that individual being connected to the network. A price set at marginal cost will discourage some subscriptions whose total benefits—to the subscriber and
others—would have exceeded the cost. The economically efficient price would therefore have to be below marginal cost.\footnote{122}

Consumer advocates generally contend that the external benefits are quite large, and so justify pricing basic telephone service well below marginal cost. The assertion is usually based on the observation that, since telephone service would be worthless if only one individual subscribed to it, the external value of a second subscriber must be large. What is relevant to pricing, however, is the external value not of a second or third subscriber, but of the marginal subscriber when penetration is, rather, eighty-five or ninety percent.\footnote{123}

It is simply not clear on \textit{a priori} grounds how the large increase in telephone penetration over the last half century has affected the external value of adding a marginal subscriber to the network. On the one side, it would tend to increase as subscribership approached universality, because the larger the body of existing subscribers, the greater the number of beneficiaries. On the other hand, it could be that, as subscription has exceeded 90\%, marginal subscribers have come to consist disproportionately of people relatively isolated from society generally, to whose hypothetical addition to the network existing subscribers would impute progressively smaller values.

It is possible to extract from Perl’s study of the residential demand for access an estimate that, in an exchange comprising 10,000 households, the external value of another subscriber is $2.60 per month when penetration is eighty-five percent, and rises to $3.00 per month at a penetration of ninety-five percent.\footnote{124} These estimates should of course be accepted, at best, as suggesting only rough orders of magnitude. At the same time they imply that the marginal externality benefits of even the relatively small

\footnote{122. The question of how far below is a complicated one. At a first approximation, it would presumably be sufficiently below to keep on the system all those for whom the private plus the external benefits of subscription exceed the marginal cost. A basic charge set uniformly at that level would, however, be likely to encourage subscriptions that do not meet that test, unless all potential subscribers would confer equal benefits on others. In principle, therefore, the first-best solution might require discrimination in the basic charge.

In any event, the market failure consequent on setting the charge at marginal cost will undoubtedly be remedied to some extent by voluntary private arrangements, diminishing the need therefore actually to reduce the price: potential recipients of the external benefits—parents of impecunious children, for example—can to some extent be counted upon to defray part or all of the cost of subscribing on behalf of people they want to be able to reach or be reached by.

123. The demand for basic service being highly price-inelastic, it appears that even a doubling of the basic charge would still leave penetration rates within that range. L. Perl, Residential Demand for Telephone Service 2 (Dec. 16, 1983) (unpublished study on file with the authors).

124. \textit{Id.} at 30-36. Using Census observations on whether individual households have telephones, Perl estimated the demand for basic service as a function of price, the number of subscribers in the local calling area, and demographic characteristics of the household. He found, as one would expect, that the probability of a household having a phone was higher (i.e., the value of service greater) the greater the number of other subscribers in the local area.
percentage of subscribers who would drop off the system if basic service charges were substantially increased cannot be ignored, even on grounds of economic efficiency alone—an inference that is reinforced, of course, if one takes into account the desirability on social, political, and income-distributional grounds of keeping telephone service widely affordable. In light of the tendency of some recent empirical investigations to produce estimates of the marginal or current costs of subscriber access on the order of only one-half or less of the more familiar twenty to thirty dollar estimates of associated revenue requirements, it appears that current charges may not on average be as far below economically first-best levels as has heretofore been widely believed.

Lest this observation be misconstrued, we must emphasize that, for reasons we will elaborate below in Part V, the economic case for much higher basic monthly charges remains extremely compelling. But as our consignment of the demonstration to that section suggests, the case may rest more on second-best considerations than on wide discrepancies between present charges and marginal costs, as modified by marginal externalities.

E. *The Proper Recognition of Accelerated Depreciation*

The increasingly competitive character of the telephone business has increased the rate at which assets depreciate, in economic terms. Under the previous regime, the monopolist could control the introduction of new technologies; typical depreciation rates based on estimated useful lives of 30 and 35 years could well have been accurate. In addition, the telephone companies presumably found it in their interest to accept such slow rates, because they entailed a correspondingly slower rate of decline (actually, a more rapid rate of increase) in their rate base, on which they generally earned something more than the cost of capital.

The introduction of competition and acceleration of innovation clearly require a more rapid depreciation of existing facilities, reflecting the more
rapid economic obsolescence of existing plant.\textsuperscript{127} Competition also changed the motivations of the telephone companies correspondingly. In these altered circumstances, the overvaluation of assets on their books\textsuperscript{128} became a source of intense concern, for two reasons. First, because regulators forced them to set their rates at levels incorporating capital charges on the inflated book values, the companies found themselves handicapped in competing with rivals using modern low-cost technology. Second, it raised questions about their ability ultimately to recover their sunk investments.\textsuperscript{128}

It is not clear whether or to what extent increased depreciation rates necessitated by the introduction of competition may properly be regarded as a real increase in economic costs rather than a belated recognition that those costs had previously been underestimated. The point is that investment and the true economic value of telephone company plant (the decline in which is the measure of economic depreciation) should always have been based on the cost of the most efficient available alternatives. So long as monopoly prevailed, neither regulators nor the companies themselves were forced to recognize that decline fully in realistic depreciation charges and rates.

On the other hand, monopoly may have retarded the process of innovation itself, and therefore the true economic depreciation of existing assets. Moreover, the process may well have been self-reinforcing or self-justifying: low depreciation rates tend to discourage regulated public utilities from replacing economically obsolete but still incompletely depreciated facilities, for fear that, once replaced, they will be dropped from the rate base; and the slower rate of innovation in turn could have lent apparent


\textsuperscript{128} The roughly 80% ratio of net to gross book value of telephone company assets contrasted with 60% to 70% in more competitive industries.

\textsuperscript{129} There have been some recent far-sighted agreements between Bell companies and state regulators that have mitigated the problem. Taking advantage of the sharp decline in the cost of capital and the impending reduction in corporate income tax rates, the companies accepted lower rates of return, the commissions refrained from demanding corresponding rate decreases, and the companies were permitted to use the windfalls to write down their overvalued assets—i.e., to make good a large portion of their accumulated depreciation reserve deficiencies. For examples in Maryland and Pennsylvania, see \textit{Telecomm. Rep.}, Sept. 22, 1986, at 11 (Maryland); \textit{Telecomm. Rep.}, Sept. 15, 1986, at 2 (Pennsylvania). If the companies would in any event have been entitled ultimately to recover their investments, customers will in these circumstances be rewarded for sacrificing present rate reductions by future reductions in the carrying charges on the diminished rate base. In other words, they will in effect earn a return on their foregone rate reductions equal to the return the companies would otherwise have been allowed on their equivalently larger rate bases.
justification to the long estimated service lives and low depreciation rates of existing equipment.\textsuperscript{130} The proper recognition of accelerated depreciation has two conceptually quite separate aspects. The first has to do with the inadequacy of past depreciation rates, which have left the telephone companies with assets grossly overvalued on their books.\textsuperscript{131} The question here is to what extent the companies are entitled to recovery of these sunk costs, and to what extent that recovery requires second-best pricing. If the companies have been regulated on an original cost or prudent investment basis, they can make a strong case that they are entitled to recover all of those costs, except to the extent it can be demonstrated that they were imprudent in incurring them. On the other hand, in view of the fact that the question is one of fairness to investors, it is necessary to ask also whether it is fair to permit full recovery of the sunk costs entirely from the diminishing residuum of captive ratepayers, whose demand alone may be sufficiently inelastic to permit it.

The second aspect relates to future depreciation, which is clearly part of long-run marginal costs,\textsuperscript{132} and therefore properly reflected in economically efficient prices. But in prices of which services? Specifically, are these higher costs properly incorporated in the charges for POTS? If they are the consequence of the introduction of competition, it would seem neither economically efficient nor equitable for them to be levied on the monopoly service. It would seem, rather, that they would constitute part of the incremental costs of providing the competitive services, and should be borne by them alone.

We have already enunciated the applicable principle: If the accelerated incorporation of new technology in new investments is economically efficient—or, even if inefficient, not reversible\textsuperscript{133}—then the marginal costs

\textsuperscript{130} See 1 A. KAHN, supra note 6, at 117-22, for a fuller explanation of the principle. For an assertion that it did in fact apply in the telephone case, see Irwin, \textit{The Telecommunications Industry}, in \textit{The Structure of American Industry} 274 (W. Adams ed., 7th ed. 1986).

There is some uncertainty about the factual premise underlying this deduction. According to \textit{AT&T Engineering Economy} 157 (3d ed. 1977), when an item of plant is retired, both the plant account and the accumulated depreciation account are reduced by its original cost. It also states that total assets equal total plant minus accumulated depreciation. If these statements properly describe the universal regulatory practice, retirement of incompletely depreciated plant would not, in fact, deprive companies of a return on the unrecovered portion of that investment. See also W. Bolter & D. Irwin, \textit{Depreciation Reform: A Crucial Step in Transforming Telecommunications to a Free Market} (1980) (commissioned by GTE) (arguing that both an accelerating rate of technological innovation and competition have made historic depreciation policies progressively unrealistic).

\textsuperscript{131} See supra note 74. The $25 billion estimated overvaluation is over and above the $7.3 billion of writeoffs that AT&T Communications took on its assets, mainly in customer premises equipment, after divestiture. \textit{Wrong Number: AT&T's Earnings Shocker and What It Means}, BARRON'S, Oct. 24, 1983, at 15-28.

\textsuperscript{132} If investment decisions are efficient, these costs will equal short-run marginal costs as well.

\textsuperscript{133} See supra note 92.
of the various services are what they are; the fact that people who want only POTS may not benefit from the new technology is irrelevant to the determination of the respective marginal costs imposed on the economy when they decide to subscribe to POTS or other more sophisticated services.

These principles are clearly easier to enunciate in *a priori* terms than to apply in practice. The efforts of the California Public Utilities Commission to assess the merits of Pacific Bell's requested represcription of depreciation rates and amortization of past undercollections illustrates the kind of quagmire of claims and counterclaims into which the attempt to apply them is likely to lead. We confess to having emerged from a reading of the record of that proceeding in a state of profound skepticism over the possibility of successfully adapting the traditional system of public utility regulation to the new environment of turbulent competition. The primary reason for deregulation is the compulsion that competition exerts on managements to engage in bold, risk-taking innovation. To subject such initiatives to regulatory second-guessing is to defeat that objective. And yet, to complete the dilemma, the case for regulation is the perceived need to protect from exploitation captive ratepayers who are inadequately protected by competition; and that surely requires regulatory determination of the reasonableness of the costs that companies propose to pass on to them.

The essential solution to this dilemma created by partial deregulation, it seems to us, is in one way or another to break the link between the prices to the monopoly customers and the revenues and costs ascribed to the competitive operations. As long as the regulated prices continue to be set, directly or indirectly, on the basis of total company costs and revenues, or on the basis of some continuing process of allocation of cost between regulated and unregulated operations, there will always be the danger, in principle, of the former subsidizing the latter. In those circumstances, conscientious regulators will not be able to refrain from setting floors under the competitive prices, as well as ceilings over the putatively monopolistic ones, and second guessing the companies' investment decisions, as they are doing in California today.

134. But on this factual question, see *supra* text accompanying notes 64 and 65.

135. *In re Application of Pacific Bell § I*, No. 85-03-078 (Aug. 21, 1985) (opinion on capital recovery/depreciation issues); see also *supra* note 65.

136. For a forceful exposition of the thesis that this clash between the imperatives of technology and of regulation ultimately dooms continued state regulation, see M. Irwin, *The Demise of State Telecommunications Regulation* 11 (Apr. 30, 1985) (presented at 14th Annual Telecommunications Policy Research Conference) (arguing that “state public utility commissions are today's castles; technology is the gunpowder”) (unpublished manuscript on file with the authors) (published in revised form as Irwin, *The Demise of State Telecommunications Regulation*, TELECOMM., Dec. 1986, at 70).

246
Pricing

V. The Applicable Principles: Second-Best

It appears that first-best prices of telephone services are unlikely to provide the revenues to which the telephone companies are entitled under traditional regulatory principles, for some combination of the following reasons. First, economies of scale may in important parts of the business cause marginal costs to fall short of average costs. Second, regulatorily permissible capital charges are based on book values of assets that far exceed current replacement costs reflecting the latest technology. Since the discrepancy between book values and the capital costs of reproducing these services with current technology is evidently enormous, it appears this is the preponderant factor. Finally, the external benefits of subscribership would call for pricing subscriber access below the marginal cost of providing it.

In view of the increasing uncertainty about the extent to which, or even whether, subscriber access charges (under LMS) or flat rate local service charges fall short of their marginal costs, the present gross overcharging of long-distance calling and evidently of local calling as well, under LMS, both strongly suggest that first-best prices would for these various reasons fail to produce sufficient revenues. A reflection of subscribership externalities in basic charges below marginal costs would further exacerbate the discrepancy.

A. Second-Best Pricing of Independent Products

Ignoring for the moment the interdependence of demands for telephone services, the first approximation to a resolution of this dilemma with minimum loss of economic efficiency is by now familiar in the regulatory literature; its purport for the pricing issues with which this paper is concerned is fundamental. The formula—named for its original expositor, Frank P.

137. See supra note 74 and accompanying text. We make no attempt in this Article to assess the prudence of the original outlays and therefore the extent to which the companies may not be entitled to recover them under traditional regulatory principles. To the extent, moreover, that the companies' current investment decisions indeed are inefficient, as critics maintain, and portions of the associated carrying charges are therefore disallowed, this too would reduce the size of the required markups above marginal costs.

138. See supra note 76.

139. See supra note 21.

140. One reason for this is that local usage charges are often not differentiated by time of day, and when they are, the off-peak price is typically set well above marginal cost, which is close to zero. According to L. Perl, LMS in Kentucky, supra note 21. LMS usage charges for peak and off-peak service in Kentucky were 2.7 and 1.5 cents per minute; the respective marginal costs were 0.837 and zero cents per minute.

141. That is to say, if we had firm reason to believe that the overcharging of usage and other services relative to their marginal costs were no more than necessary to offset an equivalent undercharging of subscriber access relative to its marginal cost, we could not dismiss the possibility that first-best pricing of all services would be consistent with total revenue requirements.
Ramsey—calls for markups above the marginal costs inversely proportional to the elasticities of demand for the several services, in order to elicit the requisite increase in total net revenues. Manifestly, the less elastic the demand for a service, the larger the markup above marginal cost it will accept while minimizing the consequent inefficient discouragement of consumption. Conversely, the greater the demand elasticity, the smaller the markup sufficient to produce proportionately equal damage.

The historic pattern of telephone pricing obviously conflicts violently with that prescription. It imposes the largest markup above marginal cost on long-distance calling, the service whose demand is generally believed to be the most elastic, and the smallest markup—quite conceivably, on average, a negative markup—on subscribership itself, the demand for which is the least elastic. Moreover, it recovers as much as forty percent of the total revenue requirements of subscriber plant costs from the ten percent or so of total calls represented by long-distance usage, the most demand-elastic service. It is this perverse pattern of markups, far more than the mere necessity for moving from first- to second-best, that explains the enormous welfare losses caused by the present price structure.

The standard formula for Ramsey pricing assumes a monopoly supplier. The competition in telecommunications markets is likely to alter the prices that satisfy the Ramsey principle. How it alters them will depend on whether regulation is confined to the incumbent firm or extended to the competitive entrants as well.

Limiting application of the Ramsey rule to the incumbent (or dominant) firm is, in principle, less efficient than comprehensive Ramsey

143. See supra note 40.
144. The figure in 1984 was 8.5%. Federal Communications Comm'n, Statistics of Communications Common Carriers table 7 (1985). Cf. note 105 supra (giving somewhat different estimate for minutes of calling in 1983).
145. We see no reason to doubt the showing of the statistics that, at the respective prices prevailing during the periods studied and largely continuing today, the demand for toll calling is markedly more elastic than for local calling. The present relative markups on these two services and on pure access therefore reverse the order required by Ramsey principles. It seems highly possible, however, that it is only at their present, economically indefensible respective rates that the demand for long distance calling is so much more elastic than for local. At rates more closely resembling their respective, apparently very low marginal costs, see supra notes 21 and 140, the two elasticities might not differ much.
146. For example, according to the estimates by L. Perl, supra note 38, at 15, uniform pricing at marginal cost would have produced an annual welfare gain of 5.5 billion dollars in 1984 dollars; moving to second-best, but with optimal markups, would still have produced a gain of 5.1 billion dollars as compared with the results produced by the prices actually in effect.
147. For a detailed discussion of these alternatives, see Braeutigam, Optimal Pricing with Intermodal Competition, Am. Econ. Rev. 38 (1979).
Pricing

When entrants are unregulated, their prices restrain the dominant firm's ability to resolve a revenue deficiency by large markups on services whose market demand is relatively inelastic. Since competition limits the ability of the firm to concentrate its markups in markets where the demand is relatively inelastic, larger markups become necessary in relatively elastic markets, where the consequent distortions are more costly. In contrast, comprehensive application of Ramsey pricing can efficiently require every firm to charge a relatively high price for services whose market demand is relatively inelastic. This does not mean that comprehensive Ramsey pricing is necessarily preferable, however, since its implementation requires more extensive information, and there may be—indeed, we believe there are in telecommunications—other, more compelling considerations that favor continuing to exempt entrants from price regulation.

If Ramsey pricing is limited to the incumbent's services, the markup of price over marginal cost in inelastic markets is reduced by competition, which increases the incumbent's elasticity of demand. The resulting reduction in the efficient price will probably be smaller if the entrants possess market power, because that will incline them to price above their own marginal costs as well.

In short, the effect of competition on Ramsey prices depends on whether entrants are also subjected to price regulation and whether they possess market power. For the most part, regulators have chosen not to regulate the prices of entrants in telecommunications, for good reasons. This regulatory abstention, in conjunction with the observed concentration of entry in long-distance markets, suggests that competition has had the effect of reducing the (second-best) efficient prices for interexchange calling and (since the deficit must be made up somewhere) increasing the Ramsey prices for services sold in less competitive markets. This reasoning reinforces our observation that the present pattern of pricing is extremely inefficient.

There are alternative systems of second-best pricing capable in principle of producing results superior to those produced by Ramsey principles. Ramsey pricing assigns one and only one price to each service, and

---

148. This can be seen most clearly by considering a multi-product monopoly confronting a perfectly inelastic demand curve in one market. If marginal cost pricing produced inadequate revenue, the monopoly could capture the additional necessary revenues by increasing its price for the service whose market demand is perfectly inelastic, thus avoiding any distortions in customers' choices. But competitive entry into that market will limit the extent to which the incumbent can profitably increase its price in that market, and thus may require it to increase its prices in other markets where (because market demand is not perfectly inelastic) pricing above marginal cost does induce distortions. This problem is asymmetrical in the sense that, if pricing services at marginal cost produced excessive revenues, competition would not prevent the incumbent from selling services for which demand is inelastic for well below marginal cost. See also Damus, Ramsey Pricing for U.S. Railroads: Can It Exist?, 18 J. TRANSPORT ECON. & POL. 51 (1984).
represent the most efficient solution consistent with that ground rule. A two-part tariff, in contrast, charges customers a fixed fee for the option of taking any service at all, and an additional, separate charge for each unit actually purchased. A multi-part tariff introduces additional steps, pricing different blocks of usage at different levels: declining block rates for electricity are a familiar example. So, if first-best rates for telephone service produced inadequate revenue, one could price some initial block of calls well above marginal cost, with the rates for additional blocks descending progressively to that first-best level.

The reason for the potential superiority of these more complex systems is that the additional prices allow marginal consumption to be priced closer to marginal cost than Ramsey principles would permit, with the revenue deficit made up by the inframarginal price—the fixed fee or the price for the initial usage block or blocks. Thus, the problem of insufficient (or excess) revenue can be dealt with more efficiently by concentrating the necessary departures of price from marginal cost where they are less likely to affect demand: a service covered by the fixed charge or the price for some initial block of usage is likely to be demanded in any event.

A two-part tariff might indeed achieve first-best results. This would be the outcome, for example, of pricing toll calls at marginal cost and using the lump sum charge to make up the revenue deficit—but only in the unlikely event that the fixed fee did not discourage some customers from making long-distance calls altogether.

If, as seems more likely, the best a two-part tariff could produce is second-best pricing, some additional benefit could probably be achieved by introducing multi-part usage rates. Suppose, for example, that the marginal cost of long-distance calls is five cents, and the price under the best achievable two-part tariff is ten cents. If the largest customer demands one hundred calls at that price, the telephone company could safely create a three-part tariff by introducing a price slightly below ten cents for calls in excess of one hundred per month. The lower price would cause the largest customer (and probably others as well) to increase their calling; and the additional calls, all priced above marginal cost, would bring in additional profits, which could be used to reduce the prices for the preceding usage blocks closer to marginal cost. Clearly a good deal of welfare-enhancing consumption, inefficiently discouraged by Ramsey prices and simple two-part tariffs, could in principle be elicited in this way, and second-best would come closer to first-best.

In practice, however, these more sophisticated pricing schemes could be less efficient than Ramsey pricing. They require substantially more
information to calculate the second-best price levels.\textsuperscript{149} Which system—Ramsey pricing or multi-part tariffs—would be superior on balance is therefore ambiguous. This uncertainty does not, however, justify the scant attention that has been paid to the possible use of multi-part tariffs for telecommunications services.

If multi-part tariffs were to prove superior to Ramsey prices, it would not change in any fundamental way our conclusions about the directions in which the current telecommunications price structure should be reformed. It would still almost certainly be imperative to reduce toll and local usage charges from their current, egregiously inflated levels and to raise access prices.\textsuperscript{160} This strong conclusion must confront three possible reservations, to which we turn, all of which suggest the possible desirability of pricing subscriber access below its marginal costs.

**B. Second-Best Pricing of Complementary Products**

The Ramsey inverse elasticity rule does not apply without revision to situations in which the demands for two or more of the services in question are complementary,\textsuperscript{181} as of course are the demands for mere access to the telephone office, on the one side, and for local and long distance calling, on the other. This consideration was presented to one of us, naively but perceptively, by a regulator, who asked how we reconciled our advocacy of raising the basic monthly charge to recover the full non-traffic-sensitive costs of subscriber access with the practice by most shopping malls of not charging people for parking their cars, choosing instead to recover the costs of providing that facility in the prices of the goods and services purchased in the stores.

There are indeed circumstances when first-best pricing is not feasible and some of the products or services are complementary, under which the second-best alternative is to price below marginal cost the one that gives

\textsuperscript{149} Formulating Ramsey prices requires information on characteristics of market demands. Multi-part tariffs require as well information on how demands for telephone services vary among individuals.

\textsuperscript{150} A multi-part tariff might, however, involve not only much greater reductions than Ramsey prices in the high usage blocks, but higher prices for the initial blocks of calls, if demand for them were sufficiently inelastic—to such a degree as to justify a lower price for access as well. In this way, these tariffs might extract from the initial usage blocks some of the revenues that Ramsey prices would obtain from the access charge. On the other hand, the demands of low volume long-distance callers, whose marginal rates would fall within the initial, high-priced blocks, are likely in our view to be more elastic than the demand for subscriber access generally.

Manifestly, there is insufficient information available to permit a conclusion about how great a difference there would be in the price structures under these two pricing schemes. One promising application of the multi-part concept would be in the pricing of local service: subscribers could be offered a choice of LMS schemes, one with a low access charge and a high price for an initial block of calls, others with the opposite configuration.

\textsuperscript{151} See Baumol & Bradford, supra note 142.
access to the others and recover the revenue deficiencies on the basis of usage or purchases of the others. The logic is not difficult to grasp intuitively. The goal of second-best, when marginal cost pricing produces inadequate revenue, is to minimize the uneconomic discouragement of consumption resulting from prices above first-best levels, by discriminating among customers on the basis of their willingness to pay. Where services A and B are complementary, it becomes necessary to consider the effect of the markups on each on purchases of the other—that is, their cross-elasticities of demand.

Specifically, pricing access to a shopping mall (or amusement park) at or above marginal cost may discourage purchases that are worth more to customers than their combined marginal cost, including the marginal cost of access. And it is possible that pricing the goods or services in the shops (or the features at the amusement park) above their marginal costs, while providing entry or parking below marginal cost, would cause less of a distortion. Here, as in the simpler Ramsey case of independent demands, the solution is to exact the necessary revenue contribution above marginal cost from customers in proportion to their willingness to pay—in this case in proportion to the value each customer places on the combined service. This value can be measured by the extent to which the customer is willing to pay separately for actually using the facilities—admission to the individual attractions, or purchases at the stores.

Whether this reasoning justifies the present system of pricing subscriber access and usage is an empirical question, whose answer depends on the relevant elasticities and cross-elasticities of demand for access and for use. Specifically, usage would have to be comparatively highly sensitive to the admission fee—the basic monthly charge—so that setting the latter at or above marginal cost would do more economic damage than overpricing the former. While the respective elasticities in the telephone case might well have justified low basic monthly charges in the past, in order to encourage widespread subscribership and use of the system, it seems highly unlikely that they continue to do so today. Even large increases in the basic charge today are unlikely to cause more than a small percentage of subscribers to drop service; moreover, as we will point out, most of that attrition can be prevented at far lower cost than is entailed by the present arrangement.

153. Observe the similarity of this practice to the use of tie-ins as a counting device, to which we have already referred, supra at note 32.
154. For a similar judgment to the effect that, telephone service having become something close to a necessity of modern living, the historic economic benefits of pricing subscribership below marginal costs are now far outweighed by the costs of overpricing usage, see A. Phillips, Welfare Aspects of Telecommunications Services: Externalities and Related Issues (June 1, 1985) (unpublished...
C. Subscribership Externalities Revisited

Although in a first-best world the external benefits of subscribership would call for a basic monthly charge below marginal cost, this would, in a second-best world, produce inefficiencies far grosser than the one it is intended to correct. Suppose those benefits were indeed worth three dollars a month per marginal subscriber. First-best pricing would—roughly\(^1\)—require a monthly service charge to all hundred million or so subscribers three dollars below marginal cost and, it appears, much further below revenue requirements, in order to make certain that not a single subscriber discontinued service the total marginal benefit of whose remaining on the system equalled or exceeded marginal cost. This alone would produce a revenue deficiency of about 3.5 billion dollars annually, which would have to be recovered from much more highly demand-elastic usage, at an immense net loss in social welfare.

The inefficiency of such a “solution” is further highlighted by two additional considerations. First, losses of external benefits consequent on pricing access at or above marginal cost would not be large in the aggregate, because relatively few subscribers would in fact drop the service in response to even quite large increases in the flat monthly charge. And, second, there are far less costly ways of keeping those relatively few subscribers on the system—for example, direct subsidies or subsidized rates targeted at low-income subscribers.

D. Income-Distributional Considerations.

Whether either government or the telephone company has some responsibility to hold down the basic residential charge is unlikely to be decided on the basis exclusively of economic efficiency considerations. Instead, it will continue to be made very largely on the basis of more broadly political considerations—in both the best and the most vulgar sense.

Economists universally recognize that moving from one set of prices to another will have different effects on different people, and they accept the relevance of those income-distributional consequences in deciding whether such a change is socially desirable. Our colleague, Lewis Perl, has attempted to estimate the effects of moving toward marginal cost pricing of telephone services on households with various levels of income.\(^2\) He found that while it would produce very large social welfare gains, in the

---

155. In fact, a non-discriminatory charge three dollars below marginal cost would err in the opposite direction, for the reason we have already given: it would encourage some subscriptions the marginal benefits of which (internal and external) fell short of marginal cost. See supra note 122.

156. See L. Perl, supra note 38.
aggregate, the lowest third of the population in income would actually suffer a decline in welfare, and about eighty percent of the benefits would flow to the highest-income third. This was so for essentially two reasons. First, the principal upward correction of price would be a uniform increase in the basic monthly charge to all families, a burden inversely proportional to income. Second, the principal downward correction would be a sharp reduction in the present overcharging of long-distance calling. A large share of the present burden is borne by businesses and probably shifted to consumers, who bear the burden more or less in proportion to their incomes. The higher the income, therefore, the greater the benefit from its elimination.  

The possibility that more efficient prices will have unfortunate income distributional consequences does not, however, suffice to justify the present economically irrational price structure, and the enormous social welfare losses that it entails. It requires us, rather, to seek ways of forestalling or reversing those consequences that would be less costly than the present system. The answer is of course the same as the solution to the problem of externalities: direct subvention of the relatively small number of low-income subscribers, in order to cushion them from the injury they would suffer from efficient pricing.

157. The reader will observe that we do not include, as a third reason, that higher-income families spend a larger proportion of their income on such calls than do families in more modest circumstances. While this is widely believed to be the case, in fact the evidence is by no means clear. See the various estimates of income elasticities in L. Taylor, supra note 40, at 169-74. The AT&T Long Line FIRM Model, whose results were not available at the time Taylor’s book was published, estimates an income elasticity of 0.63 for interstate MTS messages in 48 states and Washington, D.C. This means that as income rises, expenditures on long-distance calling rise less than proportionally. The following statistics of intrastate toll calling in Oklahoma by families of different income levels support this finding:

<table>
<thead>
<tr>
<th>Annual Income</th>
<th>Intrastate Call Minutes Per Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - $10,000</td>
<td>37.4</td>
</tr>
<tr>
<td>$10,000 - $20,000</td>
<td>47.9</td>
</tr>
<tr>
<td>$20,000 - $30,000</td>
<td>66.1</td>
</tr>
<tr>
<td>$30,000 - $40,000</td>
<td>57.4</td>
</tr>
<tr>
<td>Over $40,000</td>
<td>61.8</td>
</tr>
<tr>
<td>Average</td>
<td>49.9</td>
</tr>
</tbody>
</table>

Figures provided by Southwestern Bell Telephone Company for the year 1982.

These figures suggest that families of progressively higher income levels actually spend a declining proportion of their income on intrastate toll calls.

Similarly, Satellite Business Systems (which was acquired by IBM in 1985), in its submission to the Federal Communications Commission in Docket No. 78-72, observed that the average monthly bills of the rural subscribers served by telephone companies receiving assistance from the Rural Electrification Administration ranged from $20.41 in North Carolina to $58.85 in Alaska; the flat rates for local exchange service fell in the six to eight dollar range. The difference between these figures, it observed, demonstrates that people in rural areas, often of very modest means, make many inter-exchange calls, precisely because of their comparative isolation.

254
Pricing

The total subsidy would not have to be large; it would be required for only a small percentage of households. Since it is highly unlikely to be forthcoming out of general revenues, what would be the least inefficient way of raising it within the industry? As the logic of the foregoing discussion clearly establishes, by far the (second-) best solution would be a slightly elevated basic monthly service charge to all the other families and/or business subscribers.

Conclusion

The emergence of competition in telecommunications markets has led to strident disagreements about the prices that continue to be regulated. Some disagreements are over the applicable pricing principles; but many turn out, on close examination, to involve instead contentions about the nature of telephone company costs, the range of services the companies offer, and whether the companies are minimizing their costs, particularly of “basic service.” The only measures of cost germane to efficient prices are the marginal costs that the company will actually incur in supplying its full panoply of services, and its aggregate revenue requirement. Accurate estimates of these costs are vital, and armchair empiricism is no substitute for engineering and statistical studies. Minimizing cost and offering customers a suitable array of services and service qualities are important objectives, but they are independent of the issue of how services should, in principle, be priced.

The introduction of extraneous issues into these controversies is particularly unfortunate because the pricing issues themselves are already sufficiently complex; fallacious statements of economic principle only confuse matters more. It is only in the simplest of circumstances that it is efficient to price at marginal cost. The rule does not apply to telecommunications markets for three reasons: the “network externality,” the likelihood that marginal cost pricing would not provide the firm with adequate revenue, and the administrative cost of literally pricing every service at its marginal cost. When a departure from marginal cost pricing is necessitated by the company’s revenue requirement, the objective becomes one of finding the set of prices that minimizes the resulting distortions, either through Ramsey prices or multi-part tariffs. The network externality, by itself, makes it efficient to depart from marginal cost pricing in order to provide individuals with suitable encouragement to connect to the network. And since the costs of many telecommunications services vary from one moment to the next and also from one customer to another, the administrative costs of strict marginal cost pricing would be prohibitive.

Balancing these considerations is not easy. We do, however, have a clear idea of the directions in which prices should be moved. Whatever the
historic justification for the system of pricing still in effect today, it has long since disappeared. Its social cost today is to be reckoned not merely in terms of a multi-billion dollar annual static welfare loss, but, perhaps even more important, in the ways in which it has discouraged the exploitation of one of our most dynamic, versatile technologies. The residual need served by it can be achieved in other ways at only a small fraction of that cost. The former Chairman of the Federal Communications Commission and his colleagues have provided us with an apt summary of our argument:

The public switched network is a critical national resource that has been underutilized because of inefficient pricing. . . . [T]he key objective for public interest regulation in the transitional marketplace should be to stimulate use of the public switched network to the efficient levels that would be attained in a competitive marketplace. 158

The economic principles we have elucidated point the way to achieving that objective.

158. Fowler, Halprin & Schlichting, supra note 8, at 166.