2005

**Justified Monopolies: Regulating Pharmaceuticals and Telecommunications**

Richard A. Epstein

Follow this and additional works at: [https://chicagounbound.uchicago.edu/journal_articles](https://chicagounbound.uchicago.edu/journal_articles)

Part of the Law Commons

**Recommended Citation**


This Article is brought to you for free and open access by the Faculty Scholarship at Chicago Unbound. It has been accepted for inclusion in Journal Articles by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.
My topic for this Lecture is one that on first inspection is easily dismissed as an oxymoron. Why seek to justify monopoly when so much of my own work has been devoted to an attack on various monopoly arrangements that have been created or propped up by government?1 We know that the dangers of state monopoly are commonplace in every area of productive endeavor. Today, it is difficult for firms to receive outright protection from the state. But a preferred monopoly position is often obtained by indirection. Restrictions on the importation of goods from other states or overseas are quite often meant to throttle competition for the benefit of local businesses. Yet, they are often advertised as efforts to preserve the health and safety of citizens against dangerous products that might otherwise enter the stream of commerce or, worse, the air and water supply. Efforts to keep minimum prices on milk, for example, have been defended on just that ground.2 These attacks on imports cannot be dismissed auto-

1 James Parker Hall, Distinguished Service Professor of Law, The University of Chicago; Peter and Kirsten Senior Fellow, The Hoover Institution. This paper is a revision and expansion of the Dean Lindsey Cowen Lecture that I delivered at Case Western Reserve University Law School on March 30, 2005. I have sought to take into account developments after the lecture was given. My thanks to Justin (Gus) Hurwitz, University of Chicago Law School, Class of 2007 for his persistent and able research assistance. For the record, I have consulted on many of the issues raised in this article for the pharmaceutical and telecommunications industry.


[m]ilk is an essential item of diet. It cannot long be stored. It is an excellent medium for growth of bacteria. These facts necessitate safeguards in its production and handling for human consumption which greatly increase the cost of the business.
matically, for in some cases they may prove to be true.\(^3\) Although, all too often these health and safety claims are pursued by local competitors whose interests are not aligned with those of the consumers in whose name they purport to speak. Whether we are working in the domain of constitutional law, legislation, or common-law decisions, cutting through this bloated rhetoric of public health risks is the first critical step in figuring out how to combat the mischief of these legal monopolies. All too often, the failure of our legal institutions stems from the eagerness with which monopolists of all stripes hide behind protective barriers provided by supine or misguided legislators.

Given this general orientation, how can we use this odd term "justified monopoly" without succumbing to the inveterate vices of protectionism? The first part of the answer is conceptual. The standard libertarian, taking his cue from Thomas Hobbes, believes that the use of force and fraud in human affairs is the primary threat to social life.\(^4\) But it hardly follows from that sensible presumption that the use of force and fraud are never justified in human affairs. That absolutist position cuts against the intuitions of every reflective person. Surely force must be justified in some cases, such as self-defense, even if there are some important limits on the scope of the privilege. And by the same token, who would not lie to keep a kidnapper away from his helpless prey? Lest one fall victim to a stirring but misguided form of Kantian absolutism, there are cases of justified fraud just as there are cases of justified force.

Ordinary discourse, then, recognizes that a limited number of exceptions are consistent with the articulation of our basic rules on force and fraud. I shall argue that this same approach fruitfully carries over to our thinking about monopoly, even state-sponsored monopoly. Here, I hope to explain the basic logic of the term "justified monopoly" and then show the stunning ways in which it plays out when dealing with the pharmaceutical and telecommunications industries.

I. JUSTIFYING MONOPOLIES

The first point in dealing with the task of justifying monopolies requires offering some explanation as to why these operations should be

---

Failure of producers to receive a reasonable return for their labor and investment over an extended period threaten a relaxation of vigilance against contamination. *Id.* at 516-17.

\(^3\) See, e.g., *Maine v. Taylor*, 477 U.S. 131 (1986) (upholding state ban on imported bait-fish because it, inter alia, protects the local environment from parasites and invasive species).

\(^4\) For my recent defense of this view, see RICHARD A. EPSTEIN, SKEPTICISM AND FREEDOM: A MODERN CASE FOR CLASSICAL LIBERALISM (2003).
JUSTIFIED MONOPOLIES

regarded as wrongful in the first place. To a libertarian (or more accurately a classical liberal, like myself) this is not an easy task given that the general prohibition against force and fraud done in aid of liberty does not contain any obvious injunction against monopolies. A person is normally in a position to follow the advice of Henry Higgins, and do "precisely as he likes" so long as he or she does not cross that border that separates him from another person. And no matter how much we dislike monopolies, we have to recognize that changes in the form of the marketplace are not tantamount to beating some other person over the head with a brick or committing an elaborate fraud to bilk innocent investors out of millions of dollars.

Yet ordinary business individuals operating within the same market do not violate the constraint against either force of fraud when their sole objective is to combine their operations in order to reduce output and raise price relative to their levels in a competitive market. Long before the rise of the New Deal, the common law had a tough attitude that denied the enforcement of those contracts in restraint of trade that called for market divisions. The basic attitude depended on two factors. First, members would "cheat" on the cartel by offering hidden concessions to buyers in order to increase their market share. Second, the artificially high market price would draw new firms into the market, because they could easily "steal" market share from the firms in the overpriced cartel. The empirical assumption (which need not be true in all markets) was that these two elements supplied sufficient discipline to obviate the need for extensive government actions to ferret out price-fixing and other forms of anticompetitive arrangements.

On this view, the libertarian prohibition against force and fraud only comes into play when the government uses legal rules to confer a monopoly on a particular party in a given market. Since the ordinary definitions of liberty and property include the right to sell to willing buyers, the legal prohibition, backed by force, acts as a necessary restraint against the freedom of other individuals to carry out their trade in a form that meets their satisfaction and that of their customers. If one person could not block a rival from selling his goods, then the state cannot use force on his behalf.

At this point, we do not have to deal with the tricky economic rationales used to attack simple private contracts between private parties. Rather, we can be much more categorical in the judgment that

---

5 See, e.g., Mitchel v. Reynolds, 24 Eng. Rep. 347 (K.B. 1711) ("[A]ll contracts, where there is a bare restraint of trade and no more, must be void; but ... in cases where the special matter appears so as to make it a reasonable and useful contract, it should ... be good ... ").
the creation by law of a monopoly position in one person necessarily results in the infringement of the like liberty of other individuals. Most of the sensible decisions that struck down state-created monopolies under the banner of substantive due process were aimed at this form of abuse.\textsuperscript{6}

The basic logic of this position strikes me as impeccable. The monopolist is someone who thinks that he can profit by taking a larger share of what he knows (if he cares to reflect on the consequences of his action) will be a smaller pie. Concede that this calculation is correct for the monopolist, and it follows necessarily that others are left with a smaller share of a smaller pie. The larger slice explains why the monopolist will expend resources to gain his position. The total loss in social welfare is the shrinkage in the size of the pie, which explains why there is widespread opposition to monopoly by those who do not treat the libertarian position as the last word on the subject. This last judgment is supported by looking at the point from behind the Rawlsian veil of ignorance.\textsuperscript{7} If a single rational actor did not know whether he would be the buyer or seller, would he prefer a world of perfect competition or one of legal monopoly? Since the person is ignorant of which slice of the pie is his, he will chose the larger pie, which is the competitive solution.

At this point, this simple critique of monopoly explains what must be established to show that some monopolies are justified. The proponent of that position must, in some way, demonstrate that certain monopolies create larger pies. The question is whether the monopolist can show that its privilege produces positive gains for society as a whole. Getting a larger share of a smaller pie is doubtful to say the least. But getting a larger share of a larger pie is fine, especially if the share held by others does not shrink as well. A monopoly is then justified when the consequences of the arrangement deviate from the usual pattern by supplying some overall social gain shared by the public at large.\textsuperscript{8} This task is no small order given the vast number of

\textsuperscript{6} See, e.g., New State Ice Co. v. Liebmann, 285 U.S. 262 (1932).
\textsuperscript{7} See JOHN RAWLS, A THEORY OF JUSTICE (1971).
\textsuperscript{8} It might be informative to note here that at one time, in England, monopolies were granted on case-by-case basis by letters patent. Adam Mossoff, Rethinking the Development of Patents: An Intellectual History, 1550-1800, 52 HASTINGS L.J. 1255, 1259-64 (2001). Consider, too, Adam Smith's classic take on the question: When a company of merchants undertake, at their own risk and expense, to establish a new trade with some remote and barbarous nation, it may not be unreasonable . . . to grant them, in case of their success, a monopoly of the trade for a certain number of years. It is the easiest and most natural way in which the state can recompense them for hazarding a dangerous and expensive experiment, of which the public is afterwards to reap the benefit. A temporary monopoly of this kind may be vindicated.
government initiatives that go astray. So our task is to identify the initial conditions in which some monopolies meet the same standard of social welfare while ordinary monopolies in foodstuffs and commodities flunk. What monopolies, then, make for larger pies?

II. MONOPOLY, CONTRACT, AND PROPERTY

The initial difficulties with this inquiry start at the conceptual level. In dealing with ordinary monopolies, that is, those not backed by the power of the state, we face a constant challenge. Does the particular agreement between two individuals just create a firm that works well by allowing for specialization of activities under a single roof? Or does the agreement amount to a cartelization of the industry as a whole? Surely any agreement between two lawyers to start a firm counts as the creation of a firm rather than the establishment of a monopoly given that thousands of other lawyers can, and routinely do, form hundreds of firms of their own. It is only when we see combinations of large firms with huge market shares that we start to worry about the ability of the firm to control price and quantity—if we worry at all. We can ignore these complications, however, because our focus is on those arrangements that receive special legal protection from the state in the form of the right to exclude others.

When we turn to the class of legal monopolies, the question of definition is more difficult to resolve. It is one thing to attack a monopoly on the ground that it gives a single person the exclusive right to sell some product in a particular location. But it is quite another to distinguish monopolies that we fear from ordinary forms of private property that we generally welcome. After all, the standard definition of property is expressed in terms of exclusion: property is that over which the owner has the exclusive right to possess, use, and dispose of. How, then, are these conflicting impulses reconciled?

Begin by thinking about this in connection with land. Anytime one individual has ownership of a plot of land, he also has the right to exclude others and, therefore, holds a monopoly position with respect to that land. If the plot is characterized by the slightest bit of uniqueness, even as simple as its location, then its owner enjoys some element of monopoly power.

Why, then, is this institution of property, at least with respect to the single plot of land, justified? The answer is a strong one that de-

pends on the key role of incentives. In the formative period of the hunter and gatherer, land was not typically reduced to ownership. Most people only wandered through the land and claimed ownership of coconuts on one hand and wild animals on the other. Keeping possession of a particular plot of land was of no particular advantage but only a drain on resources. Even the establishment of large territories in which members of a tribe or a clan wandered was not an easy task given the difficulties of demarcating and defending borders against other wanderers. If you had to move on to follow the herd or the crops, why enervate yourself in trying to keep hold of land that was remote from your current location? But the moment agriculture began, these same individuals had to create exclusive rights—a legal monopoly of sorts—over a single plot of land. Otherwise, to quote the conventional wisdom, the nominal owners of property would not cultivate it, because they knew that someone else would come in and reap the harvest after the hard work was done.  

Therefore, the reason monopoly is justified in this particular case is because it creates the appropriate incentives for people to sow today and remain secure in the knowledge that they can harvest tomorrow. There is no doubt this rule is tempered at the margins to take into account, for example, the position of the poor who are allowed to gather the gleanings from the field. This kind of support for the poor reflects the traditional Lockean norm that some steps should be made for persons in condition of extreme want. But those distributional issues are tightly constrained and in no way allowed a stranger to barge in to harvest the entire crop that someone else planted on his own grounds. The creation of permanent interests in property negated what might be termed a temporal externality, whereby labor invested at time one could be expropriated at time two.

So why has this monopoly over finite plots of land not turned out to be a terrible thing? Because, essentially, the definition of competition points to the situation in which private monopolists who operate near each other or side by side in many cases, are all competing for

---

9 As William Blackstone put it:
And the art of agriculture, by regular connection and consequence, introduced and established the idea of a more permanent property in the soil than had hitherto been received and adopted. It was clear that the earth would not produce her fruits in sufficient quantities without the assistance of tillage; but who would be at the pains of tilling it, if another might watch an opportunity to seize upon and enjoy the product of his industry, art, and labor?

2 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND *7 (1799). No modern economist could say it better.

the same class of customers. I own my shop, which is my castle and my local monopoly, but my neighbor owns his, his neighbor owns his, and so forth. My ability to raise the price for services rendered by my "monopoly" (for example, my exclusive right to plant and harvest on this land) is constrained, because my neighbor enjoys a small ration of monopoly power that enables him to compete with me. Assemble enough legal monopolies (that is ordinary exclusive property rights in land) side by side and, lo and behold, in economic terms there is a trade district with strong competition. (Indeed as effective transportation expands the scope of the relevant market, even the side-by-side constraint can be relaxed.) The short analysis of market power in connection with the law firm, discussed earlier,\(^{11}\) applies with equal force to real estate. The only substantive difference is nominal: when large numbers of people who have the exclusive right to deal with some resource operate in competition with one another, we call it private property instead of monopoly. It requires, at the very least, some careful social judgment to figure out which description applies in a given case. This judgment is surrounded by ambiguities and uncertainties for both intellectual property and telecommunications. It becomes, therefore, exceedingly important to figure out the lay of the land in order to understand the dynamics of both fields. That knowledge is critical for evaluating a range of possible regulatory reforms in these two areas.

III. PATENT MONOPOLIES: PHARMACEUTICALS

The patent situation offers an instructive parallel to the real estate example. We can forget about the nature of the underlying assets, as both areas give rise to the same problem. Imagine an inventor hits on a wonderful idea and then builds a new invention never before seen anywhere on the face of the globe. Applying the old common-law rules, the inventor only has possession of the tangible object he has made by virtue of having identified the invention. After he has built his invention, his friend sees it, buys it, and then, after studying it for a bit, the friend decides to replicate and sell that invention himself. After the fact, everyone says that the multiple sources of potential suppliers are a wonderful way to spread that invention far and wide. Conditional upon both manufacturers being in the field, the price will go down, allowing for the sale of more units. If the inventor's friend can do it, everybody else can do it, and eventually a competitive mar-

\(^{11}\) See supra p. 107.
ket develops on top of the wonderful invention created by virtue of the inventor’s sole labor.

A moment of reflection indicates why, like everything else, this happy tale is only half true. The story assumed that the initial inventor was oblivious to the events that would follow once his invention was disclosed to the world. In a common-law world without state-created patent monopolies, the inventor has in effect bestowed free gifts upon his competitors. The competitors can produce the same product for less, because they do not bear any of the original front-end costs of development. Yet the inventor would have to load his costs of discovery into his price. The inventor who does all the work, therefore, gets driven out of the market. This pattern of behavior is eminently foreseeable, and it will induce the astute inventor to change course. He is better off hoarding his capital and, therefore, he will not start the process of discovery in the first place. The predictable result will be the scoreboard that everybody hates to see in the field of intellectual innovation: no runs, no hits, no errors, because of the utter absence of productive activity. The rules against force and fraud do not guard against this unhappy contingency.

The creation of the patent monopoly, with its exclusive right to sell, is the only way to avoid this dreaded scenario. This situation is vastly different from one in which producers form a cartel to rig prices. The patent stimulates production in the way that the garden-variety cartel does not. Stated otherwise, the patent monopoly is fully justified by the increased productivity that it spawns. That patent does not come easily to the applicant. The state will not give it to someone whose invention is not new12 or does not advance the state of knowledge appreciably over the prior art. The new invention has to be “nonobvious,”13 or if you like the European term, it must “involve[] an inventive step” that enables others to discern why the patent has been granted.14 Next, the inventor must explain to the state with sufficient fullness, inter alia, how he made the invention and how he has enabled others to make and use it.15 These steps are required so that

13 Id. § 103.
14 See Agreement on Trade Related Aspects of Intellectual Property Rights art. 27 ¶ 1 n.5, Apr. 15, 1994, 33 I.L.M. 1197. [hereinafter TRIPs agreement] (calling for the inventive step, and noting that “[f]or the purposes of this Article, the terms ‘inventive step’ and ‘capable of industrial application’ may be deemed by a Member to be synonymous with the terms ‘non-obvious’ and ‘useful’ respectively”). The terms “useful” and “non-obvious” are covered in the U.S. Patent law. 35 U.S.C. § 101 (2000) (referring to “any new and useful process”); id. § 103 (covering “non-obvious subject matter”). There is no obvious reason why treaty signatories should have the option to deem the two sets of definitions as unequal given the importance of uniform rules in a global economy.
the state can, in effect, take the inventor’s knowledge and place it into the public domain, thereby allowing other people to invent around him—that is, to build a house next-door.

There are many libertarians who hate patent law, because it necessarily creates exclusive rights in the form of a state monopoly. If there is one thing good libertarians know, it is that although voluntary monopolies are fine, state monopolies are horrible. But libertarians who really believe this proposition without qualification cannot believe in private property either: I may take first possession of my land, but if the only thing I can do to protect it is self-help, then it is a rather precarious title. Therefore, the state must intervene so that the landowner may say, “If I occupy the land today, I am going to have a bigger, richer, and longer title that will allow me to stay on this land in perpetuity.”

As indicated above, what justified the monopoly over land was the increased productivity from agriculture and industry. What limited this monopoly was the ability of the neighbor to do exactly the same thing on his land. Standing in the wings, antitrust laws remained available to prevent collusion in the unlikely event that too many owners came together. The same argument turns out to carry over, word for word, to the field of patent protection.

The ultimate deal that the state gives for patent protection is not going to be the same deal that it offers with respect to the land. For patents, the inventor has to show a little more ingenuity and wit, because the scope of the monopoly the state gives him is a little bit broader than that for land. Hence, the real question involves finding out whether the creation of a new set of property rights, sitting uneasily on top of the old property rights, will improve overall social welfare. In some sense, it is instructive to think of the situation as a bargain. The state has its monopoly power that it is prepared to confer on the inventor. What ought it demand in exchange from the patentee for the benefit of the public at large? The task here is not an easy one, because the transaction is not simply one in which the parties immediately exchange goods (or cash) with each other and go their separate ways. As Hobbes recognized long ago, a dilemma arises in connection with private transactions because sequential performance invites default at stage two that in turn leads back to nonperformance.

at stage one. Hence, the state must secure compliance to avoid the no runs, no hits, no errors outcome. This very temporal issue arises in the patent context. The standard utility patent lasts for twenty years. The system is, therefore, predicated on the assurance that the state will honor its monopoly grant during that period—so long, of course, as the patentee complies with all conditions of the grant. At the end of the deal, the invention covered by the patent falls into the public domain.

Confidence in the system is critical for innovation. If they have no confidence that their patents are enforceable, inventors will stop inventing because they fear that interlopers will swoop down to steal their markets. No runs, no hits, no errors possibly becomes a dominant solution. The market will shut down, except to the extent that people can find ways to use their inventions without disclosing them to others, say, by treating them as trade secrets. But even with that option (which should be available, but not required) the need for state intervention is great. Trade secrets typically have to be shared with licensees, most likely through confidentiality agreements, which again require state enforcement. So long as trade secrets are an imperfect substitute for patents, the central challenge remains: given that the state is a party to the agreement, will it perform its obligations once it has already received the benefit of the other party’s invention?

By allowing others to enter the field before the end of the patent term, there is a political risk of expropriation of the invention. We know that this risk is a serious one from the history of monopoly regulation in the United States. The basic problem associated with patents also arises with respect to railroads, power, telecommunications, and the like. The scenario unfolds as follows: Suppose I put all the pipes in the ground at a cost of $1,000. Each individual unit of power costs $0.10 to transmit, so that the price has to be in excess of that to recover the initial investment, plus interest, before the franchise expires. Suppose that an additional $0.05 is needed to cover the fixed costs, on the assumption that 20,000 units will be sold over the period. The initial deal allows the price to be $0.15 per unit, plus

---

17 THOMAS HOBBES, LEVIATHAN (Richard Tuck ed., Cambridge Univ. Press 1990) (1651) Hobbes wrote:

For he that performeth first, has no assurance the other will performe after; because the bonds of words are too weak to bridle men’s ambition, avarice, anger, and other Passions, without the fear of some coercive Power; which in the condition of meer Nature, where all men are equall, and judges of the justnesse of their own fears, cannot possibly be supposed. And therefore he which performeth first, does but betray himselfe to his enemy; contrary to the Right (he can never abandon) of defending his life, and means of living.

Id. at 96.
some amount to cover interest and to allow for a reasonable profit. But once the line is finished, the state reverses and says that it will allow only $0.11 per unit, or $0.15 for a shorter period of time. Whether we talk about time or amount, the opportunity to chisel remains with the state. If rates are cut to $0.11, without shortening the period, the firm will continue to produce; even though it will not be able to recoup its front-end investment, it can at least make a profit relative to its operating expenses. The same happens if the state cuts the duration and leaves the price untouched. But recall Hobbes's intuition: this scheme will not work twice. Once it happens, prospective innovators will be twice as shy, and we are back to the scenario of no runs, no hits, no errors in the next period. The stability of this monopoly is as important as the stability of possession was (and is) to the traditional common law.

What does one do about all of this? The first part of the problem is to figure out whether one uses rate control (as with electricity) or durational controls (as with patents). That is a legislative decision that sets the basic rules of the game. And here, there is a built-in political incentive to get the rules right, because if the government offers too little, there will be too few investors. But for our purposes, the key element is that once the government sets the rules of any particular game, it holds itself to those rules. And, given the pressures of politics, the best shot at achieving this result is through some degree of constitutional protection. One way is to think that these protections stabilize the public utility or the patent bargain by ensuring government performance. Another way is to think of the issue as a takings question.

Which way is the best way to think about it? Oddly enough, the choice of doctrinal homes does not matter all that much. A simple example illustrates the point. I sell you goods for $10. I deliver the goods and you pay me the $10 and then turn around and snatch the $10 back again. That is a taking of $10. But why should you bother giving me the $10 only to take it back? That is just too messy. Rather, you will take delivery of the goods and not pay me the $10. So instead of having the money sashaying first to me and then back to you, it just stays put. Functionally, the result is exactly the same. If the first case is a taking of $10, then the breach of contract is also a taking of $10. Each scenario implicates the same issue—how we secure government performance in the second period. Therefore, there is no sharp dividing line between what counts as a contracts case on the one hand and a takings case on the other hand.
The question of which doctrinal home is best is easy to state and hard to answer because reneging on promises is a fine art in which governments have broad experience. (Consider their willingness to pass retroactive laws.) In particular, there are two general ways that the government can renege. One way, which a conniving political actor will never use, is simply to announce, in bold terms, his exact intentions: "We like your invention so we are going to take it by passing a statute that reneges on your patent." There is a strong argument that this bald approach becomes a taking on the ground that you can change the patent law going forward, but you cannot simply repeal the law after somebody has acted in reliance and has taken out the patent. This is somewhat like a section 90 promissory estoppel theory against the government with respect to existing patents. The federal government understands that the political cost of blatant turnarounds is high. Even the Supreme Court is alert to this particular form of malfeasance when it comes to property rights. Therefore, the government must abandon open repudiation and instead use more nuanced strategies to generate the same or similar result at much lower political cost.

Such subtle approaches are not exclusive to the patent area. Once again the land analogy is instructive. The government knows that if it occupies land, then it must pay full market value. But under the current law, a restriction on land use, such as zoning, may not be compensable even if it results in a substantial diminution in land value. The Supreme Court requires different prices that the government has to pay when it takes different actions. A full price is required if it occupies the land. The state may very well prefer this deal over the status quo. The land could cost the government $100 and be worth $150 when dedicated to public use. But regulation could easily offer the government an even better deal. Of course, the government that cannot occupy land may find that it is worth less than $150. Say it is worth $100, if kept as open spaces, subject to a conservation easement. Once that easement is imposed, the value of the private land drops to $10. But now suppose that for this action, it need not pay anything at all. Here its gain is $100 dollars, so it will prefer this action. But the overall social pie (the sum of the public and private benefit) is $150 in the first case but falls to $110 in the second. The regulation induces the state to follow a course of action that produces a state gain but a social loss. The reason for the difference is that the

---

government must respond to the full extent of the private losses in the first case, but may blithely ignore them in the second. The temptation to take this regulatory shortcut will be too great for many governments to resist.

Accordingly, the savvy government will game the system: it does not take private land to build a public park but instead forbids the private landowner from building on his land. He remains in possession but is deprived most of the rights of use. The total prohibition on all use will not work because it is too close to the taking, so subtlety again takes over. If the landowner wants to exercise his right to sell the land, these rules require him to run through a gauntlet of hearings to get all the necessary permits in a process that can easily run a generation. Now the question becomes: if you leave somebody the outer rind of the orange but suck out all the juice inside, is that a permissible taking? The Supreme Court’s answer is, yes, so long as the rind is somewhat damp, which is said to happen when an owner has some (undefined) viable economic use, even if he suffers a substantial capital loss. Regulation, in effect, can take away a large fraction of property rights without compensation. The result is exactly what is feared: an unwillingness of people to invest, precisely because the legal framework has been undermined.

The same approaches may well be available for nontangible properties. The motivation to cut corners is high, and the legislative imagination is often equal to the task. First, I will consider the motivation and then, I will go through the imagination.

The motivation follows clearly from what I have already said. When creating a patent monopoly, the first question to ask is: how strong is that patent monopoly? The question is just like the real estate case. If a firm has the only statin in town that controls cholesterol levels, the monopoly is very powerful given the want of close substitutes. Yet, if, as is the case, there are six or seven different statins available in the market, there will be some product differentiation, just as with ordinary homes. The advantage of having the first patent is reduced by the competition available, which is an advantage to the public at large. That result holds even if the drugs that are covered by two patents operate in an identical fashion in all cases. That, of course, never happens because some individuals will respond better to


21 See Lucas, 505 U.S. at 1030 (“When, however, a regulation that declares ‘off-limits’ all economically productive or beneficial uses of land goes beyond what the relevant background principles would dictate, compensation must be paid to sustain it.”).
one treatment than another, so that the introduction of these "me-too" drugs increases the number of options that ordinary people have. It is true that the costs of parallel development are greater than that for a single drug, but this concern also applies to competition between products that are not protected by patents. The key point is this: even if one concedes, as I would not, that the FDA does a good job in policing for health and safety, there is no reason why it should make an ill-considered venture into industrial policy by seeking to pick winning and losing investments. It has no special competence on issues of market concentration.

Even if the second drug is superior to the former, the first-mover advantages make it likely that the second-in-class will, at the margin, be less valuable than the first-in-class. This simple fact should remind us how critical it is to get the first-in-class to market as quickly as possible. The simple point here is that consumers do far better when they have a product sold for a monopoly price than they do if they have no product at all. Some consumer surplus (equal to the reservation price paid by consumers over the market price) is better than that dreaded outcome of no runs, no hits, no errors. Clearly, the more that patents expedite new inventions, the greater the overall social gain—a gain that includes the net profit of the drug manufacturer and the multiple parties (shareholders, employees, suppliers) who claim profits through the manufacturer. Accordingly, it is in the public interest to give strong protection, once a product is made, in order to speed it to market.

Yet there is a hitch to this story. In order to encourage a party to bring a product to market, it has to be able to cover the cost of the first pill, which includes all development costs. A crafty legislator, however, will realize that he only wishes to buy the second pill. Why is that the case? Because the first pill will cost, depending on who you ask, somewhere between $500 million and $1.3 billion. I have a headache, or even a tumor, but treatment at that price is too steep. Once everybody lines up to buy the second pill, nobody is going to buy the first pill. Unless this problem is solved, we are back to our familiar no hits, no runs, no errors scenario.

So this simple economic demonstration sets up one of the most tenacious economic problems. If the first user does not pay the full cost of the initial pill, then some other party, or parties, will have to do so.

---

22 See Richard A. Epstein, Regulatory Paternalism in the Market for Drugs: Lessons from Vioxx and Celebrex, 5 YALE J. HEALTH POL'Y L. & ETHICS 741, 747 (2005) (arguing that the FDA’s "entire effort to make better judgments on what treatments should be used and why smacks of an unthinking paternalism that reveals its own institutional shortcomings, as well as those of its critics who plump for stricter regulation").
That initial cost is in the millions and has to be spread over all subsequent pills. What does that mean? In fancy jargon, this means that a producer can never engage in marginal cost pricing if he is in the patent business. Since the first pill is below marginal cost, then the second pill has to be above it, or otherwise the patentee will never recover his front-end investment. So, in effect, prices are not driven down to marginal cost. Once that happens, the producer has the unhappy task of figuring out how much of the fixed cost he is going to allocate to every pill, between $2 and $200, or more. The patentee immediately discovers that there is no unique allocation, a point that quickly becomes equally apparent to everyone else. In consequence, gamesmanship starts. Everyone tries to figure out the following truism: is there any way I can get my pills at marginal cost and make sure some other party is going to have to pay for the fixed cost of development? The wise firm will try to price discriminate, so as to get higher prices from parties with greater demand. They, in turn, will disguise their preferences or play tough in return. The resulting marketplace is chaotic and costly to operate. The temptation to regulate the pricing in these circumstances is great.

The temptation of the state to regulate is equally great, which brings us right back to the problem of sequential performance. The patentee performs first, by investing the $800 million required to develop the new pharmaceutical product. The government performs second. To be sure, the government cannot simply strip the producer of the patent, but, just as with the real estate case, it can start to maneuver. The various approaches that have been proposed over time do not deny the patentee the exclusive right to sell. But they are motivated to constrain the set of prices (or cut corners).

Let us now turn toward the imaginative ways that legislatures have used to cut corners. The following are some of the proposed variations that I have encountered in my work as a consultant for the pharmaceutical industry, especially for the trade association PhRMA. Most of these proposals have not been enacted into law, but one, in Washington, D.C. was, and we should be thankful that it was struck down in a pointed decision by the D.C. District Court. Even so the

---

23 See Ronald H. Coase, The Marginal Cost Controversy, 13 ECONOMICA 169 (1946) (articulating, for the first time, the basic insight that applies to all cases of high cost for the initial unit and lower costs thereafter); John F. Duffy, The Marginal Cost Controversy in Intellectual Property, 71 U. Chi. L. Rev. 37 (2004) (exploring the novel solution required by the marginal cost problem in intellectual property).

issue promises to be a perennial one that merits closer discussion. Here is a rundown of some of the low points in this struggle.

One early maneuver was proposed by Senator Debbie Stabenow whose proposed legislation sought to fix prices as follows: once you sell a single pill to a nonhumanitarian, nongovernment agency, that sale would set the price at which you must sell to all other customers. More simply, the first sale sets the price and, from then on, the producer has to charge that single price. This single price approach is a mistake because, in many cases, the demands will be different. If all customers pay a single price, then high demanders, who might have picked up a large portion of the fixed costs are no longer going to do so. Since high demanders pay a lower price, the producer will have to raise the price to the low demanders, some of whom will be shut out of the market. This is made worse because while the government prohibits producers from engaging in price discrimination, it also picks the point at which producers have to set prices. The bill did not even address the clear question of whether or not sales prior to the passage of the Act would be the benchmark.

Essentially, if the Stabenow proposal had passed, a producer generously giving a small clinic a great deal, because it knew that the clinic would not resell the product and undercut his market, would rue the day. The price set in a friendly $10,000 transaction would set the prices for the other $10 billion worth of products that the producer thought it could sell over the life of the patent. The situation differs, but only by degrees, if the first-price sale postenactment governs. At this point, there will be no more discounts, so that many marginal buyers will be forced out of the market until some new price control scheme takes over. This pricing restriction will feed back into the development process, where the lower rate of expected returns will reduce investment and delay the entry of new products to market. This is a bad deal all around. I believe that depriving the owner of the right to set prices counts as a taking of the property, but the law on this question is so murky that it becomes almost impossible to figure out which way courts will go on the question.

There is second way to regulate patents, called parallel importation, that exploits the price differentials between the United States and overseas markets. An additional complexity in pharmaceutical markets is that national governments can set the prices at which goods are


sold in their countries, by precluding sales to competitive private parties. To make matters more complicated, it is difficult to just say no in these cases because there is always the threat that the foreign government will allow generic production of the patented drug if it cannot purchase the needed drug. It takes only a little imagination for some American tourist in Canada to stock up on a drug sold there that is of use in the United States. And it takes just a bit more to forget the side trip and to order direct from Canada. American companies are alert to those difficulties and seek to develop ways to prevent such reimportation. Two kinds of reasons are relevant. The first is that the long chain of custody introduces myriad risks of potentially deadly adulteration, which are hard to counter, especially if the importation is done in bulk. The second is that the American firm could restrict the supplies sold to foreign nations to an amount that covers its internal demand, leaving little for sale back to the United States.

The price differential thus sets up a domestic storm, and in 2004 one piece of legislation, the Kennedy-Dorgan\textsuperscript{28} bill sought to cut back on the patent monopoly in the United States. Under that bill, firms that do business in the United States would have been required to sell to foreign buyers whatever quantities they demand. The established price for these mandated transactions would have been equal to the price at which the firm sold drugs to that country for its internal consumption. Thus, if the price for a pill in the Canadian market was $1.00, then that is the price that the drug would have had to be sold at to all qualified parties in Canada for export into the United States or, even under the statutory language, for sales into third countries. The drugs, when transported, could have then undercut the higher prices that are found elsewhere. Under these schemes, the buyers could have then competed with each other to sell these goods in the United States. The imported drugs could have been in sizes or forms that are not approved by the FDA. If so, then the American firm would have been required at its own expense to help its new rivals secure FDA compliance for those nonstandard products.

The precise details of the scheme are not critical to this analysis. What is critical is that these schemes are klutzy and expensive forms of domestic price controls, set without reference to the cost of making these drugs. And just as price controls produce shortages in other industries, shortages will happen here as well. Working at no runs, no hits, no errors, is better for a drug producer than going into a loss situation. Once again, this program all looked just like confiscation,

analogous to the forced sale of land at a fraction of the price. But again, the legal position is sufficiently murky that it is a far riskier business to figure out how courts will respond to this extreme destabilization of property rights.²⁹

In 2005, Congress brought forward a variation of the earlier system, which in one sense at least acknowledges the constitutional infirmities in the earlier legislation. Unlike the Kennedy-Dorgan bill, the Pharmaceutical Market Access and Drug Safety Act of 2005 (the “Market Access Act”)³⁰ does not flat out require all pharmaceutical firms to sell their products in Canada (and several other nations like Australia, New Zealand, and most European Union Countries).³¹ Rather this obligation to sell is conditional upon the decision to sell for consumption in that domestic market. Once those sales overseas are completed, then the payoff is simple enough. The Market Access Act contains a provision that holds explicitly that resale of goods sold overseas shall not be treated as acts of infringement.³² The firm that is willing, for example, to forgo sales in Canada, need not comply with the other provisions of the proposed statute. The ostensible purpose of this provision is to beat back the charge that participation in the overall scheme is mandatory. But that conclusion is simply erroneous. The choice that is offered in this case is best described as a “Hobson’s choice” that is comparable in all its particulars to one that allows individuals to choose between their money and their life. Stated otherwise, the fact that there is a choice tells us nothing about the legitimacy of the private or government decision that leads up to the choice in question. An individual being robbed, for instance, is offered the choice between his life and his money. To treat that offer as legitimate will allow everyone to get the property of another by a threat of murder. The legislative case is not quite so bloody but it is every bit as unprincipled. Before the passage of the Market Access Act, a patentee could exercise both its American and foreign patents. Now it is forced to choose between them. To be sure, there is no question that the modern takings law is more lax than it ought to be, with its easy

³¹ The Market Access Act would “appl[y] only to the sale or distribution of a prescription drug in a country if the manufacturer of the drug chooses to sell or distribute the drug in the country.” Id. § 804(n)(3)(A) (proposing to amend 21 U.S.C. §§ 381-384). It also states that “[n]othing in this subsection shall be construed to compel the manufacturer of a drug to distribute or sell the drug in a country.” Id.
³² “It shall not be an act of infringement to use, offer to sell, or sell within the United States or to import into the United States any patented invention under section 804 of the Federal Food, Drug, and Cosmetic Act that was first sold abroad by or under authority of the owner or licensee of such patent.” Id. § 204(d)(1) (proposing to amend 35 U.S.C. § 271(h)).
substitution of "investment-backed expectations" for private property. But it is hard to think of a firmer expectation, backed by more extensive investment, than those associated with worldwide patents. The burdens that this statute imposes on the exercise of both foreign and domestic patent rights should spell its welcome constitutional doom. It is just not appropriate to encourage the heavy investment in scientific research under one regime and then just yank out the rug by the adoption of a punitive measure under another regime that prevents the realization of the promised gains.

There are other schemes that could be adopted as well. The District of Columbia recently passed legislation—the Prescription Drug Excessive Pricing Act of 2005 (the "Excessive Pricing Act" or the "D.C. Act")—that was intended to curb the excessive prices for prescription drugs in the District of Columbia. Its operative provision did not give a definition of an excessive price, but did create a presumption that any such price would be excessive if its price was "over 30% higher than the comparable price in any high income country in which the product is protected by patents or other exclusive marketing rights." The high income countries were the United Kingdom, Germany, Canada, and Australia. The presumption in question could only be rebutted by an elaborate and costly showing that takes into account its costs of invention, global sales, past profits, and government research support.

This last qualification opened up an enormous can of worms, for it attempts to impose a rate of return regime on the costs of new drugs. But this brief description seems to exclude the possibility that the high rates of return on successful drugs have to cover the costs of those drugs that make no return at all. And the reference to accumulated profits to date carries with it the implication that once the costs of recovery are recovered in a particular case, the permitted rates in question will be cut as close to the marginal cost as possible. The en-

35 Id. § 28-4554(a).
36 Id. § 28-4554(b).
Where a prima facie case of excessive pricing is shown, the burdens of providing evidence and proving by a preponderance of the evidence shall shift to the defendant to show that a given prescription drug is not excessively priced given demonstrated costs of invention, development and production of the prescription drug, global sales and profits to date, consideration of any government funded research that supported the development of the drug, and the impact of price on access to the prescription drug by residents and the government of the District of Columbia.

Id.
tire edifice looks, therefore, like a crude effort to invoke the rate regulation systems that have been used to deal with natural monopolies, such as gas and electrical, without any sense of the real differences in the process and patterns of development that exist in these various markets: gas and electrical lines are rarely built on spec, so that the false starts do not pose as a great a problem for determining a rate base.

These differences have genuine constitutional significance. As noted earlier, many patents have close substitutes. This makes any form of rate regulation highly suspicious since in many instances the firms are already earning only the competitive rate of return. In any event, unless this D.C. statute supplies some guarantee of an appropriate "bottom line" rate of return, consistent with the risk of the industry, then it runs into serious due process and takings objections. The baleful influence of this and similar proposals has already taken its toll on the pharmaceutical industry, whose stock prices and market capitalization has fallen dramatically in the last several years. (Remember that patents have short shelf lives, so that the profits today are not capitalized indefinitely.) No one could claim that all of the distress of the major pharmaceutical houses is a function of their exposure to price controls. Other factors include the prospect of huge liabilities and the evident decline of the blockbuster drug model, by which huge sales from a few key products are responsible for overall profits. But despite the interplay of other factors, any consistent threat to the ability of companies to retain pricing freedom over their products will necessarily contribute to long-term malaise. In the world of patents, companies have to perform first, and the government has to follow suit by enforcing the promise of patent protection. We do not want to learn through experience about the negative consequences flowing from governmental defection from its half of the patent bargain.

The Excessive Pricing Act, moreover, carries with it an enormous threat to innovation, because its influence is not confined to the Dis-

---

38 For commentary, see, for example, Amy Barrett, Pfizer's Funk, BUS. WEEK, Feb. 28, 2005, http://www.businessweek.com/magazine/content/05_09/b3922001_mx001.htm.

Big Pharma's sales are expected to grow by a sluggish 2.2% annual average through 2010, according to Datamonitor PLC. Pfizer's top line over that period will actually decline, by an estimated 1.5% a year. While Pfizer has posted strong earnings growth during Chairman and Chief Executive Officer Henry A. "Hank" McKinnell Jr.'s four-year tenure, much of that has come from cost-cutting in the wake of big acquisitions. That has caused investors to flee the stock, which has dropped 45% since he became CEO.

Id.
istrict of Columbia. The situation is such that if the bill comes to pass, customers from all states will seek to purchase their drugs through the District, which will, in turn, raise the same questions with respect to parallel importation discussed above. Could the drug companies restrict sales for local use, or could they be compelled to sell by further legislation? Yet even that might not take place if other states, feeling similar dangers, adopt the same strategy. Were this to happen, the combined price reductions would place an enormous dent on returns, and, hence, on the very incentives that justify the patent system in the first place.

The Act also raises obvious questions of whether it improperly negates the pricing authority that federal patent law normally gives to patent holders. It also raises questions of whether its limitation on the power of sale counts as a taking of the property in question. In addition, these forced price cuts raise serious questions under the TRIPs agreement. The justifications for lower prices put forward in the District of Columbia’s proposal cannot amount to the kind of public health crisis that might allow the United States to disregard the patent protection as a contracting party to the TRIPs agreement. That prices are higher than what some people are willing or able to pay cannot be grounds to deal with TRIPs because differing people’s ability to pay is a permanent condition of social affairs, not a once-in-a-lifetime scare, like a full-fledged anthrax attack.

Fortunately, in this instance it was possible to sidestep thorny issues that are likely to recur in the future. The statute was struck down on a collateral ground that avoided these larger questions of principle. The prohibition in question made it “unlawful for any drug manufacturer or licensee thereof, excluding a point of sale retailer, to sell or supply” that patented drug. The obvious political motivation of the italicized phrase was to immunize local sellers from a political thrust that was aimed at out-of-state manufacturers and their licensees, which led the District Court to strike down the statute on the grounds that it is preempted by federal law because “the D.C. Act, as drafted is a clear obstacle to the accomplishment and execution of the purpose and objectives set by Congress in passing federal patent laws relating to prescription drugs.” For good measure, the District Court held that the D.C. Act impermissibly sought to exercise its powers

---

39 TRIPs agreement, supra note 14, art. 3 ¶ 1.
40 Id. art. 8 ¶ 1.
beyond its jurisdiction as applied to commerce that took place solely outside its borders. 43

Neither of these arguments required the slightest modification of previous law, and it is a sobering commentary that the Council of the District of Columbia would choose to pass a statute with such obvious constitutional infirmities. In one sense, however, the D.C. Act was the soul of moderation, because earlier version of the legislation that would have allowed its Department of Health to declare a drug emergency with respect to patented drugs, after which the Department could issue licenses for generic production for sale within the District of Columbia or any other state that adopted similar generic licensing provisions. 44 Any proposal of this sort runs into the obvious difficulty that its patent negation is a matter of federal, rather than state law, so that the entire scheme would likely be preempted by federal patent legislation. And even if it were not, the decision to allow others to manufacture in defiance of an exclusive right to produce should be condemned as a per se taking under the Fifth Amendment. 45

This last point, which is also likely to arise again in the future, requires some explication because standard Supreme Court doctrine reserves the description of a per se taking to the physical occupation of land, 46 (which is not possible in the case of an intangible) while applying a much more complex balancing test for statutes and regulations that only restrict various forms of land use. 47 But this initial proposal for generic production goes far beyond restrictions on prices that could otherwise be charged. So long as others are allowed to use the product during the patent period, the patent holder is denied the right to recoup his front-end investment. This case, therefore, looks like one in which I am allowed to remain on my farm, but everyone else is also allowed to enter against my will. The imposition of that legal regime counts as a clear taking because there is a loss of all exclusive rights, 48 even if the original owner is allowed to share with others. And the same logic should apply with equal force to intangible

44 See supra note 34 and accompanying text.
46 Loretto v. TelePrompter Manhattan CATV Corp., 458 U.S. 419, 430 (1982) (explaining that "permanent occupations of land . . . are takings even if they occupy only relatively insubstantial amounts of space and do not seriously interfere with the landowner's use of the rest of his land").
48 Kaiser Aetna v. United States, 444 U.S. 164, 179-80 (1979) (holding that the "right to exclude," so universally held to be a fundamental element of the property right, falls within this category of interests that the Government cannot take without compensation").
forms of property that are equally protected against confiscation under the takings clause.

IV. NETWORK MONopolies: TELECOMMUNICATIONS

In terms of product line, few industries could be more disparate than the pharmaceutical and telecommunications industries. But despite the differences in their products, their uneasy relationship to state power is remarkably similar. Competitive solutions do not work with either patents or network industries. In both settings the key firms have an element of monopoly power. In telecommunications, the monopoly resides (or more accurately, resided) in the occupation of the local exchange carrier (LEC), through whose facilities all landline calls must (or had to) be made. In the patent case, the constraints on monopolies come through limitations in the duration and the scope of the patent. In contrast, with LECs the most notable constraints come through the various duties to interconnect, to supply unbundled network elements to rivals, or to engage in universal service.

The nature of the quid pro quo, however, is incidental to the main theme of this Lecture, which is the sequence of performance in situations where firms arguably have monopoly power. Telecommunications companies have to install their networks today and recover the installation cost over time. Once installed, they may well have monopoly power in certain markets—a proposition that was far truer in 1996 when the Telecommunications Act was passed than it is today—and that, in turn, raises the question how government regulation responds to that power. Will it allow for the appropriate rate of return? Or will its actions amount to a form of confiscation? To be sure, the questions here have some manifest technical differences. Interconnection obligations, which are key to the telecommunications industry, have no obvious parallels in patent law, just as the role of limited patent protections has no obvious parallel in telecommunications law. But in both cases, the creation of a legal monopoly is justified. With patents, it is to encourage innovation, whereas with telecommunications, it is to prevent the redundant duplication or destructive Balkanization of the basic network.

49 For my more detailed analysis of the telecommunications issue, see Richard A. Epstein, Takings, Commons, and Associations: Why the Telecommunications Act of 1996 Misfired, 22 Yale J. on Reg. 315 (2005).

To see how this all plays out, put yourself back in the mindset of 1900 or even 1985. In those days, there was only one way to connect any particular home to the communications network, and that was to bury a wire that went from that home to the central facility and connected it to the rest of the network. This was called the "last-mile" problem, in light of the obvious redundancy of running two or twenty wires to any given subscriber. It was this hard-wired feature that led to the system of cradle-to-grave regulation that characterized the AT&T monopoly before its 1982 judicial breakup. The reason this industry became a monopoly is because it was extremely expensive to run a second wire after the first. In addition, since the second wire ran to a different exchange, it would be necessary to connect the two exchanges, one to another, otherwise people who were on exchange one could not call people who were on exchange two.

Everybody in the telecommunications business knows the vital importance of the network effect, which is the positive correlation between the value a subscriber derives from joining a given network and the number of people that subscriber is able to reach. If the subscriber can only reach four other persons by walkie-talkie, the network is of limited value. If the subscriber can reach anyone with a telephone anywhere in the world, the value is far greater. The basic objective of any system of telecommunications is to make sure that this last goal is reached at the lowest possible cost. Anyone in the world with a phone should be able to link up with anyone else with a phone.

How is this to be done? Earlier I spoke of the patent bargain. Now it is time to speak of the telecommunications bargains that allows for the creation of that unified network after the break up of the single Bell system. What the government says to a given telephone company is this: we will give you a legal monopoly in a given territory to induce you to build this network; in exchange you will accept some degree of regulation once that network is in place. Why should we have such a regulation? Because unlike the patent system, the government cannot use time limitations to control this monopoly, and unlike the patent situation, the government never thought there would be another LEC to come along and erode that initial monopoly.

So the quid pro quo is the company builds it, the government gives the company a legal monopoly in a given territory, but in exchange it subjects the company to regulation, the extensible purpose of which is

---

51 The present system that rose from the ashes of the monopoly's breakup is characterized by regional Bell operating companies, providing local service that is supplemented by a separate set of long-distance carriers, such as the spun-off AT&T.
to bring the rate of return that the company receives down to a competitive level. It is as though we relive the ancient myth of Scylla and Charybdis. We must avoid the risk that the firm will claim monopoly profits from its position. Yet, we must also avoid the opposite peril, that the rates allowed will be so low, or on terms so onerous, as to confiscate the initial investment of the firm.

Every serious student of regulation asks, "What system of regulation manages to avoid these two perils simultaneously?" Taking aggressive steps to solve the first problem necessarily creates or exacerbates the second problem. Essentially, in the pre-1996 era, the government used the system of direct rate regulation, which made a certain amount of sense. In telecommunications, one feature, which is not present with a lot of other industries, including pharmaceuticals, is a long-term declining cost structure: everyone knows to a moral certainty that the price of executing a phone call tomorrow is going to be cheaper than it is today. This is an industry in which all technology moves in one and only one direction, so a simple but effective form of rate regulation is to look at the prices today, increase them three percent for inflation, and decrease them three percent for ordinary technological improvements. The result is steady rates under a system of price caps—a very easy system.\(^5\) You can tweak it in one way or another, but the problems are not all that acute. And it is far easier than the traditional systems of rate regulation, which seek to set a rate base of invested capital used in the business on which the appropriate rate of return could be established.

The 1996 Telecommunications Act sought to avoid the difficulties of rate regulation by introducing a form of competition between rival phone companies.\(^5\) But the ceaseless trumpeting of the virtues of competition should not be allowed to conceal the enormous difference between competition by way of ordinary markets, on the one hand, and competition over a network on the other. In the former, the sellers are wholly independent from one another, so that the survival of the one does not depend on the survival of its rivals. In the latter, the fortunes of "competitive" firms are interdependent: multiple firms can compete only if some institutional arrangements make it certain that traffic originating on one network can be carried onto another network and vice versa.

At this point, we face one of the great challenges of economic organization. The cooperation needed between competitive firms in a

\(^5\) For discussion, see, for example, Nat'l Rural Telecom Ass'n v. FCC, 988 F.2d 174 (D.C. Cir. 1993).

\(^5\) Telecommunications Act, supra note 50.
network industry necessarily requires some profound deviation from the pure competitive model in which isolated firms do their own thing. So what form of government coercion is the most effective and least intrusive way to secure network integration? The menu of possible solutions is highly technical and is likely to numb any popular audience. But perhaps it will stoke some interest in the uninitiated to say that the mistakes here have probably been worth a trillion dollars over the nine- or ten-year period the statute has been in effect because of (until late) our unerring instinct to reach for the wrong solution. The problem here is not a logical one, because there is no solution that does not carry with it some heavy costs. But it is an empirical question, capable of intelligent estimation, as to which set of costs is higher.

On this point, there are two alternatives: interconnection and the sale of unbundled network elements, or UNEs. The first of these

54 See, e.g., Nat'l Cable & Telecom. Ass'n v. Brand X Internet Servs., 125 S. Ct. 2688 (2005) (holding that the FCC acted within proper limits when it classified cable-based broadband connections as "an information service" not subject to mandatory FCC regulation, as opposed to a "telecommunications service" that was subject to mandatory FCC regulation). After Brand X, the FCC then declined to exercise its regulatory oversight. The removal of this restraint prompted an upward surge through the industry. Following on the heels of this decision, on August 5, 2005, the FCC classified all telephone-based DSL broadband connections as "information services." See David Ewalt, FCC Chair Martin Deregulates DSL, FORBES.COM, Aug. 5, 2005, http://www.forbes.com/2005/08/05/fcc-dsl-internet-cx_de_0805autofacescan11.html.

55 For this definition see 47 U.S.C. § 251(c)(2) (2005):
   (c)(2) Interconnection. The duty to provide, for the facilities and equipment of any requesting telecommunications carrier, interconnection with the local exchange carrier's network—
   (A) for the transmission and routing of telephone exchange service and exchange access;
   (B) at any technically feasible point within the carrier's network;
   (C) that is at least equal in quality to that provided by the local exchange carrier to itself or to any subsidiary, affiliate, or any other party to which the carrier provides interconnection; and
   (D) on rates, terms, and conditions that are just, reasonable, and nondiscriminatory, in accordance with the terms and conditions of the agreement and the requirements of this section and section 252 of this title.

   (c)(3) Unbundled access. The duty to provide, to any requesting telecommunications carrier for the provision of a telecommunications service, nondiscriminatory access to network elements on an unbundled basis at any technically feasible point on rates, terms, and conditions that are just, reasonable, and nondiscriminatory in accordance with the terms and conditions of the agreement and the requirements of this section and section 252 of this title. An incumbent local exchange carrier shall provide such unbundled network elements in a manner that allows requesting carriers to combine such elements in order to provide such telecommunications service.
   (d)(2) Access standards. In determining what network elements should be made available for purposes of subsection (c)(3) of this section, the Commission shall consider, at a minimum, whether—
   (A) access to such network elements as are proprietary in nature is necessary; and
JUSTIFIED MONOPOLIES

says that competition is wonderful and any new entrant that wishes to
develop its own system is free to do so. The cost of duplication is the
price of admission. The state assistance for the new entrant comes in
the form of an obligation on the part of incumbent carriers to allow
the new entrant to interconnect with the existing network on nondis-
criminatory terms. That obligation is reciprocal, so that each new en-
trant has to allow its own customers to reach the customers of all
other carriers. Indeed, under this rule each of two network entrants
has an interconnection obligation to each other.

Focus for the moment on the connection between the new and es-
tablished carrier. At the beginning of this relationship, the new entrant
receives a large, but implicit subsidy. If it carries only 1 percent of the
traffic, it now has guaranteed access to the other 99 percent of system
users. The incumbent carrier with 99 percent of the market only gains
new access to the 1 percent of hard souls who choose to do business
on the new network. Yet this imbalance is offset, in part, by the dif-
ferential costs of complying with the interconnection obligation that is
likely to be higher on a per capita basis for the incumbent. But note
that the situation is largely self-correcting. As the new entrant gains
market share, the imbalances on both dimensions fade: the access to
the new network is worth more to the incumbent, and the cost of po-
licing the interconnection obligation for the new entrant falls on a per
customer basis.

When the basic system was under consideration in the pre-1996
era, everyone put on their best technological blinders. The general
view, which I shared, was that this last-mile problem would remain a
serious obstacle to competition for a very long period of time. Setting
about to overcome the problem had a certain sense of urgency be-
cause of the perceived high cost of a duplicate network. Hence, there
was much pressure to "jump-start" the competitive process by finding
some alternative to the "facilities-based" competition that is contem-
plated by the interconnection approach. 57

The method hit upon to achieve that goal was to develop a scheme
that allowed for the forced purchase of UNEs at a price set by the
FCC in cooperation with the state commissions that retained some
uncertain authority of the regulatory process. 58 The task of buying and


58 See AT&T Corp. v. Iowa Utilities Bd., 525 U.S. 366 (1999) (holding that the FCC had
rulemaking authority to promulgate and enforce the price regime).
selling these components is far more daunting than it sounds, for an ordinary switch can have seven or more tiny components that, when put together, will allow you to actually make the call. The outsider, or competitive local exchange carrier (CLEC), essentially has an option to buy one or more of these from the incumbent local exchange carrier (ILEC) and build the others himself or purchase them from third parties. This statutory option is exceedingly valuable no matter how the costs are ultimately figured: the CLEC can exercise the option when it is cheaper and decline the option when it is more expensive.

In effect, with the compulsory purchase regime, a CLEC can form an ideal network that combines all his own low-cost elements with all the low-cost elements of the incumbent. In principle, that competitor should do better than the incumbent, but there are a series of obstacles that come his way. What happens is that the regulator now faces the second problem: how should it price these components in the absence of any pricing information that could be gleaned from voluntary markets?

The task here, in effect, is to use a version of the eminent domain power to order the transfer of some network elements from A to B. There are no situations in which forced exchanges work as well as voluntary ones. But for these purposes, the central insight is that the eminent domain system functions more easily in some settings than it does in others. The best proposition on the effectiveness of the eminent domain power can be neatly summarized in a single sentence: it works well for big plots of land with high values that can be taken over in single transactions. If the government wants to condemn land for a highway, the public use is allowable under any interpretation of the takings clause. On the other hand, when the government starts running condemnations at the microlevel between private carriers, matters become somewhat more complex, even if we overlook any latent issues with the public use requirement. The transactions cost for per unit of condemnation becomes enormous.

The basic problem was compounded when the FCC, under Reed Hundt, took an aggressive position on the pricing question to increase the jump-start that the ILECs could receive. The device that was hit on was the decision to price the transfer of individual UNEs on the assumption that their historical cost—the amounts actually paid—was

---

59 For a longer version of this account, see Epstein, supra note 45.
60 The broad expansion of that clause was announced by the Supreme Court in its five-to-four decision in Kelo v. City of New London, 125 S. Ct. 2655 (2005) reh'g denied 74 U.S.L.W. 3113 (2005).
61 For a sharp criticism, see Peter Huber, Telecom Undone: A Cautionary Tale, 115 COMMENT. MAG. 36 (Jan. 2, 2003).
irrelevant to the issue. What really mattered was the cost that would have been required, counterfactually, to construct each individual unit on the strength of what has become known as TELRIC pricing—total element long-run incremental cost—on the assumption that the incumbent had made, as of the moment of this decision, all optimal decisions on network configuration, using the best available technology at the time. In effect, all risk of error for the system design fell on the incumbent, who was not, however, given any higher implicit rate of return (in the form of higher network element prices) to compensate for the risk that it had to bear. The gist of the system was that it would not be possible for any ILEC to recover its historical cost of putting the network together if the CLECs (each one acting individually) had decided to appropriate each and every network element.

The net effect of these developments was that the transfer from the old historical cost systems under rate regulation to the newer TELRIC pricing left the incumbents bearing the full risk of the new system, a problem that was compounded because many of the historical costs had, under the traditional systems of rate regulation, been postponed to later periods in order to reduce the interim rates. But the promised increments in future years were negated by the TELRIC regulations. Ironically, however, it is not clear that the systematic price break for UNEs do any good for the new entrants. The FCC rules allow anyone to enter the market, so that no individual CLEC that gets a price break has an advantage over any other CLEC. In effect, therefore, the subsidy is peeled away from each of them. The tough competition among CLECs led to high failure rates. At the same time, it imposed heavy financial costs on the ILECs, who were, in general, required to sell the UNEs to the CLECs on the strength of unsecured credit (much of which became worthless in bankruptcy). The ILECs had to become the involuntary creditors to their competition.

Yet, after much tribulation, and countless false starts, it turned out that, in the last lap, the ILECs won the war while losing the pricing battle. How? The tests for selling UNEs did not only have the pricing component, which raised the conflict between historical cost and TELRIC. The language in the UNE section also required that incumbents could be forced to enter into these transactions only in some limited set of circumstances. Thus, the statute provided that "the Commission shall consider, at a minimum, whether . . . the failure to provide access to such network elements would impair the ability of

---

62 Id. TELRIC is a means of pricing a network element.
the telecommunications carrier seeking access to provide the services that it seeks to offer." The problem was that nobody knew (or knows) what "impairment" meant. One possibility is that the CLEC is impaired whenever it is required to pay more money to buy something somewhere else. Yet that cannot be right, for then the requirement has no bite at all. In the 1999 AT&T v. Iowa Utilities Board decision, the Supreme Court invalidated the initial round of "impairment" regulation and told the FCC to try again. But after several shots, the FCC had not improved the test enough to satisfy the Circuit Court for the District of Columbia. As a result, once the Solicitor General decided (after extensive political lobbying on all sides) not to seek Supreme Court review of the decision, the FCC's new regulations were invalidated, which in turn dammed up the stream of new UNE purchases. In the end, therefore, the pricing battle essentially was won by the new entrant but the access battle was won by the incumbents. The price does not matter if you get access to nothing.

The question here, of course, should involve more than political wrangling for partisan advantage. How we decide to form networks out of separate forms is a matter of first principle. Recall that the implicit assumption of the 1996 Telecommunications Act was that the last-mile problem was regarded as a long-term technological obstacle, and not without reason. Remember that in 1996 there were many more landlines in the market than cell phones, which, at the time, were regarded as an added and expensive extra. This is an indispensable element. The first thing you have to do is tell the FCC not to mess up with the cellular system because you can connect to a cell phone user without building any miles at all, unlike traditional landline phones. For those who care about the numbers, there were a bit over 44 million cell phone subscribers in 1996, and as of June 2005, that number had reached 194 million. Yet the number of landlines in the United States has dropped in recent years, from about 191.7 million in 2001 to about 181.4 million in 2003. At present, there are more cell phones in the United States than landlines, and nothing will reverse that trend in the foreseeable future. For so many people—young singles who travel on business, construction workers...
who operate in the field, students constantly on the go—a single portable phone is far more preferable to keeping tabs on a dozen different landlines. There is, therefore, at least one viable portal to every person, through the creation of a second network that in no sense duplicates the first. In addition, the rise of VoIP (voice-over-internet protocol) telephony, and the potential use of cable and electrical networks for phone transmission also offer new ways to erode the local monopoly of the original ILEC. The lesson here is too important to overlook. Changes in technology blur the lines between different industries, and in so doing, undermine the market power of established firms. The 1996 Telecommunications Act came just before these new technologies bore fruit. If matters had been delayed perhaps two or three years, Congress could not have adopted a system of regulation that presupposed the permanence of that last-mile issue. Viewed with hindsight, then, it seems clear that those individuals—and there were not many—who thought that the only appropriate response to the network problem was to require interconnection, had it right. The efforts to force-feed competition proved to be both expensive and ruinous. It turns out that it makes a difference which form of regulation is used when competitive solutions are not possible—a whole lot of difference.

V. CONCLUSION

This discussion of the monopoly problem as it relates to pharmaceutical patents and to telecommunication networks raises questions that must be answered in order to determine the appropriate pattern of interaction between the creation of a monopoly on the one hand and its regulation on the other. The state, for its part, is constantly worried about the abuse of monopoly power, and no one can deny that it is a legitimate fear. Structuring sensible terms for these monopolies takes a good deal of thought. But the job is not finished once that initial task is discharged. Equally important is the question of sequential performance, with which I started this Lecture. The rules of the game set out a bargain between the regulated firm and the state. The firm must perform first, and, therefore, its success is wholly dependent on the willingness of the state to keep to its half of the bargain, even when the state has every incentive to renege on its promises. The pattern of state defection can differ as we move from one institutional setting to another, because the justifications for monopoly are not the same in all contexts. Yet in the grand scheme of things, these variations matter relatively little. In all cases, the long-term effects of state
defection have to be deleterious because they create massive confusion in the short-run and undermine the incentives to invest in the long-run. To control these tendencies, it is critical to adopt simple rules of the game so that all know what is expected over the life cycle of the monopoly. For patents, especially in pharmaceuticals, this includes clear terms and general pricing freedom. For telecom carriers, it includes interconnection obligations, and little else. The issues raised in these cases may well be technical, but their sound resolution is key to the prosperity and well-being of us all.