

Laurence J. Kotlikoff

Chairman and Professor of Economics, Boston University

January 21, 2003

Natural Monopoly and the Definition of “Impairment”¹

This memo describes and provides an economically sound and administratively workable definition of impairment. The definition melds the principles set forth in the Circuit Court’s *USTA* decision² and Dr. Robert Willig’s recent discussion of the meaning of impairment, which highlights the importance of applying established economic theory and legal principles of competition to the Commission’s analysis.³

Background

The goal of the Telecommunications Act of 1996 is to “promote competition and reduce regulation in order to secure lower prices and higher quality services for American telecommunication consumers and encourage the rapid deployment of new telecommunications technologies.”⁴

The Act specifies that in determining what network elements should be made subject to unbundling, the Commission “shall consider, at a minimum, whether (A) access to such network elements as are proprietary in nature is necessary; and (B) the failure to provide access to such network elements would impair the ability of the telecommunications carrier seeking access to provide *the services that it seeks to offer*.”⁵ [emphasis added]

In the *USTA* decision, the Circuit Court held that impairment refers to higher costs faced by competitors. Indeed, the decision states that “... any cognizable competitive ‘impairment’ would necessarily be traceable to some kind of disparity in cost.”⁶ The Court also pointed out that “average unit costs are necessarily higher at the outset for any new entrant into virtually any business” and that suitable impairment cost criteria should be “linked (in some degree)” to characteristics of a natural monopoly in which a single provider enjoys “economies of scale over the entire extent of the market.”⁷ The Court further stated that “The classic case where competitor duplication would make no

¹ This research has been funded by AT&T. The views expressed are my own, and may not always reflect those of AT&T.

² *United States Telecom Ass’n v. FCC*, 290 F.3d 415 (D.C. Cir. 2002).

³ Robert D. Willig, “Determining ‘Impairment’ Using the Horizontal Merger Guidelines Entry Analysis,” attachment to ex parte letter from C. Frederick Beckner III, Sidley Austin Brown & Wood, to Marlene H. Dortch, dated November 14, 2002.

⁴ Pub. L. 104-104, 110 Stat. 56, codified at 47 U.S.C. § 151 et seq.

⁵ 47 U.S.C. § 252(d)(2).

⁶ 290 F.3d at 426.

⁷ *Id.* at 427.

economic sense is where average costs are declining throughout the range of the relevant market.”⁸

In addition, the Court indicated that, whatever cost criteria are used to determine impairment, such criteria should be assessed on a market-specific basis. Thus, it criticized the FCC for choosing “to adopt a uniform national rule, mandating the element’s unbundling in every geographic market and customer class, without regard to the state of competitive impairment in any particular market.”⁹ Moreover, the Court stated that “To rely on cost disparities that are universal as between new entrants and incumbents in any industry is to invoke a concept too broad ... to be reasonably linked to the purpose of the Act’s unbundling provisions.”¹⁰

The FCC is now reconsidering its unbundling rules in light of the *USTA* decision. In doing so, it is important that the Commission’s revised rules comport closely with the economic foundations of the unbundling requirements and how they relate to both the *USTA* Court’s decision and other existing court-accepted principles of competition regulation.

Organization and Overview of Findings

I proceed by presenting the standard economic definition of “natural monopoly” and show that this definition cannot be meaningfully applied by focusing solely on the individual network elements that are used as inputs to finished services. Rather, I show that for the natural monopoly concept to be useful in identifying impairment faced by CLECs (in providing the finished services they seek to offer), natural monopoly must be evaluated *both* with respect to individual elements *as well as* to the production process used to combine these elements into finished services.

The principles that can be applied to operationalize this determination of impairment can be found in established antitrust law and economics. Applying these principles appropriately requires a two-part test. The first would determine whether the provision or delivery of one or more telecommunication services exhibits sufficient characteristics of natural monopoly in the relevant market, including persistent economies of scale and high fixed costs, such that it would be socially inefficient for another party to produce that service or those services using its own facilities. The second would determine whether, under the circumstances that a substitute for an incumbent element could be used to provide a finished *service*, it is cost efficient and otherwise practicable for a competitor to do so. I propose an impairment standard below that applies these dual criteria and, in particular, relies upon Dr. Willig’s discussion of the Department of Justice’s *Horizontal Merger Guidelines* in assessing the latter set of issues.

A key determinant of whether governmental intervention is appropriate and socially beneficial is whether the ILEC retains market power in the relevant product and

⁸ *Id.* at 426.

⁹ *Id.* at 422.

¹⁰ *Id.* at 427.

geographic market. In this regard, I would note that the mere fact that there are one or two competitors that use their own facilities to provide a given telecom service does not suffice to eliminate real economic concerns about the exercise of market power. The reason is that it ignores the possibility that the incumbents, because of their small number, can collude to exercise such power. Hence, my proposed definition of impairment applies established economic and antitrust principles to assure that incumbents cannot exercise market power so as to frustrate the Act's aim to "secure lower prices and higher quality services for American telecommunication consumers." Such concerns about market power, whether it be exercised by monopolies, duopolies, or oligopolies, lies at the heart of Dr. Willig's recommendation that the FCC's new impairment standard make use of the Justice Department's and FTC's horizontal merger guidelines.¹¹

The Economics of Natural Monopoly

The standard definition of a natural monopoly relates to the average costs involved in producing a particular product or combination of products. If costs decline over the range of possible demand in a given market, then a single large firm will be able to produce a given quantity of output at a lower total cost than two or more smaller firms. This typically occurs either because (a) it is possible to produce or otherwise secure the inputs to the production process more efficiently when larger quantities of these inputs are demanded to satisfy higher demand for the final good or (b) because larger scale production allows the dominant supplier to use more efficient techniques for combining these inputs into final goods.

The first source of natural monopoly may arise when there are significant scale economies in the production or acquisition of the inputs to a telecommunications service.

A classic example is loops. The ability to serve all of a neighborhood's demand out of a single cable in a single trench causes significant drops in the average cost of producing a loop as the total number of loops served in a neighborhood rises. Thus, if there are scale economies in the provision of a particular UNE input, a CLEC's inability to purchase this UNE from the largest, most efficient producer (i.e., the ILEC) will force the CLEC to incur higher costs than the ILEC over the entire range of demand in the provision of the final telecom service, and the CLEC will be impaired in its offering of this final telecom service.

The second source of natural monopoly arises when a large producer of telecom services can use a production process for combining inputs that is significantly more efficient than the most efficient alternative process available to smaller competitors.

¹¹ The danger of unregulated duopolists and oligopolists is that they can collude to reduce supply and raise prices. Indeed, the most likely outcome of such collusion is the monopoly level of price and output. Alternatively, if the firms compete at all, they may well do so in a manner that leads to a) consumer prices far in excess of marginal cost, b) significant economic distortions, and c) major losses in consumer welfare.

For example, because of an ILEC's scale advantages in producing loops, it can terminate very large numbers of loops at its wire centers and then combine these loops with switching inputs by simply running a short, cheap jumper pair between the line-side and the switch-side of its main distribution frame. However, because new entrants operate at a lower scale, they cannot economically build their own loop network, but must lease these inputs from the ILEC. However, the ILEC's loops all terminate at its wire center location, not at the wire center where the CLEC's switch is located. Thus, in order to combine the leased loops with its own switching to provide service, the CLEC must (1) establish a collocation at the ILEC wire center serving each of its customers, (2) install digital loop carrier and related transmission equipment in this collocation, (3) incur both the ILEC's charges and its own internal costs to complete a hot cut of each of its customers' loops, and (4) backhaul these loops to the distant location where it maintains its switch. Because this hot cut/collocation/backhaul technology is not nearly as efficient in absolute terms as the short jumper pair technology, the CLEC is impaired relative to the ILEC in combining loop inputs with switching inputs for such customers.

Regardless of whether the source of the natural monopoly in the provision or delivery of a particular telecom service or services is related to 1) the production or acquisition of particular inputs, 2) the production technology governing the combined use of inputs, or 3) both factors, the cost of providing the service(s) will decline over the demand range of the relevant market. Thus, natural monopoly must be assessed based on the cost curve of the finished product, and not just the cost curves of providing or acquiring its individual inputs.

Implications for Telecommunications Services

Telecom services almost always require multiple inputs, especially switched services. These services use (at a minimum) loops, switches, transport, signaling and databases. Whether or not the provision of telecom services is a natural monopoly depends on the production and acquisition characteristics of each of their inputs *as well as* the available technologies for combining these inputs into finished telecom services. Thus, both factors must be assessed in determining whether a CLEC would be impaired without cost-based access to an individual UNE. If, however, the FCC were to narrowly review whether the average costs of procuring a single input declined over the range of market demand, it could – incorrectly -- eliminate access to a critical element at TELRIC-based rates even though that element cannot be replaced by the CLEC at comparable total costs in a manner that enables it to provide a finished service to retail users.

For example, CLECs have repeatedly demonstrated that (due to hot cuts, collocation and backhaul) they face substantially higher costs than incumbents in using non-ILEC switches to provide finished telecommunications services to customers served by voice-grade loops. If the Commission looked only at the costs of provisioning CLEC switches, and thus ignored the additional costs that the CLECs face in using such switches, this partial analysis could lead to a false conclusion that CLECs are not impaired without access to ILEC switches.

Competitive Supply of Inputs

The proper regulatory response to the presence of natural monopoly characteristics or to market power arising from implicit or explicit collusion is to ensure a) that prices are not manipulated by a restriction on output; and b) that whatever is produced is done at minimum cost. Allowing CLECs to use UNE-P to serve natural monopoly markets (e.g., analog line customers) achieves these ends. The retail competition it fosters keeps a lid on prices because entrants will be able to price their retail services based on input costs that reflect the incumbents' efficient costs, and these services will continue to be produced efficiently on the ILEC's network.

Is there an alternative to UNE-P that allows efficient self-provision by CLECs of one or more inputs/elements while continuing to use natural monopoly inputs such as loops or transport? The answer, in principle, is yes – but only if there is no loss of output or economic efficiency or higher costs to competitors in supplying or utilizing self-provided inputs to produce the output. In the case of telecommunications services, however, that's a very big *if*. CLECs have, for example, tried to combine their own switching with unbundled analog loops, and have found that the current hot cut, collocation and backhaul process make it uneconomic. Notably, if a reasonably priced electronic loop provisioning process (ELP) were available that allowed unbundled ILEC loops to be combined efficiently with CLEC switching, this could significantly reduce the CLECs' impairment in using alternatives to ILEC switching and possibly permit switching to be delisted as a UNE.¹²

The Circuit Court's Bridge Analogy

To make the foregoing discussion more concrete, consider the example raised by the Circuit Court of a natural monopoly associated with a bridge. Assume a monopolist owns a large and very expensive bridge connecting two cities and that it also has trucks and drivers to deliver goods between the cities.

A separate examination of each of these delivery service inputs suggests that while trucks and drivers are available at comparable costs to all delivery service operators, the bridge has strong natural monopoly characteristics because the average cost of shipping goods across the bridge declines continuously with the amount shipped.

Although other bridges across the river may exist at distant locations up or downstream from the two cities, it is not efficient for the competing *local* delivery services to send their trucks and drivers across these distant bridges. This is because their customers are located in the two cities at the opposite ends of the first bridge. Thus, the competitors would face significantly higher costs than the monopolist if they had to deliver packages between the two cities using these alternative circuitous routes.

¹² For a discussion of ELP, see Laurence J. Kotlikoff, "Telecommunications Policy--Promoting Investment and Vigorous Competition," mimeo, 2002. <http://econ.bu.edu/kotlikoff/TelecomLJK4-23-02.pdf>.

There are several potential regulatory solutions to this impairment. One option is to require the monopolist's delivery service to deliver packages collected by the competitive services at a wholesale price that matches its efficient costs (a "UNE-P" type solution). But a second option may also be available. If it were possible for the competitors' trucks and drivers to use the bridge at similar cost to what the monopolist experiences when its trucks and drivers access the bridge, it may suffice to provide competitors cost-based access to the bridge without also providing them access to the monopolist's trucks and drivers.

But such a solution may be impossible. Suppose, for example, the incumbent's loading terminal is located convenient to the bridge on the only designated truck road with access to the bridge. Further suppose that competitors are only permitted (because of new zoning ordinances) to locate their loading facilities in an industrial park several miles from the bridge, without a direct truck access road. If reaching the truck road that provides access to the bridge from this industrial park requires a two-hour drive through heavy traffic around the city, competitors cannot use the bridge without incurring additional expensive capital, labor, and operations costs (similar to hot cut/collocation and backhaul) compared with the incumbent. As a result, the added costs (and time) of shipping could well make the competitors' costs for the local delivery service too high for them to compete on reasonably equivalent terms in the retail market for finished delivery services. In such a case, the UNE-P-type solution might be the only viable one, unless competitors were allowed to use a different local road running straight from the industrial park to the bridge (an ELP-type response).¹³

As this example indicates, a pro-competitive decision-making process to determine whether entrants are impaired depends on the specific circumstances applicable in each relevant market – and requires examination of both the supply circumstances surrounding each input as well as the technology for combining inputs into final telecom services.

Other Forms of Market Power

In the above example, it made sense to have only one bridge. But suppose the river is sufficiently narrow and traffic sufficiently robust to permit efficient construction of one or two more bridges. Would competition flourish? Not necessarily. If the bridge owners explicitly or implicitly collude and set very high tolls, they could still achieve an "unnatural" monopoly outcome. Moreover, the likelihood of such collusion rises if the incumbent bridge owners can credibly threaten potential bridge builders with, for example, a toll price war once they enter the market after incurring the fixed costs of a) obtaining the rights to build a bridge and b) actually constructing it.

¹³ In certain circumstances, the competitors' higher costs of transfer may be reduced to manageable levels if, for example, the volumes of deliveries for specific large customers were so high that the cost disparities applicable to the total delivery service were a relatively small proportion of the total cost of shipping such large volumes. This explains in part why CLECs have not sought to use a UNE-P type arrangement for high volume customers of local telecommunications services.

Since lowering prices of telecom services and generating other consumer benefits is one of the key goals of the Telecom Act, any new impairment standard must take into account the potential effects of other forms of anticompetitive market power, including duopoly and oligopoly. This is where Dr. Willig's recommendation to use the Justice Department's and FTC's *Horizontal Merger Guidelines* ("*Guidelines*") becomes highly relevant.

Willig's Test for Impairment

Relying on the *Guidelines*, Dr. Willig proposes that a network element should be available if denying competitive carriers unbundled access to that particular element at cost-based rates would prevent those carriers from offering effectively competitive alternative retail services. Such a standard, which rests on established economic theory as well as basic regulatory and legal competition principles, would clearly meet the statutory standard of "impairment" as well as economists' understanding of the term. Moreover, as Dr. Willig indicates, another key advantage of such a standard is that it would help to "drive retail rates toward costs."

Dr. Willig's proposed impairment test would come into play whenever there is evidence that the market is not competitive as defined by the *Guidelines*. In this regard, Dr. Willig's test adds important pro-competitive criteria that go beyond a mechanistic determination of whether average cost is declining over the range of market demand.

As Dr. Willig explained, the *Guidelines* consider several barriers to entry: inherent incumbency advantages and other barriers to entry such as significant sunk costs or a high level of minimum viable scale.

Incumbency advantages may have a direct preclusive effect on competitive firms by completely preventing or substantially impeding these firms' ability to offer services that are competitive with those offered by the incumbent.

The most compelling example of an entry barrier in local telephony that is based on inherent incumbency advantages is that incumbents' loops terminate at the same central office where their serving local switch is located, whereas CLECs must always extend their customers' loops from the ILEC central office where they terminate to an external site where the CLEC switch is located. This characteristic of ILEC networks requires entrants to incur many costs and other service-affecting problems that the incumbents do not. Other examples of incumbency advantages include ubiquitous and favorable physical location (particularly in privately owned buildings),¹⁴ government provisions that discriminate against entrants,¹⁵ and name recognition. These and other incumbency advantages leave new entrants particularly vulnerable to price wars initiated by incumbents that would quickly put them out of business if they were

¹⁴ It is well documented that building owners frequently impose restrictions on new entrants that incumbents do not face.

¹⁵ For example, municipalities may not let new entrants put up new telephone poles or to dig up streets to lay new underground conduits or use existing conduits without charging considerable fees for the privilege.

forced to incur costs that are materially higher than the incumbents' efficient (*i.e.*, TELRIC) costs.

But no less injurious to competitive entry are barriers such as significant sunk costs or a large minimum viable scale (relative to the total market size). Significant sunk costs refer to major, one-time, expenditures associated with entry that are unrecoverable if the firm fails. Minimum viable scale refers to the percentage of the market that a firm must gain before its average costs compare reasonably with those of a larger incumbent. The existence of either significant sunk costs or high minimum viable scale raises the risk of entry. The reason is that if these barriers exist, competitors must not only incur significant, and possibly unrecoverable upfront costs, but they also must enter the market at such a large scale (to be cost-competitive) that their entrance likely would so flood the market with product that the market price will be bid down to the point that full recovery of their costs can no longer be expected.¹⁶

Dr. Willig reviews these three factors as they affect the three most significant elements in this proceeding, *i.e.*, loops, transport and switching. He then demonstrates that each of these three elements satisfies one or more of the above-stated criteria for impairment. As a result, Dr. Willig correctly advises the Commission to consider the aforementioned barriers in assessing impairment. Impairment, for Willig, would arise, in both a legal and economic sense, if a) the CLEC must incur significant sunk costs to supply the element or constellation of elements in question and/or the CLEC would have to obtain more than a modest minimum viable scale to recoup the fixed cost of the element or constellation of elements, or b) the incumbent has an inherent advantage in employing the element or constellation of elements in the production of finished telecom services.

The *Guidelines* use the following focal points to measure when an entry barrier is a substantial impairment to competitive entry. First, they consider a competitor's ability to enter a market to be impaired if a 5% improvement in its potential profitability (due, say, to a 5% increase in market price) would not be adequate to make its entry profitable. And if the market price remains fixed by the incumbent, this impairment is analogous to what would occur if the entrant's costs were 5% higher than those of the incumbent.¹⁷ But because elimination of scale disadvantages by entrants might require them to place

¹⁶ The risk that entrants will face dramatic price reductions is compounded by the fact that incumbents' short-run marginal costs are often much lower than TELRIC, so that CLECs often are not in a comparable economic position to the incumbent even if they can obtain access to UNEs at TELRIC rates. This situation leaves CLECs open to a classic "price squeeze" situation in which an incumbent lowers its retail prices to its short-run marginal cost in order to force out new entrants and later raises its retail rates to supra-competitive levels.

¹⁷ Although a new entrant also incurs higher costs than the incumbent to market its new services and to establish customer relationships, in light of the Circuit Court's ruling that such costs should not be considered in determining impairment, I do not reference them below. That said, I do not believe the Circuit Court is correct in differentiating such costs from other one-time start-up costs. From an economic public policy perspective, the preclusive effect of these costs on competition may be just as injurious as the preclusive effects of other impairments.

substantial extra product on the market, any scale economies whose realization requires more than a 5% increment to market demand also create an impairment.

Given the above, it would be conservative to establish as benchmarks for “material” economic impairment a 5% cost differential based on the incumbent’s TELRIC costs for providing a comparable service or a minimum viable scale that exceeds 5% of current market demand. In addition, if a CLEC cannot provision or maintain its services as quickly, efficiently, and accurately as the incumbent, customers will reject its offers.¹⁸ Thus, any material disadvantage in such service quality also legitimately demonstrates economically cognizable impairment.

A Proposed Definition of Impairment

The above discussion of the sources of natural monopoly and of the Justice Department’s and FTC’s *Guideless* suggests the following definition of impairment:

Impairment

A requesting carrier is impaired in supplying a telecommunications service if

(a) the production or delivery of a specific category of telecommunications service or combination of services in a relevant product and geographic market exhibits characteristics of a natural monopoly. A natural monopoly is deemed to exist if the incumbent’s TELRIC-based unit costs of providing these services decline over the range of relevant market demand; or

(b) there are one or more of the following barriers to entry with respect to the individual or joint use of one or more unbundled network elements used as input(s) to a telecommunications service or combination of services in a relevant product and geographic market:

- (i) A requesting carrier must incur sunk costs that cannot reasonably be expected to be recovered within 1 year of entry even assuming it achieves a market share as high as 5%;
- (ii) Achievement of minimum viable scale requires the entrant to serve more than 5% of current market demand.
- (iii) Material barriers to entry resulting from inherent incumbency advantages exist, including the following:

(A) A requesting carrier’s additional costs of using an alternative to a TELRIC-priced UNE to produce a final telecom service

¹⁸ Such disparities may result from inefficient processes that do not result in equivalent service quality or from external barriers created by third parties such as governmental entities or landlords.

amount to 5% or more of the incumbent's TELRIC costs of providing the equivalent final service; or

(B) Existing conditions prevent a requesting carrier from provisioning a telecommunications service in a manner equivalent in quality to the manner in which the incumbent provisions such service.