Inconsistency in Antitrust

Ramsi A. Woodcock

Inconsistency in Antitrust

RAMSI A. WOODCOCK*

I. INTRODUCTION

When the price of a good is too high, consumers who can afford to pay cost, including enough profit to make production worth the manufacturer's while, but cannot pay enough to meet the high price, are forced to do without. Economics teaches that efficiency would increase if price were to fall to cost because at cost the manufacturer would still be glad to produce and consumers could now afford to purchase more of the good. Efficiency requires that where more for less is possible, more must be had for less.

If the point of antitrust is to maximize efficiency, then the point of antitrust is to drive prices down when they are too high. It follows that antitrust should make the charging of a monopoly price illegal, at least where such a price is not necessary to cover costs. Strangely, antitrust does this for cartels but not for monopolies. Cartels that fix inefficiently high prices are illegal (often they are illegal regardless even of the prices they charge), but courts and commentators have long been at pains to stress that a monopoly does not violate the law simply by charging an inefficiently high price.

Antitrust has typically explained this hole in the law in three ways. First, it has argued that market forces compensate for the absence of antitrust in this area (the "entry argument"). Antitrust currently prohibits monopolies from taking affirmative steps to exclude competitors, such as tying, exclusive dealing, or predatory pricing. So long as a monopoly does not engage in such prohibited conduct, competitors will enter whenever the monopoly charges an inefficiently high price and drive price back down. Antitrust therefore does not need to intervene. Second, antitrust has argued that monopolies are more efficient than cartels (the "efficiency argument"). Monopolies have higher costs because they spend more on research and development ("R&D"), so they need to be able to charge higher prices. More cannot be had for less. Third, antitrust has argued that even if the hole is a real problem, there is nothing that antitrust can do about it (the "breakup argument"). Breaking up cartels is cheap (just dissolve the cartel agreement); breaking up monopolies is expensive.

This article shows that these are flawed arguments. The efficiency argument is flawed because, among other reasons, even if we accept that monopoly profits finance R&D that leads to innovation, it does not fol-

low that this relationship does not hold for cartels as well[1] or that there is no longer such a thing as too high a price.[2] The breakup argument is flawed because, among other reasons, breaking up monopolies means breaking up essential inputs, not the more daunting task of trying to reallocate all of a firm's inputs between competitors by fiat. The market sorts that out.[3]

The entry argument fails because it does not recognize that monopolies can exclude competitors passively, simply by using property rights to deny competitors access to essential inputs, without having to engage in affirmative acts of exclusion. There is no reason to think that in the absence of affirmative acts of exclusion, the market will always naturally respond to inefficiently high pricing by ramping up output and driving price back down.

This article proposes input control as a unifying concept in antitrust and shows how existing categories in the field can be mapped onto various legal and extra-legal forms of input control.[4] A necessary condition for monopoly power, whether exercised by a cartel, monopoly, or other entity, is control by the entity over inputs that competitors might otherwise use to enter the market. Without such control, competitors can use the inputs to enter and drive price down. From the perspective of input control, cartels and monopolies differ only in the legal form of their control over inputs: cartels use contract to control inputs (by using contract to control the inputs' ultimate owners), whereas monopolies use property to control them. When a cartel forms, it uses contract to centralize control over inputs in the hands of a single competitor; when a monopoly forms, it uses property to do the same thing. In this sense, collusion and exclusion are the same. You might say that collusion excludes, or exclusion colludes.

This article argues that antitrust breaks up overcharging cartels but not overcharging monopolies because it harbors a bias in favor of property-based exclusion. A cartel can charge a monopoly price because, via contract law, it controls enough inputs to prevent competitors from entering to drive price down. In breaking up cartels, antitrust in effect recognizes that this contract-based control is exclusionary. A monopolist can charge a monopoly price because, via property law, it controls enough inputs to prevent competitors from entering to drive price down. It is a major blindspot in antitrust that it fails to recognize that such property-based control is equally exclusionary.

---

1. *See infra* Part II.
2. *See infra* Part III.
3. *See infra* Part V.
4. *See infra* Part IV. Chart 2 summarizes this mapping.

The doctrinal expression of antitrust's bias in favor of property-based exclusion is the refusal to deal doctrine, which, with some small exceptions, guarantees the right of firms to refuse to share their resources with competitors, regardless whether this leads to inefficiently high prices. Plugging the hole in antitrust means doing away with the general rule that firms have a right to refuse to deal. This has important consequences for intellectual property ("IP") law. The bias in favor of property-based exclusion has prevented antitrust from coming into direct conflict with IP in cases in which firms use IP to exclude. Without the bias, antitrust can take on the role of preventing inefficiently high pricing by IP-based monopolies.[5]

The unjustified favor antitrust accords monopolies relative to other forms of input control is expressed in other areas of antitrust in addition to refusal to deal. Antitrust regulates monopolies only when they form through merger or are created or maintained by affirmative acts of exclusion other than the mere exercise of property-based control over inputs. In the case of affirmative acts (what antitrust calls "exclusionary conduct"), antitrust applies a balancing test that approves all acts that increase efficiency, regardless whether they maximize it. But many acts that increase efficiency do not maximize it. As a result, current law approves of many more affirmative acts of exclusion than would be permitted under an efficient antitrust regime.[6] In the case of merger, antitrust scrutinizes merger to natural monopoly for efficiency. But it fails to break up preexisting natural monopolies even though the efficiency rationale for doing so is identical to that associated with preventing mergers to natural monopoly.[7]

The article is organized as follows. Part II introduces the notion that there is a hole in antitrust enforcement, calling it a per se rule in favor of monopolization. It argues that the hole is expressed in the doctrines of the conduct requirement and refusal to deal and that refusal to deal amounts to an exemption for property-based exclusion. Part II also introduces the notion that collusion and exclusion are conceptually identical for antitrust purposes and argues that there is no basis for treating cartels and monopolies differently. Part III argues that the need for rewards does not eliminate the need for antitrust to constrain monopoly power. Part IV introduces the metaphysics of input control and uses it to give a more precise statement of some of the claims in Part II. Part IV also maps current law onto the input control paradigm. Part V attacks a pillar of the refusal to deal doctrine by arguing that the cost of breakup

---

5. *See infra* Part VI.
6. *See infra* Part VIII.
7. Part VII discusses this and other inconsistencies.

need not be excessive as a rule. Part VI explores the consequences of the demise of refusal to deal for IP. Part VII identifies additional inconsistencies in antitrust enforcement, including its approach to natural monopoly.[8] Part VIII critiques the balancing test.

II.   A First Look at Inconsistency in Antitrust

A.   *Pro Monopoly and Anti Cartel, Without Basis*

It is apparent to any careful student that, as far as textbook economics is concerned, the organizational structure of firms should be irrelevant to antitrust. Textbook antitrust tells us that the goal of an efficiency-oriented antitrust is to eliminate the black triangle in Diagram 1 on page 110.[9] The socially optimal price is C. If the price in the market is P, then consumers cannot afford goods corresponding to QQ' even though they are willing to pay more than their cost of production in order to obtain them. The black triangle represents the value consumers lose from not being able to enjoy these goods, which also happens to be the loss to society generally from this underproduction. That is all textbook economics has to say about efficiency and antitrust.[10]

---

8. One inconsistency that is not discussed at length in this article is the exemption for collusion when organized as oligopoly but its condemnation when organized as cartel. This inconsistency has been much discussed elsewhere. *See, e.g.,* RICHARD A. POSNER, ANTITRUST LAW ch. 3 (2d ed. 2001).

9. The simple model described by Diagram 1 ignores economies of scale and product and process improvements. Improvements are dealt with in Part III. Economies of scale and the problem of natural monopoly are dealt with in Parts II.B, IV.A (including, in particular, note 87 therein), and VII.B.

10. *See* William J. Baumol, *Horizontal Collusion and Innovation,* 102 ECON. J. 129, 129 (1992) ("The familiar textbook description of the social cost of monopoly applies also to horizontal collusion."). Textbooks do not generally make this point explicitly. Instead, it is to be inferred from the fact that textbooks usually only provide a single model of market inefficiency (e.g., Diagram 1). The reader is left to infer that any inefficient organizational form must be inefficient in the style of that model. This is starkest in the way textbooks generally treat monopolies and cartels. The textbook will explain inefficiency in the context of a monopoly, and then subsequently remark that cartels have power over price as well. The reader is left to infer that cartels are therefore inefficient in the same way. *See, e.g.,* DENNIS W. CARLTON & JEFFREY M. PERLOFF, MODERN INDUSTRIAL ORGANIZATION 95–96, 122 (4th ed. 2005) ("A cartel that includes all firms in a market is in effect a monopoly . . . ."); F.M. SCHERER & DAVID ROSS, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE 235, 661–67 (1990); HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY 19–21, 158 (4th ed. 2011) [hereinafter HOVENKAMP, FEDERAL ANTITRUST POLICY]; HAL R. VARIAN, INTERMEDIATE MICROECONOMICS 431–33, 438 (7th ed. 2006).

The point is also captured in Richard Posner's call for the repeal of all antitrust laws other than Sherman Act Section 2 because "[i]f by 'monopolizing' we were to mean simply conduct that unjustifiably promotes supracompetitive pricing, we would cover all practices . . . that pose a threat to the maintenance of competition in the economic sense." POSNER, ANTITRUST LAW, *supra* note 8, at 260.

DIAGRAM 1

P

C

Q   Q'
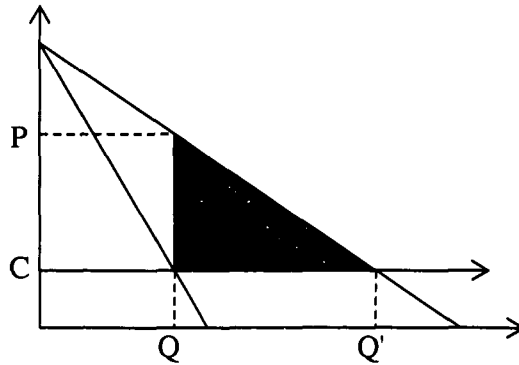
Textbook economics has no interest at all in how price might have ended up at P instead of C. Price might get there because independently owned firms agree to charge the same price. Or because for whatever reason there is only one firm in the industry and it chooses to charge P. Or because gangsters threaten to kill anyone who does not sell at P. Or because all the firms in the industry have delegated pricing authority to an independent trustee who has chosen P.[11] Or because all the firms in the industry are community-oriented folk interested in being good neighbors to the other firms in the industry rather than competing hard with them on price.[12] Or it might get there by accident.

The point is that if the goal of antitrust is to make the black rectangle go away,[13] antitrust does not care whether it got into Diagram 1 through cartelization, monopolization, gangsterism, neighborliness, or

11. *See* HERBERT HOVENKAMP, ENTERPRISE AND AMERICAN LAW, 1836–1937, at 249–50 (1991) (describing the operation of a stock-transfer trust).

12. In this article, "oligopoly" means not just textbook models in which price ends up somewhere above C and below P, but any kind of monopoly inefficiency due to the behavior of multiple firms that are not explicitly communicating to coordinate their actions (i.e., those that are not cartels). This includes oligopolies that can charge P. Economics knows that any price can be an oligopoly equilibrium. DAVID M. KREPS, A COURSE IN MICROECONOMIC THEORY 525 (1990). "[E]xogenous norms of cooperative behavior" determine which equilibrium obtains, and the right norms (neighborly norms) can get you to P, even in a relatively unconcentrated industry. Albert O. Hirschman, *Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble?*, 20 J. ECON. LITERATURE 1463, 1470 (1982). In my view, "oligopoly" used in this way corresponds to what antitrust lawyers mean by "tacit collusion." *See generally* CARLTON & PERLOFF, *supra* note 10, at 127 n.8 (discussing collusion terminology).

13. This article assumes that the goal is total welfare (producer plus consumer), as opposed to just consumer welfare. However, the consensus in antitrust is that consumer welfare is the standard. *See* Richard Schmalensee, *Thoughts on the Chicago Legacy in U.S. Antitrust, in* HOW THE CHICAGO SCHOOL OVERSHOT THE MARK: THE EFFECT OF CONSERVATIVE ECONOMIC ANALYSIS ON U.S. ANTITRUST 11, 13 (Robert Pitofsky ed., 2008). As a rhetorical matter, this article wishes to make the argument that even the strongest case for monopolization is flawed. Because the total approach makes the strongest case for the efficiency of monopolization, the article therefore assumes the total approach.

accident. It just cares that it is there.[14]

It turns out, however, that antitrust law discriminates heavily based on how an industry gets to P. The bias is strongly in favor of punishing cartels and gangsters, and ignoring inefficiency associated with oligopolies (e.g., good neighbors) and monopolies. The bias is extraordinarily pronounced in enforcement practice, the case law, and scholarship.[15]

If textbook efficiency does not account for the difference, what does? The argument in the literature is that you often can have P without

---

14. Likening monopolization to the creation of an involuntary cartel, Jonathan Baker observes that "[i]t does not matter to buyers whether the cartel is voluntary or involuntary; either way, the same firms collectively reduce output and the price that buyers pay increases." Jonathan B. Baker, *Exclusion as a Core Competition Concern*, 78 ANTITRUST L.J. 527, 557–58 (2013).

15. Gangsters. State criminal law prohibits the use of violence, threats, or theft, regardless whether used to restrict industry output. *See, e.g.*, MODEL PENAL CODE § 211.1 (assault); *id.* § 223.2 (theft). Federal criminal law also reaches gangsterism relevant to monopoly inefficiency. *See, e.g.*, 18 U.S.C. § 1951 (2006) (imposing criminal penalties for use of threats or violence to obstruct commerce); Ray V. Hartwell III, *Criminal RICO and Antitrust*, 52 ANTITRUST L.J. 311, 312–19 (1983) (discussing application of RICO to prosecute attempts to raise prices).

Cartels. The courts read the Sherman Act to prohibit per se all agreements between competitors with the purpose or effect of raising price, meaning that such agreements are prohibited regardless of efficiency. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 279. Even though the criminal provisions of the Sherman Act may be applied to monopolization, the U.S. Department of Justice ("DOJ") applies them exclusively to cartels that fall under the per se rule. *See* 2 ABA SECTION OF ANTITRUST LAW, ANTITRUST LAW DEVELOPMENTS 956 & n.9 (7th ed. 2012) (and sources cited therein). Cartel fines have exploded in recent years; not so for violators of the antitrust laws that have other organizational forms, because they are never prosecuted. *See Criminal Enforcement Fine and Jail Charts*, U.S. DEP'T OF JUSTICE ANTITRUST DIV., http://www.justice.gov/atr/public/criminal/264101.html (last visited July 1, 2013) (showing that total criminal cartel fines increased tenfold to $1.1 billion in the ten years ending in 2012). DOJ filed one civil monopolization case in the ten years ending in 2012 and challenged 123 mergers over the same period; but it filed nearly three times as many criminal cases against cartels (345). *See* U.S. DEP'T OF JUSTICE, ANTITRUST DIVISION WORKLOAD STATISTICS FY 2003–2012, at 4–5, 7, *available at* http://www.justice.gov/atr/public/workload-st atistics.pdf. Despite a largely successful effort by antitrust opponents starting in the 1960s to reduce the scope of the antitrust laws across the board, the Supreme Court firmly rejected attacks on the per se rule against cartels and a consensus in favor of the rule now prevails among both opponents and defenders of antitrust. *See* Arizona v. Maricopa Cnty. Med. Soc'y, 457 U.S. 332, 347–48 (1982) (affirming continued vitality of the rule); Schmalensee, *Thoughts on the Chicago Legacy in U.S. Antitrust*, *supra* note 13 (identifying Chicago School victories); Eleanor M. Fox, *The Efficiency Paradox*, *in* HOW THE CHICAGO SCHOOL OVERSHOT THE MARK, *supra* note 13, at 77, 98 n.18 (describing consensus of left and right in favor of the per se rule against cartels).

Oligopolies. Firms that raise price as a group without explicitly coordinating with each other are exempt from antitrust scrutiny. *See* Matsushita Elec. Indus. Co. v. Zenith Radio Corp., 475 U.S. 574, 588 (1986) (insisting on evidence that conspirators did not act independently).

Monopolies. As we shall see in Part II.B, much monopolization is exempt entirely from the antitrust laws. When it is not, courts impose no blanket rule but instead test for efficiency on a case-by-case basis. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 214 (acknowledging that "unilateral conduct receives the lowest level of antitrust scrutiny").

(The emphasis of enforcers on prosecuting cartels instead of monopolies might be due to factors such as efficacy of deterrence, rather than bias. *See* Baker, *supra* note 14, at 577–78. But taken together with the bias in the case law and scholarship, the unequal enforcement is certainly suggestive.)

the black triangle for monopolies, so it does not make sense to discourage monopolization.[16]

The logic goes like this. Most monopolies do not just have marginal costs; they also have fixed costs, which are the amounts you have to pay for factors of production regardless how much you actually produce or sell. Let us assume that you have a monopolist whose fixed costs exactly equal the monopoly profit (i.e., they equal the area of the box created by the y-axis, Q, P, and C in Diagram 1). Then the monopolist will go out of business unless she can charge P and earn the monopoly profit. That means that the extra goods amounting to QQ' could not be feasibly produced at a price below P and therefore the black triangle can no longer correspond to lost social or consumer value. Now P is efficient. We can stop coloring the triangle black.

The trouble with this argument is that there is no reason to think that cartels do not also have fixed costs.[17] The notion that they do not, but monopolies do, seems to arise because antitrust participants are in the habit of telling themselves ex post stories about cartelization but ex ante stories about monopolization.[18]

The ex post story about cartelization is that initially you have a bunch of firms in a competitive industry, each of which fully expects to remain in a competitive industry in the future. This expectation means that the firms all made investment decisions that assumed that they would only need a price of C to cover their costs. They did not expect to be able to charge P and they ensured that they would be able to remain in the market without doing so.

The price is at C. At some point, however, cartelization happens. Price jumps to P. We know that the black triangle must appear as a

---

16. *See, e.g.*, CARLTON & PERLOFF, *supra* note 10, at 99 (stating that the benefit of monopoly is that it allows innovative firms to cover R&D costs). Modern arguments for monopoly are traceable to the work of Joseph Schumpeter, who believed that monopoly is necessary for innovation and growth. *See, e.g.*, JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM AND DEMOCRACY (Harper Torchbooks 1975) (1942). The fixed cost argument described in this article is a generalization of one such Schumpeterian argument, which is that monopoly profit is needed to cover the costs associated with R&D. *Id.* at 87–89. Because this is the only Schumpeterian argument to have survived empirical review, it seems reasonable to focus on it in this article. WESLEY M. COHEN & RICHARD C. LEVIN, *Empirical Studies of Innovation and Market Structure*, *in* 2 HANDBOOK OF INDUSTRIAL ORGANIZATION 1059, 1060–61, 1070, 1074–75, 1078 (Richard Schmalensee & Robert D. Willig eds., 1989).

17. *Cf.* MICHAEL D. WHINSTON, LECTURES ON ANTITRUST ECONOMICS 16–17 (2006) (sketching a model in which cartelization allows an industry to cover fixed costs and relating it to the 19th century "ruinous competition" argument made by a railroad cartel).

18. I am not aware of any clear instantiation of these stories in the antitrust literature (see *infra* note 19 for an example that hints at them). But I believe that they are what antitrust scholars and practitioners think about when they think about the different treatment of cartels and monopolies.

result of this price jump because we know that all the firms could have continued to charge C and produce Q' going forward.

The ex ante story about monopolization is that initially you have an entrepreneur who is deciding whether to create a superior product that no rivals are sophisticated enough to imitate. Making better products is expensive. But she realizes that if she can count on charging P for the product, then she will be able to afford to develop it. She invests in the project fully expecting, and on the condition, that she will be able to charge P when she enters the market. The fact that her willingness to enter the market is conditional on being able to charge P implies that she faces fixed costs equal to the size of the monopoly profit at P.

When she enters the market, she charges P. No black triangle appears, however, because we know that if she had expected to be forced to charge C, she would not have entered the market in the first place and there would be no superior product for anyone to buy at any price.[19]

The ex post telling of the cartel story makes it sound like P is not necessary, while the ex ante telling of the monopoly story makes P essential. What this masks is that there is no reason to think that firms cannot take cartelization into account ex ante when deciding whether to take on high fixed costs. Similarly, there is no reason to think that monopolies always need P to cover their costs. It is perfectly reasonable to think that each member of a cartel might only produce a superior product (relative to the products of non-members) if it can expect to be able to work together with other members to fix the price at which it sells it at P.[20,21,22] Similarly, it is perfectly reasonable to think that a

---

19. I read Herbert Hovenkamp's argument that cartels deserve more scrutiny because they can be created "very quickly" whereas monopolies are the product of hard work to contain an element of this ex post/ex ante story. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 214. It seems obvious to Hovenkamp that cartels form quicker than monopolies because monopoly power "ordinarily takes many years of innovation and aggressive production and marketing." *Id.* In other words, cartels do not need planning; firms that expect to face a competitive market can turn around and enjoy P instead, whenever they wish, by cartelizing. By contrast, to charge P, a single firm must plan and fight for it.

But the fit is not precise. The argument also rests on the notion that speed of formation matters for efficiency. But Hovenkamp gives no explanation for why this should be so. What matters for efficiency is the existence of fixed costs; if cartels are as likely as monopolies to have them, then they should be protected, regardless whether they happen to form faster than monopolies.

20. The odd thing about antitrust's treatment of cartels and innovation is that while it professes to ban price fixing per se, it permits it when the price fixing is used to fund fixed costs associated with cooperative R&D. There is no other way to understand the membership fees that open-membership R&D joint ventures are permitted to charge other than as a way of encouraging members not to compete price too low (there would be no incentive to join the joint venture if the fees were thought to be ineffective at encouraging pricing discipline). *See* HOVENKAMP, FEDERAL

superior product monopolist might never have expected to win big and would have stayed in the market anyway at C.

Of course, the argument for the cartel/monopoly distinction is not that all cartels have black triangles but all monopolies do not. The argument is that cartels tend to have black triangles and monopolies tend not to.[23,24] Because of enforcement costs, the argument goes, it is more effi-

---

ANTITRUST POLICY, *supra* note 10, at 248 & n.53 (membership fees permissible for open-membership joint ventures).

The inconsistency here is that antitrust recognizes that R&D joint ventures will not come to pass unless antitrust allows them to fix prices necessary to cover fixed costs, but it does not recognize that independent innovation by cartel members may not come to pass unless antitrust allows the cartel to fix prices. The inconsistency might make sense if there were reason to believe that cooperative R&D by cartel members tends to be more efficient than independent R&D by cartel members. But I am not aware of any study that establishes this.

21. For example, the members of the DVD patent pools presumably incurred fixed costs in inventing the DVD technology for which they hold patents. *See* Complaint at 25–27, Florida v. Hitachi-LG Data Storage, Inc., No. 13-1877 (N.D. Cal. Apr. 24, 2013). Because this technology is superior to VHS or VCD technology, the members of the pool are able to exclude nonmember competitors from the market. *See id.* at 26. However, excluding rival firms is not sufficient to allow them to cover the fixed costs associated with inventing the technology. To do that, they must also avoid competing too hard with each other; otherwise they will drive prices too low to cover those costs. But in order to reduce the transaction costs associated with coordinating their pricing, it might be efficient for these firms explicitly to work together to fix prices. *Cf.* OLIVER E. WILLIAMSON, THE ECONOMIC INSTITUTIONS OF CAPITALISM: FIRMS, MARKETS, RELATIONAL CONTRACTING 159–60 (1985) (discussing progressive integration of members of telegraph cartels in the 19th century).

The analogue of this in the single firm monopoly context is the monopoly that invents DVDs on its own. In order for it to cover its costs it too must (1) exclude rival firms from the market (that would otherwise be able to drive price down) and (2) charge a high enough price to cover costs. It is not clear why antitrust allows monopolies, but not cartels, to do this.

(Perhaps an enlightened antitrust judge would not apply the per se rule to the cartel in this example on the ground that the price fixing is in service of an innovation joint venture; but imagine that the firms first agree to fix prices, and only later realize that they can maximize profit at the new higher prices by inventing the DVD. Their price-fixing agreement would certainly be per se illegal, even though this change of timing has no effect on the efficiency of the price fixing.)

22. There is some reason to think that members of an unstable cartel might be more likely than a single firm monopoly to incur fixed costs associated with R&D because each member feels compelled to prepare for the free-for-all that will accompany the demise of the cartel. *Cf.* POSNER, ANTITRUST LAW, *supra* note 8, at 21.

23. *See, e.g.*, SCHERER & ROSS, *supra* note 10, at 335. Scherer and Ross concede that they have no "solid estimates" about the efficiency of cartels but conclude that it must be "modest" because, based on a review of railroad and soft coal mining studies, "price competition does seem to have done its job in forcing 'sick' industries to shed high-cost capacity . . . ." *Id.* They do not explain why, if this is so, price competition would not do a similarly good job in forcing sick single firm monopolies to shed high cost capacity.

24. There are three kinds of antitrust rule for any particular organizational form: per se illegal, case-by-case illegal, and per se legal. Under the first, nothing goes. Under the second, it goes only if there is no monopoly inefficiency. Under the third, everything goes. There are three corresponding justifications: respectively, that the form tends to be inefficient, has no tendency either way, or tends to be efficient. The argument for the per se rule against cartels is therefore that cartels tend to be inefficient. The argument against a per se rule against monopolies is that

cient to throw the book at most cartels and throw it away from most monopolies.[25] This suggests that antitrust must rest on a deep tradition of research into the relative incidence of monopoly inefficiency in cartels and monopolies.[26] In fact, there is no such tradition and no empirical evidence that cartels are mostly inefficient and monopolies mostly

---

they either have no efficiency tendency or tend to be efficient. In Parts II.B and III, I argue that there is a per se rule in favor of monopolization that is based on the idea that monopolies tend to be efficient. It is for this reason that here I ignore the no-efficiency-tendency argument and focus on the efficiency argument.

25. *See, e.g.*, Scherer & Ross, *supra* note 10, at 336. Scherer and Ross argue eloquently for the cost savings associated with the per se rule against cartels:

> To place upon the enforcement agencies and courts the burden of ascertaining from day to day whether . . . price has become unreasonable through the mere variation of economic conditions could break the back of an already bowed and groaning camel, and it would entail a substantial increase in the degree to which government intervenes in the affairs of business. These problems are in part symptoms of a fundamental deficiency: the unsuitability of U.S. judicial processes for making balanced judgments on issues as technical and complex as the reasonableness of a price-fixing scheme.

Id. (internal quotations, brackets, and citations omitted). What is not clear from the argument is why it should not apply with equal force to justify imposing a per se rule against monopolization.

26. I am aware of two theoretical discussions of the different treatment of cartels and monopolies. Armen Alchian and William Allen argue that it is not justified. Armen A. Alchian & William R. Allen, Exchange and Production: Theory in Use 408 n.6 (1969) ("If there is something bad about collusion, is there not also something bad whenever people voluntarily pool their private wealth to form a corporation that is big enough to affect the market price by its offerings of some good? Every corporation and partnership uses jointly owned resources in wealth-maximizing ways. Why is effective collusion agreement among several businesses different from merger or new creation of a large business? It isn't.").

William Baumol argues that cartels are more inefficient than monopolies for two reasons. First, monopolies will close their least efficient plants first whereas a cartel cannot do the same if the least efficient plants are all owned by one member. Such a member "can hardly be expected to consent to their closing—leaving it with no facilities if the cartel falls apart." Baumol, *Horizontal Collusion and Innovation*, *supra* note 10, at 130. Second, the costs of the negotiations required to hit price or output targets can be extraordinarily high. *Id.*

The trouble with the second argument is that it treats single firm monopolies as suffering only from monopoly inefficiency. There are costs associated with centralized management, not just cartel management. Williamson, *supra* note 21, at ch. 6 (attempting to answer the question, "[w]hy can't a large firm do everything that a collection of small firms can do and more?"). It is reasonable to assume that in some cases cartel management costs will tend to be higher and in other cases single firm management costs will tend to be higher. We do not know, for example, that most monopoly pricing by cartels could be done more efficiently by merging cartel members.

The trouble with the first argument is of a similar nature. It is not acceptable to assume that because a monopoly "will have a strong incentive to close its most efficient plants first," it will do so. Baumol, *Horizontal Collusion and Innovation*, *supra* note 10, at 130. Inefficient operations may have powerful stakeholders within the management of a single firm. Moreover, the instability that drives the fear of cartel members is partly due to the fact that cartels are illegal. Cartels, like monopolies, have a strong incentive to close inefficient plants. We cannot say a priori that the management costs associated with achieving this outcome are lower in monopolies than in cartels.

In any event, both arguments are irrelevant to antitrust, which polices only monopoly inefficiency. Firms can be expected to choose the most efficient method of organizing to exclude competition on their own.

efficient.[27]

B. *The Per Se Rule in Favor of Monopolization*

Antitrust has two per se rules, but it is only aware of one of them. The one it knows is the per se rule against price fixing,[28] which is the doctrinal expression of the enforcement bias against cartels. The one it is not aware of is the per se rule in favor of monopolization, which it currently sees as the unrelated doctrines of the "conduct requirement" and "refusal to deal." These two doctrines express the enforcement bias in favor of monopolies.

The conduct requirement holds that you cannot be guilty of monopolization simply by having a monopoly and charging monopoly prices. You have to do something to keep rivals out of your market before you can violate the law.[29] The rule expresses the idea that antitrust should

---

27. I am not aware of any empirical work that seeks directly to compare the efficiency of monopolies and cartels. It is an expression of the bias in favor of monopolies that there is a long tradition of empirical work that seeks to establish a link between concentration or firm size and innovation. *See* COHEN & LEVIN, *supra* note 16, at 1060–61. Because the social value of innovation is thought to be much greater than the social cost of monopoly, a strong connection between monopoly and innovation would have something favorable to say about the efficiency of monopoly. *Id.* at 1060. Unfortunately, this work has shown that the link between monopoly and innovation is "fragile." *Id.* at 1078. Regardless of the results of this work, however, it is useless as a support for the bias in favor of monopolies because no work has been done to study the relationship between innovation and cartelization. So comparison is impossible.

There is also a tradition of empirical work that seeks to establish a connection between concentration and higher prices or higher profits. *See* Richard Schmalensee, *Inter-Industry Studies of Structure and Performance, in* 2 HANDBOOK OF INDUSTRIAL ORGANIZATION, *supra* note 16, at 951, 973–77, 987–88. The evidence for higher profits is weak, and prices seem to rise within industries but fall across them. *Id.* at 976, 988; Michael Salinger, *The Concentration-Margins Relationship Reconsidered*, BROOKINGS PAPERS ON ECON. ACTIVITY: MICROECONOMICS, 1990, at 287, 310. Regardless of the results of this work, it does not tell us anything about the proper relative treatment of cartels and monopolies because there has been no work done on cartel profitability. Indeed, even the empirical evidence that cartels tend to lead to higher prices is thin. WHINSTON, *supra* note 17, at 38 ("[I]t is surprising how limited the published literature is that documents significant effects of price fixing.").

So antitrust thinks that maybe innovation is associated with monopolization, and this might mean that monopolization increases social value. But antitrust has nothing to say about innovation and cartelization, because it has not studied it. And antitrust has found a connection between concentration and price, but not profitability, but it has established no such connection between cartelization and price or profitability (in the former case because the evidence remains thin and in the latter case because antitrust has not looked). That is a very long way from being able to say that monopolies tend to be more efficient than cartels.

28. *See supra* note 15.

29. *See generally* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 296–97. The conduct requirement is sometimes associated with the following classic statement:

A market may . . . be so limited that it is impossible to produce at all and meet the cost of production except by a plant large enough to supply the whole demand. Or there may be changes in taste or in cost which drive out all but one purveyor. A single producer may be the survivor out of a group of active competitors, merely by

only use the remedy of entry to eliminate monopoly inefficiency.

There are two types of markets: natural monopoly markets for which entry is impossible regardless what the incumbent does and markets for which entry is possible unless the incumbent does something to exclude competitors. Monopolists in natural monopoly markets do not take steps to exclude competitors because they do not need to. Barriers already exist. Indeed, they could not bring about entry even if antitrust were to order them to do so.[30] Because it is not for antitrust to solve monopoly inefficiency by other means than entry (e.g., by regulating rates),[31] there is nothing antitrust can do about natural monopoly. As a result, the conduct requirement exempts natural monopolies from the antitrust laws.

By contrast, in markets for which entry is possible, monopolists must do something to exclude competition; they can raise price and restrict output because the barriers they themselves create and can tear down protect them from competition. Antitrust can solve this problem by ordering monopolists here to allow entry. So the conduct requirement does not exempt these monopolists.

Natural monopolies enjoy barriers, charge P, and, at least sometimes, make black triangles.[32] Because the conduct requirement exempts

---

virtue of his superior skill, foresight and industry. In such cases a strong argument can be made that, although, the result may expose the public to the evils of monopoly, the Act does not mean to condemn the resultant of those very forces which it is its prime object to foster . . . . The successful competitor, having been urged to compete, must not be turned upon when he wins.

United States v. Aluminum Co. of Am., 148 F.2d 416, 430 (2d Cir. 1945) (Hand, J.). In this article I treat the conduct requirement as but one part of the rule articulated in this passage. The only way I can make sense of the conduct requirement is as an exemption for natural monopolies (see *infra* note 87); all other monopolies by definition must do something to exclude, so the conduct requirement cannot exempt them. Hand's references to markets that will only support a single plant or for which changes in taste have driven competitors from the market suggest situations that correspond to this natural monopoly interpretation of the conduct requirement.

Hand also writes about monopolies due to "superior skill, foresight and industry." *Aluminum Co. of Am.*, 148 F.2d at 430. These too may be examples of natural monopoly. But they also may not. A monopoly may be able to share its superior skill with competitors, in which case its monopoly may not be natural, but built on conduct: its refusal to share. As discussed more fully in the remainder of this Part and in Part IV, if the monopoly is not natural I treat it as exempt not under the conduct requirement but under the refusal to deal doctrine.

30. To be precise, they could not bring about entry without driving up their own costs. Not being able to bring about entry without raising costs is not the same thing, however, as not being able to bring about entry without reducing efficiency. This is discussed in Part VII.B.

31. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 322 ("Antitrust . . . is designed to be a market alternative to price regulation, not merely price regulation by another name.") (emphasis omitted).

32. That is why legislatures seek to regulate them, though not usually through antitrust. *See* RONALD R. BRAEUTIGAM, *Optimal Policies for Natural Monopolies, in* 2 HANDBOOK OF INDUSTRIAL ORGANIZATION, *supra* note 16, at 1289, 1299–1309 (discussing efficiency rationales for regulating natural monopolies).

natural monopolies, it amounts to a per se rule of legality for natural monopolies that charge P. For any given natural monopoly, there may be a black triangle associated with P, but antitrust does not care.

The conduct requirement is only half of the per se rule in favor of monopolization, however, because antitrust does not actually make unnatural monopolists do whatever is in their power to induce entry. Instead, large swaths of exclusionary conduct are exempted from antitrust scrutiny under the refusal to deal doctrine.[33] The conduct that the doctrine exempts appears somewhat arbitrary, but it tends to express a bias in favor of exempting conduct based on prototypical exercises of property rights.[34]

A survey of types of exclusionary conduct is useful in understanding how much exclusionary conduct is exempted by the refusal to deal doctrine. The following are examples of exclusionary conduct by a single firm.

- Real property. A farmer just happens to own all the fertile land in the market. She deploys her property rights to deny competitors access to farmland, allowing her to charge P in the crop market.

- IP. A woman invents a superior product and patents it. No one else can figure out an alternative way of producing it. The inventor deploys her property right to deny competitors access to the market for the superior product, allowing her to charge P in the market.

- Exclusive dealing.[35] A distributor convinces the manufacturers in an industry to sell only to her. As a result, she can charge P to consumers and share the monopoly profit with the manufacturers. She can charge P because the exclusive dealing arrangement denies access to the product to rival distributors, forcing them out of the market.

- Tying.[36] A hotel ties meals to rooms. Local restaurants have no

---

33. *See generally* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 317–24.

34. *See infra* Part II.C.

35. The example of exclusive dealing that follows is a "hub and spoke" horizontal conspiracy with the distributor serving as the hub. This is intended to capture the basic idea that the ultimate value proposition in exclusive dealing is really the same thing as the value proposition for any monopoly: using input control to capture surplus from consumers. *See* WHINSTON, *supra* note 17, at 152–54 (characterizing exclusive dealing as an attempt by the parties to appropriate benefits from outsiders in situations in which externalities prevent them from simply entering into a joint profit maximizing contract). For reasons for which it might be necessary to use vertical, as opposed to horizontal, agreements to monopolize, see generally *id.* at ch. 4.

36. This example of tying that does not run afoul of the single monopoly profit problem is given in ANDREW I. GAVIL ET AL., ANTITRUST LAW IN PERSPECTIVE: CASES, CONCEPTS AND

rooms to sell. The market is such that if you cannot sell meals to hotel guests you cannot afford to sell meals to locals either. As a result, local restaurants are driven from the market; guests and locals must eat at the hotel, and the hotel charges P. By tying the meals to rooms, the hotel makes rooms an essential input for anyone who wants to compete in the meals market. Because it is the only restaurateur in the market that can also offer rooms, competitors die.

- Predatory pricing. A firm charges below cost to attract buyers from the competition. In so doing, it has turned the industry into a natural monopoly. In order for any competitor to remain in the industry, it must charge a below-cost price too and later recoup it with a monopoly profit. But only one firm can earn the entire monopoly profit available in any given industry.

- Superior products. A firm makes a better product that no one else is able to copy or steal. This might be because the product is patented (a case already discussed above), but it might also be because the secret of the product's success is not apparent from examination of the product itself and is closely guarded by the firm. The secret is an essential input. Suppose the firm guards the secret by keeping it on a piece of paper in a vault on its property. By guarding it, the firm excludes competitors.

There is exclusionary conduct in all of these examples because the incumbent could do something to induce entry. The farmer could sell her fields at a price low enough to allow entrants to compete price down to C. The inventor could license her invention on similar terms. The distributor could stop trying to organize manufacturers. The hotel restaurant could dissolve the tie. The predatory pricer could charge cost.[37] And the superior producer could share her secret with the competition.

All these examples prevent entry and result in P.[38] The doctrine of refusal to deal would, however, tend to exempt any firm for which letting competitors into the market would require it to do something that smacks of waiving a property right. As a result, three of the six types of exclusion discussed above would probably be exempt from antitrust.

PROBLEMS IN COMPETITION POLICY 811–12 (2d ed. 2008) (attributing it ultimately to Robert Gertner).

37. This is true only ex ante. After the predatory pricing starts, the market is a natural monopoly market, at least for a time.

38. With the exception of predatory pricing, they are also all in a sense "refusals to deal" because in each case the exclusion could be alleviated by the sale of inputs to competitors. This is discussed in Part IV. The arbitrariness of the refusal to deal exemption lies in the fact that it does not treat all refusals equally. *See infra* Part II.C.

Refusal to deal protects the farmer from being forced to sell her land.[39] It would protect the inventor from being forced to license her invention.[40] And it would protect the superior producer from being required to allow competitors to come into its offices and examine its secrets.[41] For any of these firms, there may be a black triangle associated with P, but antitrust does not care.

As in the case of the per se rule against cartels, the per se rule in favor of monopolies makes organizational form relevant when there is no reason to think that it should be. For textbook economics, it does not matter whether entry is blocked because an incumbent happens to own all the inputs and uses property law to deny access or because the incumbent developed a clever tie. It does not even matter whether entry is blocked because the incumbent did something to block it or because the market just happens to support only one firm. What matters is that the price is P and there is a black triangle.

The per se rule in favor of monopolization (conduct requirement plus refusal to deal doctrine) is the other half of antitrust's bias against cartels and in favor of monopolies.[42]

C. *What It Means for Refusal to Deal to Exempt Property-Based Exclusion*

I have described refusal to deal doctrine as permitting property-based exclusion. Refusal to deal doctrine itself draws no distinction between property and contract. The classic statement is: "[i]n the absence of any purpose to create or maintain a monopoly, the [Sherman Act] does not restrict the long recognized right of trader or manufacturer engaged in an entirely private business, freely to exercise his own independent discretion as to parties with whom he will deal."[43]

I call this an exemption for property-based exclusion because in

---

39. This is paradigmatic refusal to deal. For an overview of the law, see *infra* note 43 and accompanying text.

40. Miller Insituform, Inc. v. Insituform of N. Am., Inc., 830 F.2d 606, 609 (6th Cir. 1987) ("A patent holder who lawfully acquires a patent cannot be held liable under Section 2 of the Sherman Act for maintaining the monopoly power he lawfully acquired by refusing to license the patent to others.") (citations omitted); Precision Instrument Mfg. Co. v. Auto. Maint. Mach. Co., 324 U.S. 806, 816 (1945) ("[A] patent is an exception to the general rule against monopolies and to the right to access to a free and open market."); *see generally* 1 HERBERT HOVENKAMP ET AL., IP AND ANTITRUST § 13.2 (2d ed. 2002 & Supp. 2011).

41. This is another paradigmatic refusal to deal. *See infra* note 43 and accompanying text.

42. To be sure, the bias is not complete. Some cartels and monopolies are not subject to per se treatment. *See infra* Parts IV.B, IV.E & VII.C.

43. United States v. Colgate & Co., 250 U.S. 300, 307 (1919). The phrasing of the classic statement of the refusal to deal doctrine makes it sound rather more enlightened than it has in fact become because the classic statement stresses that the doctrine applies only "[i]n the absence of any purpose to create or maintain a monopoly." *Id.* If taken seriously, this qualification says that

practice the doctrine is deployed to exempt exclusion based primarily on
property law. When the farmer refuses to let competitors use her land,
she relies on her title to the property. Refusal to deal doctrine exempts
this behavior from antitrust. If she did not own all the land outright, but
instead had merely contracted for exclusive access to it, then the refusal
to deal doctrine would not apply and antitrust might intervene to dis-
solve the contract.

Of course, an exclusive dealing agreement would only be exclu-
sionary because the other owners of the farmland can use their own
property rights in the land to exclude third parties in accord with the
agreement's grant of exclusivity. What distinguishes the exercise of

---

the refusal to deal doctrine in fact creates no per se rule in favor of property-based exclusion after
all and instead requires case-by-case examination of refusals for efficiency.

But antitrust for the most part ignores the qualification. Consider Hovenkamp's response to
the reference to monopoly in the classic statement quoted above. He writes that "[i]f a unilateral
refusal to deal is ever illegal, it is when the refusal is undertaken by a monopolist, or by someone
who threatens by the refusal to become one." HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra*
note 10, at 317. If ever. It is this skepticism that the monopoly qualification has any force at all
that gives the refusal to deal doctrine its power and meaning in antitrust law. *See also id.* at 321
(characterizing any requirement that a dominant firm deal with rivals as a "severe exception").
The skepticism is reflected in the rarity of court decisions that condemn refusals. Lately, the courts
have condemned them only when the refusal amounts to a stopping of a prior course of dealing.
See GAVIL ET AL., *supra* note 36, at 636–37. They also do so occasionally under the essential
facilities doctrine. *See infra* Part V.E.

Two recent Supreme Court cases also reflect this skepticism. Verizon owned the landline
telephone infrastructure in New York State. Verizon Commc'ns Inc. v. Law Offices of Curtis V.
Trinko, LLP, 540 U.S. 398, 402 (2004). It was presumably too expensive for competitors to run
their own cables to all the homes in the state, so Verizon in effect owned an essential input in the
provision of telephone service in New York. Verizon refused to provide quality access to its
network to competing phone service providers, such as AT&T. *Id.* at 403–05. The Supreme Court
did not permit the parties to litigate the efficiency of the conduct, but instead upheld dismissal of
the claim as exempt under the refusal to deal doctrine. *Id.* at 409–10.

The Court did suggest that an intent to monopolize the market might be relevant. *Id.* at 409.
But it then embarked on characteristically fallacious and dismissive economic reasoning to
conclude that no such intent was in evidence. It recognized that Verizon's refusal to deal had
allowed Verizon to defeat "cost-based" pricing imposed by its regulator, a fact that strongly
suggests monopoly inefficiency. *Id.* But it concluded that this did not show an intent to
monopolize because Verizon's behavior had not allowed it to raise price above "its own retail
price." *Id.* The court seems to have thought that the pricing imposed by the regulator did not
include the cost of capital, and that Verizon's exclusionary conduct could therefore have simply
been a way for it to earn something like a fair rate of return, as opposed to a monopoly profit. The
court seemed unaware of the fact that the rates imposed by the regulator already include "a
reasonable profit." 47 U.S.C. § 252(d)(1)(B) (2006).

Like Verizon, AT&T owned presumably unreproducible landline telephone infrastructure, in
this case in California. Pac. Bell Tel. Co. v. Linkline Commc'ns, Inc., 555 U.S. 438, 442 (2009).
Rather than exclude competitors by denying quality access, it in effect prevented them from
competing price down by requiring them to pay something like the monopoly profit it would
otherwise earn if it were to exclude them entirely and charge P directly to consumers. *See id.* at
449. The Supreme Court again refused to allow the parties to litigate the efficiency of the conduct
and instead upheld dismissal of the claim on refusal to deal grounds. *Id.* at 449–51.

property rights here is that their exercise is only exclusionary because of some non-property-based act: the exclusivity agreement. Without the agreement, we would want the other owners to exercise their property rights to exclude because this would allow them to profit from their labor and thereby give them incentive to compete vigorously with each other.

Similarly, tying involves an exercise of property rights. It is not covered by the refusal to deal doctrine, however, because the exercise has exclusionary effect only because of other non-property-based acts. In order for the tie between meals and rooms to exclude, local restaurants must not be able to expropriate rooms from the hotel. But what makes the hotel's right to refuse to give away rooms to competitors exclusionary is its imposition of the tying term in its contracts with buyers.

The point is not that, as a dumb matter of rule, a prohibited exclusionary practice may be distinguished from a protected refusal to deal because the practice involves "property plus" while a refusal to deal always involves property and only property. If a single owner owns all the land, and she leases it all to the farmer, refusal to deal might protect the farmer from being forced to sublet a portion of it to competitors, even though the farmer's right to the land is contractual. The reason is that the farmer's refusal to deal "feels" like an extension of the owner's right to refuse to deal.[44,45]

The point is that if it seems to enforcers that the only way for the incumbent to end exclusion is by waiving something that feels like a property right, then the exclusion is property-based.[46] Antitrust enforcers

---

44. Restrictive covenants are another grey area. If the owner of the best supermarket land in the market were to open a supermarket and let the rest of the land lie fallow, refusal to deal would protect her. But if she were to sell part to a supermarket and covenant not to allow a competing supermarket to enter the rest, a court might go either way. *Cf.* Tops Mkts., Inc. v. Quality Mkts., Inc., 142 F.3d 90 (2d Cir. 1998); Acme Mkts., Inc. v. Wharton Hardware & Supply Corp., 890 F. Supp. 1230 (D.N.J. 1995). (The distinction changes nothing from an economic perspective. Either there is monopoly inefficiency in both cases or there is not.)

45. The Supreme Court has said:

> It has been argued that . . . restrictive [practices] should be treated as beyond the prohibitions of the Sherman Act, since the owner of the property can choose his associates and can, as to that which he has produced by his own enterprise and sagacity, efforts or ingenuity, decide for himself whether and to whom to sell or not to sell. While it is true in a very general sense that one can dispose of his property as he pleases, he cannot go beyond the exercise of this right, and by contracts or combinations, express or implied, unduly hinder or obstruct the free and natural flow of commerce in the channels of interstate trade.

Associated Press v. United States, 326 U.S. 1, 14–15 (1945) (internal quotations and citations omitted). The vagueness of "go beyond" perfectly captures the role of "feel" in deciding what counts as protected property-based exclusion.

46. It is for this reason that antitrust tends to associate refusal to deal with the avoidance of

have an intuitive sense of what involves a property-type exclusion and what does not. "Property plus" says only that when exercise of a property right appears to enforcers to loom large in exclusionary conduct, they will not challenge it. "Property plus" does not explain what exactly makes a particular exercise of property rights loom large relative to other causes of exclusion, and another loom small. That is for social psychologists to determine.[47]

D. *The False Distinction Between Collusion and Exclusion*

A bulwark of the bias against cartels and in favor of monopolies is antitrust's fondness for the distinction between collusion and exclusion. Antitrust believes that collusion and exclusion are two ways, different in kind, in which industries get to P.[48] This belief seems to rest on two supporting stories. In the collusion story, you start with a market for which there are barriers to entry that prevent any additional firms from entering. Price is at C despite the barriers because there are already many firms in the market. These firms compete, keeping price low. Price subsequently goes to P because these firms stop competing and cartelize instead. In the exclusion story, you start with a multi-firm market for which price is at C. One firm subsequently does something to kick the others out of the market and price goes to P.

The heart of the distinction is that collusion does not actually kick any firms out of the market. Everyone stays. What drives price to P is that firms now work together. In contrast, exclusion gets you to P by driving firms out of the industry.[49] This is an intuitively pro-monopoly and anti-cartel story because the point of competition is to have each

---

judicially-mandated "forced sharing." *See Verizon,* 540 U.S. at 408 ("Enforced sharing . . . requires antitrust courts to act as central planners, identifying the proper price, quantity, and other terms of dealing—a role for which they are ill-suited."). What antitrust seems to be trying to say by its choice of the word "sharing" is that something it feels amounts to property is involved (not necessarily as a strict matter of legal category, but as an emotional matter). We share what we feel that we own. Antitrust is only ever faced with the problem of having to force sharing when it is trying to remedy exclusion brought about by something that it feels amounts to the exercise of property rights.

47. I have ignored the fact that some exclusions may involve neither property nor property plus. An example would be some forms of gangsterism (e.g., I exclude you with a threat). It is easy to predict how enforcers will treat such exclusions (they will not exempt them under the refusal to deal doctrine).

48. *See* POSNER, ANTITRUST LAW, *supra* note 8, at 40–41 (stating that the difference between collusion and exclusion "is fundamental to an understanding of the antitrust laws").

49. *See, e.g.,* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 214 (stating that collusion "creates market power by bringing firms *into* the venture rather than excluding them from the market," as in the case of monopoly). Surprisingly, economists have seemed to embrace this distinction as well. *See, e.g.,* WHINSTON, *supra* note 17, at 133 ("While the aim of anticompetitive horizontal agreements is collusion . . . , the concern arising from the vertical agreements . . . that I focus on here is exclusion.").

firm strive to kick the others out of the market; it is not for firms to work together.[50] This suggests that the persecution of exclusionary conduct ought to be much more cautious than the persecution of collusion.

There is nothing wrong with telling these collusion and exclusion stories. They describe distinct ways in which an industry might get to P. Trouble arises only if you assume that the distinction is at all meaningful from the perspective of textbook economics. What gets price to P according to textbook economics is not whether a firm is forced to stop doing business in an industry, but that the number of competitors in an industry declines. It is the exclusion of competitors, not firms, that matters. Textbook economics teaches that collusion and exclusion both drive price to P because they do exactly the same thing: they both exclude competitors.[51]

A competitor chooses its output level in order to maximize the profit from what it sells. A colluder chooses output in order to maximize the profit from what the group sells. As such, a colluder cannot be a competitor. The only competitor in a cartel is the cartel itself, which chooses its overall output level in order to maximize the profit from what it sells. So, when a cartel forms out of a group of competitors, the number of competitors in the market falls by the number of members of the cartel, minus one to account for the cartel itself. Collusion does not expel member firms from the market, but it does expel competitors from it.

Just as it is a mistake to assume that collusion does not exclude in an economically meaningful way, it is also a mistake to assume that the entry of firms into an industry must imply entry in an economically meaningful way. The exclusion of a firm that would have colluded upon entry is not exclusion in an economic sense. Though the excluded firm would have entered the market in the sense that it would have contributed to output, it would have operated to maximize the joint profit of the

---

50. *Cf.* United States v. Aluminum Co. of Am., 148 F.2d 416, 430 (2d Cir. 1945) ("The successful competitor, having been urged to compete, must not be turned upon when he wins.").

51. Textbook economics does not say this directly, but the point can hardly be controversial once it is grasped. Hints can be found everywhere. *See* CARLTON & PERLOFF, *supra* note 10, at 125 ("A competitive firm ignores the good it does other firms by reducing its output and increasing the market price . . . . [T]he externality created by each firm in reducing its output [is] internalized by the cartel. As a result, it pays the cartel to reduce total output below the competitive level, even though it would not pay any competitive firm to reduce its output individually."); GAVIL ET AL., *supra* note 36, at 588 ("[A] similar outcome to what might arise from successful collusion can be achieved if a dominant firm or a group of firms acts to exclude rivals that might otherwise not go along with collusive conduct."). The quote in the second parenthetical is particularly sensitive to the identity of collusion and exclusion, although it does not quite make it explicit. A single firm only needs to exclude firms from the market when it cannot induce them to collude; that is because the act of inducing collusion is just another way of excluding competitors.

cartel and not its own, so its entry would not increase overall industry output or have any effect on price. For the exclusion to be meaningful, it must keep out a firm that intends to maximize its own profit, which is to say, not merely a firm, but a competitor.

In order to get to P in an industry, you need some intelligence with control over the output of the industry that is intent on maximizing profit. That is to say, to get P, you need one and only one competitor. It does not matter whether that competitor is bound together by informal contracts linking a group of corporations or partnerships, or by a single corporate charter and deeds, titles, and employment contracts, or by an ethic of neighborliness, or by fear and friendship. What matters is that it is the only competitor and has the power to stay that way, even when it charges P.

But if cartelization is as much exclusionary conduct as any of the monopolization schemes outlined above, then it ought to be treated with the same leniency as monopolization. Or you might say with equal force that if monopolization is as exclusionary as cartelization, then monopolization ought to be treated with the same severity as cartelization.

E.   *Section Two as a Dead Letter*

The first two sections of the Sherman Act[52] are often read as articulating a ban on collusion and a ban on exclusion, respectively.[53] As we have seen, this distinction is untenable from an economic perspective. A ban on exclusion is a ban on collusion. What the two sections might better be read to do is to ban "property plus" exclusion and property-based exclusion respectively. What allows a cartel to exclude is not just property but also contract, however informal, so cartelization would still fall under Section One. If we read them this way, then at present Section One is all there is to the Sherman Act. Section Two is a dead letter. As we have seen, refusal to deal eliminates all property-based claims. It thereby swallows Section Two whole.[54]

---

52. 15 U.S.C. §§ 1–2 (2006).

53. *See, e.g.*, CARLTON & PERLOFF, *supra* note 10, at 632–33, 662.

54. Posner too seems to want to think that Section 2 is a dead letter, but for quite different reasons. POSNER, ANTITRUST LAW, *supra* note 8, at 259–64. He thinks that an expansive reading of the concept of contract lets you include all recognized forms of exclusionary conduct under Section 1. *Id.* at 259–60. My point is that what distinguishes Section 1 claims is not so much that they can be related to contract in particular, but that they have some appeal as involving more than just property-based exclusion (a plus factor).

It does not really matter what that plus factor is. Posner can, if he thinks a contract nexus is necessary to bring a claim under Section 1, argue in effect that the plus factor is always contract. So, for example, as a purely formal matter he may be able to characterize predatory pricing as contractual in the sense that the low price is part of a sales transaction. *But see* S. Todd Lowry, *Bargain and Contract Theory in Law and Economics*, 10 J. ECON. ISSUES 1, 12–16 (1976)

III. No JUSTIFICATION FOR A BLANKET EXEMPTION: MONOPOLY
POWER IS STILL A PROBLEM IN A WORLD OF PROGRESS

A. *An Antitrust and Product Improvement Model*

We have seen that there is no theoretical or empirical support for treating cartels differently from monopolies. But this tells us nothing about how to treat them consistently. Consistency might mean extending the per se rule against cartelization to also prohibit all monopolization. Or it might mean extending the per se rule in favor of monopolization to cover all cartelization (i.e., doing away with most of antitrust). Or it might mean doing away with per se rules entirely and testing for efficiency everywhere on a case-by-case basis.

We saw in Part II.A that the case for the per se rule in favor of monopolization is built on the notion that monopoly power is necessary to cover fixed costs. This is a generalization from an argument, which has become increasingly conspicuous in recent years,[55] that monopoly profit is a reward for innovation in the sense that it allows innovators to

---

(distinguishing between the law of sales and the law of contract). But doing so is vacuously formalistic. There may be a contract element to predatory pricing, but there is nothing collusive about it. The economic significance of predatory pricing is that it transforms the market into a natural monopoly. That is a plus factor, though not one that is meaningfully contractual in character. To try to characterize it as contractual is to commit the characteristic intellectual sin of law scholars, which is to put the language of the law before understanding, form before substance. If we are to find an elegant way to stuff the live part of antitrust into Section One and the dead part into Section Two, better to stop thinking about "contract, combination, . . . or conspiracy in restraint of trade" as "contract" and start thinking of it as code for property plus. 15 U.S.C. § 1.

55. Most glaringly in the majority opinion in *Verizon v. Trinko*, which reads:

The mere possession of monopoly power, and the concomitant charging of monopoly prices, is not only not unlawful; it is an important element of the free-market system. The opportunity to charge monopoly prices—at least for a short period—is what attracts "business acumen" in the first place; it induces risk taking that produces innovation and economic growth. To safeguard the incentive to innovate, the possession of monopoly power will not be found unlawful unless it is accompanied by an element of anticompetitive conduct.

Firms may acquire monopoly power by establishing an infrastructure that renders them uniquely suited to serve their customers. Compelling such firms to share the source of their advantage is in some tension with the underlying purpose of antitrust law, since it may lessen the incentive for the monopolist, the rival, or both to invest in those economically beneficial facilities.

Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 407–08 (2004) (emphasis omitted); *see also* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 296 ("'[M]onopolization' and the monopoly profits that may result are essential to economic development. Firms innovate because they expect their successes to produce economic returns."); CARLTON & PERLOFF, *supra* note 10, at 99 ("[T]he prospect of receiving monopoly profits may motivate firms to develop new products, improve products, or find lower-cost methods of manufacturing. Were it not for the quest to obtain monopoly profits, firms might innovate less. The benefit of monopoly is most clearly recognized in research and development."); GAVIL ET AL., *supra* note 36, at 583 ("Making the mere possession of substantial market power illegal would ensnare firms that achieved preeminence through laudable means such as reducing costs and

cover the fixed costs associated with R&D.[56] Because innovation con-
tributes much more to social value than monopoly inefficiency takes
away, the argument goes, monopoly profit is therefore efficient.

According to this argument, so long as antitrust vigilantly polices
the exclusionary practices that do not fall under the exemption for refus-

---

improving product quality. A competition system that punished such success inevitably would
dampen the initiative to develop new products or processes that benefit consumers.").

56. This is a charitable interpretation of the argument. Sometimes, the argument appears to be
not that monopoly profit is necessary to cover fixed cost but that a reward above fixed cost is
necessary to create an incentive to innovate. *E.g.*, HOVENKAMP, FEDERAL ANTITRUST POLICY,
supra note 10, at 296 ("A rule that condemned all prices higher than, say, average cost could stop
innovation dead."). The "greater than cost" argument reflects at best a substitution of the
accounting for the economic definition of cost and at worst a misunderstanding of how economics
works.

The economic definition of cost includes sufficient profit to make the producer willing to
engage in the activity for which she incurs the cost. This means that it includes enough reward to
make the activity more profitable than any other activity available to the producer. This has been
the standard definition at least since Alfred and Mary Paley Marshall wrote in the late 19th
century:

> In deciding for instance whether to make some boxes, the manufacturer calcu-
> lates . . . their Expenses of production. He makes a definite allowance for the
> remuneration of his own labour . . . . He reckons in these expenses not only the
> outlay of money that he will make, but interest on this outlay together with his own
> Earnings of Management, or more strictly, profits. If he sees his way to getting a
> price that will cover these Expenses of production, and therefore give him adequate
> profits, he is content and continues his production.

ALFRED MARSHALL & MARY PALEY MARSHALL, THE ECONOMICS OF INDUSTRY 75 (London,
MacMillan & Co. 1879).

It is precisely because cost includes sufficient profit to induce action that it is often said that
in a competitive market, firms produce at cost. VARIAN, *supra* note 10, at 403–10. If cost did not
include all the profit needed to make firms willing to produce, which is to say, enough profit to
make the reward for producing in the particular market greater than the reward for producing
anything else, then there could be no competitive markets because no one would be willing to
produce in them.

(An important corollary of this point is that a reward above that necessary to induce action is
not part of cost. So, for example, if the producer's best alternative profit is $1.00, then a profit of
$1.01, and no more, is part of the cost of the activity, because that, and no more, is necessary to
induce the producer to prefer it. If the producer were to earn a reward of $1.02, then the extra
penny would not be part of cost.)

The accounting definition of cost does not include the return necessary to make innovation
more profitable than doing anything else (it does not include the $1.01). If the "greater than costs"
argument is simply that, using an accounting definition, something more than a price that covers
cost is necessary to induce innovation, then the error in the "greater than costs" argument is just
one of nomenclature. It is just an accountant's way of saying that innovators need to be able to
cover fixed cost in the economic sense.

But if the argument is that something greater than cost in the economic sense is necessary to
induce innovation, then the error is conceptual. Because the definition of cost includes all the
profit necessary to induce innovators to innovate, a reward greater than cost by definition means a
reward greater than necessary to induce innovation. A reward above cost by definition cannot be a
necessary reward. The whole point of antitrust is to identify and eliminate such unnecessarily high
rewards because every dollar of additional reward inflicts a loss on society in terms of monopoly
inefficiency. As soon as reward stops being necessary, a black triangle appears in Diagram 1.

als to deal, the only way for a firm to obtain a monopoly position will be through the creation of a superior product, which implies innovation of some kind.[57] So only innovators will receive the reward of monopoly profit, and a powerful incentive to innovate is created thereby. Although the argument is not generally made with respect to cartels, we have seen that there is no reason not to extend it that far.[58]

Antitrust today, then, is quite taken by an argument for monopoly that suggests that consistent treatment of cartels and monopolies would require extending the per se rule in favor of monopolization to swallow antitrust law. The trouble with the argument is that there is no reason to think that admitting that monopoly profit can have a stimulating effect on innovation implies that full monopoly power is always, or even sometimes, efficient.[59] To understand this, let us consider a model of the relationship between R&D expenditures, product improvement, and antitrust enforcement levels.[60] The model is summarized in Diagram 2 on page 129, explained in this section, and discussed in the next.

In this model, there is a trade-off between product quality on the one hand and both R&D cost and access to the product on the other.[61] You get better products through R&D. But in exchange for better products, you suffer two costs. First, there is the cost of the R&D itself.

R&D expenditures are fixed costs because you must incur the cost of inventing a product before you can start producing it.[62] In order to pay for fixed costs, you must be able to charge a higher price, which means there must be less antitrust enforcement. But if there is less anti-

---

57. *Cf.* POSNER, ANTITRUST LAW, *supra* note 8, at 113–16 ("[I]t takes a good deal of strained and ad hoc argumentation to explain persistent monopoly or concentration withòut assuming unlawful exclusionary practices, lawful patent protection, economies of scale, superior management, competitive pricing, or other factors that would not justify breaking up a firm.").

58. *See supra* note 21 and accompanying text.

59. For an outline of an alternative critique based on the argument that competition actually promotes innovation, see Baker, *supra* note 14, at 561–62. The strategy of this article is to engage in "internal critique" by accepting the assumptions of the argument in favor of monopoly as given and showing that they do not produce their intended result. This article therefore proceeds by accepting the premise that monopoly power can promote innovation and showing that it does not lead to the conclusion that there is no need for antitrust to regulate monopoly power.

60. Antitrust has traditionally modeled R&D as contributing to cost reduction rather than product improvement. Both approaches yield the same conclusions. (For the cost reduction model that inspired the product improvement model in this article, see Vincenzo Denicolò, *Do Patents Over-Compensate Innovators?*, 22 ECON. POL'Y 679 (2007).) I choose the product improvement approach because product improvement accounts for the lion's share of R&D spending in the United States. *See* SCHERER & ROSS, *supra* note 10, at 630.

61. For simplicity, the model assumes that when a product improvement is introduced, the unimproved product disappears from the market, so consumers who do not end up with access to the improved product end up with access to nothing. *Cf.* SCHERER & ROSS, *supra* note 10, at 622 ("If the product is really new and useful, it creates a wholly new demand curve . . . .").

62. For simplicity, this discussion of the model will treat fixed costs as being due entirely to expenditures on R&D.

DIAGRAM 2

──────── Firms' choice: Rising fixed cost / product improvement ────────→

Enforcers' choice: Falling strength of antitrust enforcement / rising price relative to marginal cost ──────↓

C' = actual price	C = marginal cost	P = monopoly price
Q" = actual quantity	Q' = quantity under marginal cost pricing	Q = monopoly quantity
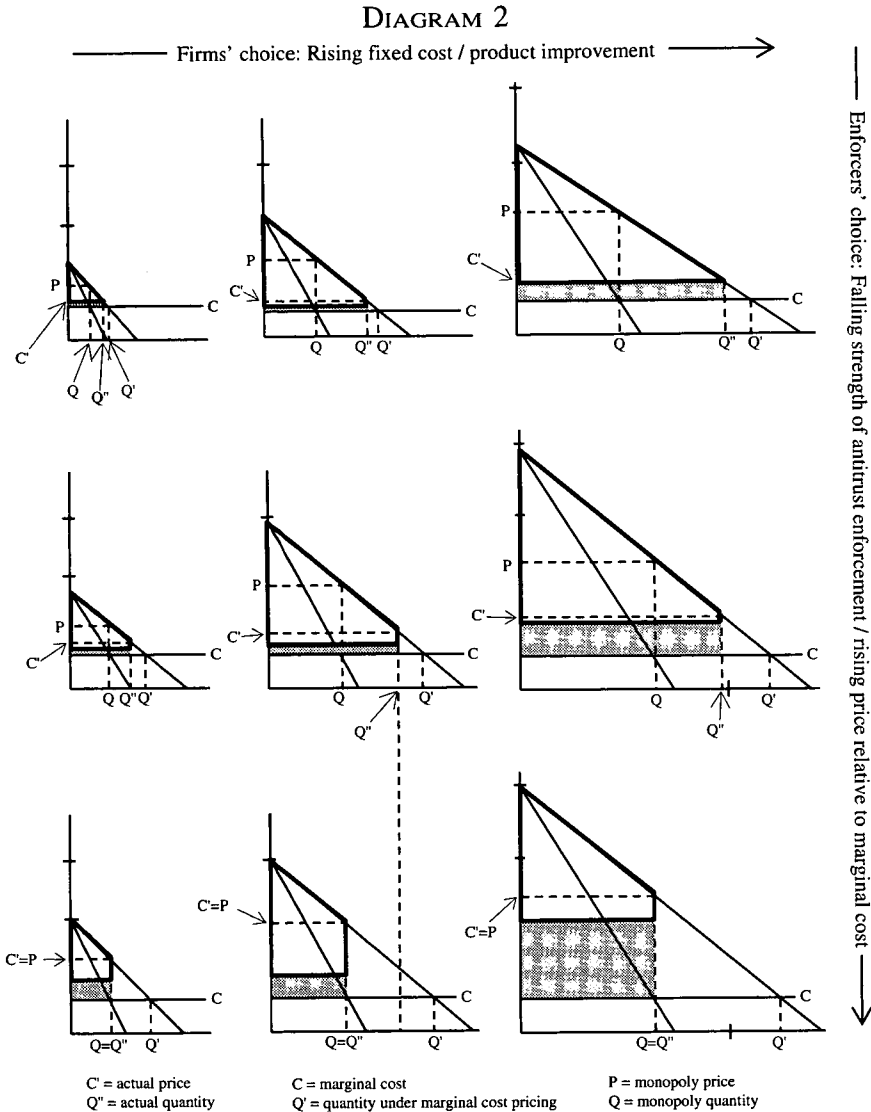
trust enforcement and prices are higher, then, if product improvement is not sufficiently high, there will be less output and therefore a loss of enjoyment associated with the lost output.[63] This loss of enjoyment due to reduced access is the second cost. There can therefore be a trade-off between improving product quality, the cost of the improvement, and allowing access to the product. Higher quality means higher research

---

63. A key assumption throughout this article is that fixed cost exceeds producer surplus. Otherwise, marginal cost pricing is sufficient to cover costs and there are no efficiency problems. This assumption is dealt with in Diagram 2 by the use of constant marginal costs.

costs and less access; more access and lower research costs mean lower quality.

Social value is the combined value of production to producers and consumers. The market that maximizes social value is efficient.[64] If product improvement is below the efficient level, then you can increase the benefit of the product to society overall by improving the product, even if that means that less of the improved product will be available.[65] If improvement is above the efficient level, then society would have been better off overall if you had invested less in product improvement and simply made more of a poorer quality product available.[66]

The trade-off between quality and access is starkest in Diagram 2 in the right column as you move from top to bottom. Product quality increases, pushing the demand line up toward the top tick mark. This tends to increase the size of the bolded area, which represents social value,[67] and accounts for the boost to social value associated with more R&D. But price, C', also increases to cover the higher fixed costs; as a result, output, Q'', falls further and further below the tick mark. This increases the size of the deadweight loss triangle created by the output level on the left, the demand line on the top, and C on the bottom. The expanding deadweight loss triangle tends to reduce the social value area. The increase in fixed costs, represented by the gray box, itself also tends to reduce the social value area. Together, these two efficiency-reducing effects account for the hit to social value associated with more R&D.

The relative size of the positive effect due to product improvement and the negative effects due to reduced access and higher fixed cost

---

64. This article takes a partial equilibrium approach to welfare economics. For welfare economics assumptions, see Arnold C. Harberger, *Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay*, 9 J. ECON. LITERATURE 785 (1971). For partial equilibrium assumptions and pitfalls, see KREPS, *supra* note 12, at 279–283; EUGENE SILBERBERG & WING SUEN, THE STRUCTURE OF ECONOMICS 498–99 (3d ed. 2001).

65. For example, you might think that the extra value to those who have smartphones of having cameras in smartphones justifies including them, even after you take into account the fact that including them makes smartphones so much more expensive that some people are unable to afford smartphones at all as a result.

66. For example, you might think that the extra value to those who have cars of making maglev technology standard instead of wheels would not justify the switch, because then only a very small number of people would be able to afford a car.

67. The area under the demand line gives the value of the product if everyone who values it were to consume it. The area under the demand line bounded by the y-axis on the left and the output actually produced on the right (i.e., Q'') gives the value of the product to those who actually consume it. The social value created by the product is this value less the cost of producing the product, which is variable cost plus fixed cost. Fixed cost is the gray rectangle. Variable cost is the area under the marginal cost line from the y-axis on the left to the level of output actually produced on the right. This means that the shape created by the y-axis and Q'' on the left and right, and the demand line and the top of the gray box on the top and bottom, represents the social value associated with the product.

determines whether the size of the social value area ultimately increases or decreases. In the right column, this area appears to increase as you move from top to middle, and then to fall as you move from middle to bottom. Maximum social value (and therefore efficiency) for this column is obtained with the level of product improvement shown in the middle graph. In fact, if you survey all the graphs in Diagram 2, the middle right graph appears to have the highest social value and therefore to depict the most efficient level of R&D spending.

Diagram 2 contains three rows, each of which corresponds to a different level of monopoly power. It also contains three columns, each of which corresponds to a different level of product improvement. The graphs in each row show the market at different levels of product improvement, taking the level of monopoly power as given. The graphs in each column show the market at different levels of monopoly power, taking the level of product improvement as given. Firms choose the level of product improvement, so they choose one of the three graphs for any given row. Antitrust enforcers choose the level of monopoly power, so they choose one of the three graphs for any given column. As we shall see, firms will always choose a graph in the middle column, because profit for any given row is maximized in that column.[68] As a result, antitrust enforcers can never bring the economy to the social-value-maximizing graph in the right column. We shall see that, forced to make do with a limited set of options, antitrust enforcers should choose to take monopoly power to the middle graph in the middle column.

Let us call the amount of money earned by firms in a market above variable cost their quasi-profit. It is quasi because the firms will want to

---

68. The nine graphs in Diagram 2 are meant to summarize the process of determining social value using two variables: R&D (fixed cost) and monopoly power. The graphs are not the outputs of any particular algebraic model. Instead they are drawn to make it easy to see by visual inspection the trade-offs involved in the optimization process. In order to show that firms do not always choose to maximize R&D, the middle column graphs are chosen to reflect intermediate levels of R&D but maximum levels of profit. In order to show that something less than full monopoly power may be an optimal solution, the middle row is drawn to reflect maximum levels of social value.

Such "interior" solutions can be obtained, for example, for a model with demand $p = i - q$, marginal cost $c = 0$, and R&D fixed cost $F = i^5$, where i is an index of product quality ≥ 0. Antitrust drives price down to $0 \leq a \leq i$. For any prevailing antitrust regime, a, firms will choose product quality to maximize profit, which will amount to choosing i^* such that $F'(i^*) = a$. Social value is consumer surplus plus profit, or $\frac{1}{2}(i - a)^2 + a(i - a) - F$. Choosing a to maximize social value subject to the profit condition gives $a^* = i^* / (F''(i^*) + 1)$. Substituting $F = i^5$ gives $a \approx 0.167$ and $i \approx 0.427$. The monopoly price in this model will always be $\frac{i}{2}$, so in this model antitrust should optimally drive price down to a level substantially below the monopoly price ($0.167 < 0.214$).

Rather than cast about for parameters and variable values for such a model that might yield graphs with easily distinguishable differences in shape sizes, I chose instead to take the heuristic approach represented by Diagram 2.

spend some of it on fixed costs. Their profit is their quasi-profit less their fixed cost. In the graphs in Diagram 2, quasi-profit is the box bounded by the y-axis and Q″ on the left and right and C′ and C on the top and bottom. The box always has a gray portion, which represents the portion of quasi-profit spent on fixed costs, including R&D. The white portion is profit.[69,70]

Firms will invest in more R&D so long as the resulting increase in demand increases their profit after taking the increased fixed cost into account. The rows in Diagram 2 show how much profit firms earn in response to different levels of R&D spending for a given level of monopoly power. In the top row, the firms' monopoly power is weak and the market price C′ is therefore low relative to the monopoly price P. The left graph shows that for a low level of R&D spending, the firms earn no profit. The middle graph shows that as spending increases, the firms eke out a small profit. But the right graph shows that too much investment in R&D will drive profit back to zero. The middle and bottom rows reflect a similar story for intermediate and full monopoly power levels, respectively.

For all three levels of power, the maximum level of profit is obtained in the middle column. Firms will always choose a level of R&D expenditure that takes them to this profit-maximizing level. The firms will never choose the high levels of R&D in the right column, even though these maximize social value for each level of power.

We can think of the profit-maximizing levels of investment for each level of monopoly power as feasible levels of investment for each power level. Other levels might be more efficient, but so long as we are committed to allowing firms to maximize profit instead of efficiency, firms will choose these levels and antitrust enforcers are stuck with them. In Diagram 2, accepting profit maximization as the ultimate determinant of firm behavior[71] forces us to cross out the left and right col-

---

69. In the left and right graphs in the top row, the firm spends its entire quasi-profit on fixed costs, so there is no white portion.

70. In this model, fixed cost includes whatever profit is necessary and sufficient to make firms prefer to produce in the market described in the graphs in Diagram 2. But in this model fixed cost does not include whatever profit is necessary to make firms produce in any particular graph in Diagram 2 (i.e., to make firms prefer one level of R&D expenditure over another). This is what allows us to distinguish between the part of quasi-profit spent on fixed cost and the part left over for profit. This approach lets us focus on how antitrust can use the regulation of monopoly power to associate the largest profit available to firms with the social value maximizing level of R&D expenditure.

(Because this approach includes part, but not all, of profit in cost, it counts as a "partial accounting" approach to defining cost. It is this partial accounting approach to fixed cost that may justify the tendency of advocates of the reward theory of monopoly to suggest that rewards ought to exceed costs. *See supra* note 56.)

71. The model assumes that there is no competition to enter ex ante, meaning that firms do

umns and consider the middle column only.

The role of antitrust is to use its enforcement powers to bring about the level of monopoly power that corresponds to the most efficient level of investment in R&D in the set of feasible levels of investment. In Diagram 2, this means that antitrust must choose the level of monopoly power that corresponds to the largest social value shape in the middle column. The graphs in Diagram 2 are drawn so that the shape is largest for the middle graph in that column. This graph reflects the most efficient feasible level of investment in R&D. Antitrust can bring the industry to this level by engaging in antitrust enforcement that drives monopoly power in this market to the intermediate level represented by the middle row.[72]

B. *The Model Shows That There Is No Reason to Assume That Monopoly Pricing Is Efficient, Even When Firms Spend Their Profits on Improving Their Products*

Defenders of antitrust tend to want to assume that there is never any trade-off between product quality and access. Firms cannot increase their profits by investing in R&D and therefore antitrust enforcers cannot induce more investment in R&D by allowing greater levels of monopoly power. As a result, monopoly power is an unconditional bad.

---

not compete to invent the best product; they only compete after a particular product has been invented and introduced. Without this condition, competition to enter the market would cause firms to get into an R&D arms race, which might drive R&D expenditure up to exhaust all available quasi-profits for any given level of monopoly power. In Diagram 2, the market would be driven as far to the right as possible for any given row. Because this article wishes to engage with the argument that profit is an incentive for innovation, it ignores the competitive case and focuses on the case in which firms have the power to maximize industry profit.

However, the conclusion of this Part, that promoting innovation does not imply that monopoly pricing is efficient, holds with equal force in a model of competitive innovation. For a discussion of the competitive and non-competitive entry cases (called "free entry in research" and "monopoly in research") in a process improvement model, see Denicolò, *supra* note 60, at 686–89, 710.

72. In the discussion of Diagram 1, I noted that if firms charge P, then output that consumers could afford to enjoy if price were at cost is not produced. *See supra* the paragraph containing note 9. I used a black triangle to represent the value lost to consumers. The black triangle is just the difference between social value at a given price and social value at the optimal price. In Diagram 1, the given price is P and the optimal price is C. At C, social value is the triangle bounded by the demand line on top, the y-axis on the left and C on the bottom. At the monopoly price, output falls to Q, lopping off the area of the black triangle from social value.

Diagram 2 does not color in losses in social value relative to the optimal level. We could still color in losses in social value if we wished, but the shapes would no longer be triangles. For example, the loss in social value associated with the price in the top graph of the middle column would be the shape left over after subtracting social value in that graph from social value in the middle graph in the middle row (which we identified above as the graph of price at the optimal level).

It reduces access without bringing about any countervailing quality benefits.[73]

For the defenders of antitrust, all the demand lines in Diagram 2 are identical and therefore so too are all the graphs and all the social value shapes in each row. As you move from the top row to the bottom, price rises and the social value shapes fall in size as the corresponding reduction in output increases the deadweight loss triangles. There is no benefit from product improvement to counterbalance this effect. Because efficiency falls with every increase in monopoly power, in this world the optimal level of monopoly power is the lowest level that covers variable costs. Society always wants to be in the top row.

In recent decades, the opponents of antitrust have had great success attacking the defenders for assuming away any relationship between monopoly power and product improvement.[74] But they deserve at least

---

73. *E.g.*, Paul Feldman, *Efficiency, Distribution, and the Role of Government in a Market Economy*, 79 J. POL. ECON. 508, 517 (1971) ("Wherever noncompetitive markets exist, government should operate to lead them to the competitive solution.").

74. Not all of the attack has been about the need for rewards in the form of monopoly profit to cover R&D fixed costs. The early attack of the opponents on the defenders was the argument that product or process superiority is tied to natural monopoly. *Cf.* Sam Peltzman, *The Gains and Losses from Industrial Concentration*, 20 J.L. & ECON. 229, 230–33 (1977) (recognizing the connection to natural monopoly). The argument, which was mostly made in process improvement terms, was that superiority is due to unshareable factors like management talent or merger synergies that cannot be reproduced without the merger. Because advantage cannot be shared, antitrust can do nothing to lower price short of regulating rates. *Cf.* JOHN S. MCGEE, IN DEFENSE OF INDUSTRIAL CONCENTRATION 21–22 (1971) [hereinafter MCGEE, IN DEFENSE].

The merger synergies argument played a central role in the rolling back of what once amounted to a per se rule against all but the smallest mergers. *See* F.M. Scherer, *Conservative Economics and Antitrust: A Variety of Influences*, *in* HOW THE CHICAGO SCHOOL OVERSHOT THE MARK, *supra* note 13, at 30, 35 (noting the "significant" change in merger law brought about by the incorporation of an efficiency defense); Herbert Hovenkamp, *United States Competition Policy in Crisis: 1890–1955*, 94 MINN. L. REV. 311, 355 (2009) (noting that in 1963 there was a "virtual per se rule" against mergers involving large market shares) (citations omitted) [hereinafter Hovenkamp, *United States Competition Policy in Crisis*]. The unshareable factors argument played a prominent role in the Chicago School's successful destruction of an attempt in the 1970s to do away with the per se rule in favor of monopolization and replace it with case-by-case analysis (the no-fault monopolization campaign). *See* William E. Kovacic, *Failed Expectations: The Troubled Past and Uncertain Future of the Sherman Act as a Tool for Deconcentration*, 74 IOWA L. REV. 1105, 1126–27, 1138–39 (1988); POSNER, ANTITRUST LAW, *supra* note 8, at 101 & n.1 (sources cited therein). For the best formal account of the unshareable factors argument, see Peltzman, *supra*, at pp. 230–33. For the way it was deployed to combat the deconcentration initiative, see John S. McGee, *Efficiency and Economies of Size*, *in* INDUSTRIAL CONCENTRATION: THE NEW LEARNING 55, 61–65 (Harvey J. Goldschmid et al. eds., 1974) [hereinafter McGee, *Efficiency and Economies of Size*]. Although the heart of these arguments was natural monopoly, rewards arguments hovered about the edges. For a discussion of these, see *infra* note 75.

Reward featured more prominently in the early arguments for relaxing scrutiny of vertical restraints. Part of those arguments was that vertical restraints are efficient because they allow monopolists to capture more of the surplus associated with product or process improvements, not only by making it easier to increase uniform prices, but also by making it easier to engage in price discrimination, and this in turn allows firms to invest more in improvements. For a discussion of

as much criticism for making the equally extreme assumption that more monopoly power and more product improvement always translate into more efficiency.[75] The opponents of antitrust tend to assume that the social value area in the bottom graph in the middle column of Diagram 2 must always be the largest of the social value areas in the middle column. Just as it seems unreasonable to assume that for all or most markets there can be no relationship between pricing power and product improvement, it is quite unreasonable to assume that for all or most markets monopoly pricing in particular induces firms to maximize social value.

---

how these arguments succeeded in both the resale price maintenance and restrictive patent licensing contexts, see Scherer, *supra*, at pp. 34–35, 37–39. The rewards argument has grown in recent years to become the central argument for monopoly, backstopping the gains won by the early arguments.

75. *E.g.*, McGEE, IN DEFENSE, *supra* note 74, at 137 ("If there is enough demand for machines, given their costs, they get produced. If large profits are earned, there is a tendency for others to try . . . ."); JOHN S. McGEE, *Commentary, in* INDUSTRIAL CONCENTRATION: THE NEW LEARNING, *supra* note 74, at 101, 104 ("[A]part from those industries dominated by State controls, there is the strongest presumption that the existing structure is the efficient structure.").

The early rewards arguments of the opponents appear not to have been based on the notion that a reward is required to cover R&D fixed costs in particular. Instead, the argument was what I call a "pure incentive" argument. It held that a monopoly profit is necessary to make innovation attractive to firms, regardless of the need to cover R&D fixed costs. According to the argument, if an innovator is not rewarded with monopoly profit, even for a short time, but instead finds herself in a competitive market both before and after innovating, then she cannot have any basis for preferring innovation to stagnation. The competitive market will push her return as low as possible, leaving her no better off after innovating than before. Because firms do only what makes them better off, if innovation does not make firms better off, it will not happen. *See, e.g.*, Harold Demsetz, *Industry Structure, Market Rivalry, and Public Policy*, 16 J.L. & ECON. 1, 3 (1973) ("[A] firm that seizes an opportunity to better serve customers does so because it expects to enjoy some protection from rivals . . . . To destroy such power when it arises may very well remove the incentive for progress."); McGEE, IN DEFENSE, *supra* note 74, at 46 ("If superior knowledge will not be permitted to pay, why strive to be first?").

This fallacious argument is closely related to the mistaken notion that reward above cost is necessary to induce production. *See supra* note 56. There is nothing wrong with the argument that an innovator must expect to earn a greater return after innovating than before, otherwise she has no reason to prefer innovation. The fallacy arises when this is used to justify monopoly profit, even for a short time (unless "vanishing" is meant by "short"). A firm will prefer innovation to stagnation whether it expects to earn a penny more than its competitive return or a billion dollars more. Because more compensation means more inefficiency, the whole point of antitrust is to ensure that production occurs for the smallest reward consistent with making producers willing to produce. The cost toward which antitrust pushes markets includes enough return to make production happen, and not a penny more.

To return to the example in note 56, if the innovator already earns $1.00 in profit in a competitive market, she is quite indifferent to innovating if she can expect to earn only $1.00 in profit in the same market after introducing her innovation. But assuming that her only alternative to innovating really is to continue to earn $1.00, she will be just as sure to innovate for a profit of $1.01 as for a profit of $1.01 billion. To fail to distinguish between the two is to ignore the entire reason for the existence of antitrust.

To the extent that the opponents of antitrust were making this pure incentives argument and not worrying about covering fixed costs, the critique in this note suffices.

Diagram 2 shows that as a theoretical matter it is possible to imagine markets for which social value is less at the monopoly price than it would be at a lower price. In the top graph of the middle column, antitrust imposes a high level of enforcement, driving price very close to C. The opponents of antitrust wish to argue that relaxing enforcement will cause firms to generate more quasi-profits, which in turn will allow them to invest more in product improvement and thereby to increase social value and efficiency. If their point is that antitrust enforcement should be relaxed to the intermediate level depicted in the middle graph, then they are right. At that level, the ability to charge higher prices has induced firms to invest in product improvements so substantial, and at such low R&D cost, that output has increased despite the firms' increased monopoly power, and social value has expanded.

But if the point of the opponents of antitrust is that enforcement should continue to be reduced below the intermediate level depicted in the middle graph of the middle column, then they are wrong. The bottom graph shows that although the opponents of antitrust are quite right that less antitrust enforcement relative to the intermediate level may drive up quasi-profits, investment in R&D, product quality, and demand, it does not follow that social value and efficiency will increase as well. In the bottom graph, demand has indeed increased relative to the middle graph, but the increase is feeble despite a much higher increase in fixed costs relative to the change between the top and middle graphs. This reflects diminishing returns to R&D. The increase in deadweight loss incident to the increase in monopoly power is also large. The effect of higher fixed costs and deadweight loss erases the benefits of product improvement and constricts the size of the social value area relative to the middle level.

The model tells us that it is perfectly reasonable to think that in some or perhaps many markets some monopoly power is good for society, but too much is bad. It makes as much of a case against too little antitrust enforcement as it makes against too much.

IV. INPUT CONTROL AS A NECESSARY CONDITION FOR ANY KIND OF MONOPOLY INEFFICIENCY AND THE INCONSISTENT WAY IN WHICH ANTITRUST RESPONDS TO IT

Part II argues that antitrust treats cartels and monopolies inconsistently because it condemns some forms of collusion per se but contains an exemption for property-based exclusion, even though collusion and exclusion are not meaningfully distinct. This Part is devoted to making these arguments more precise by restating them in relation to the notion that actual control over inputs is a necessary condition for monopoly

power. In the process, this Part shows how much of antitrust law and economics may be categorized in terms of different forms of input control and transitions thereto.

A. *Types of Exclusion: Input Control and Product Transformation*

We can think about the ability to charge P in any given market as being due to the ability of a single profit-maximizing entity to exclude competitors from all the actual and potential inputs in the market. Actual inputs are inputs actually used in production. Potential inputs are all the inputs that could be used in production but are not. You need to exclude competitors from the actual inputs to charge P because the revenues generated thereby give you a motive for charging P. You need to be able to exclude competitors from the potential inputs to charge P because if you do not then a competitor could use them to increase output in the industry and drive price below P.[76] The power to exclude competitors from potential inputs is a necessary condition for any kind of monopoly inefficiency.[77,78,79,80]

---

76. Excluding competitors from actual inputs does not create monopoly power because these inputs are actually used in production; they drive down price.

77. The necessity of input control follows directly from the definition of input. It needs neither empirical support nor any model other than the one that associates monopoly inefficiency with a reduction in output (e.g., Diagram 1). An input is any element of a sufficient condition for production. It follows that the absence of production associated with monopoly inefficiency is possible only if enough such elements are unavailable to violate a necessary condition for production. Otherwise, a sufficient condition for production would be satisfied, production would take place, and price would fall. (What about access to buyers? See note 78.)

78. When antitrust thinks about production, it sometimes thinks of the processes that lead to the creation of the product that defines the relevant market, and leaves out processes that bring about the sale of the product. It is for this reason that antitrust sometimes distinguishes between access to suppliers and access to buyers or between access to supply and access to distribution. *See, e.g.,* GAVIL ET AL., *supra* note 36, at 593–94. This suggests that we cannot attribute all monopoly inefficiency to input control; control over access to buyers would seem to be an independent source of monopoly inefficiency.

This is no objection to the necessity of input control, however, because access to buyers is a production input. In economics, as opposed, perhaps, to engineering, production does not mean only the creation of the product, but the creation of the sale of the product. It extends all the way through manufacture and all the way through distribution, and includes even the process of bending the will of the consumer to agree to the sale (i.e., advertising). That is why all these items count as producer costs. A firm that uses control over access to buyers to keep competitors from expanding production when it raises price is using input control to charge P.

The reason antitrust is interested in the production of sales, and not just the production of products, is that it is the ability to produce sales, and not just units of product, that determines price. A competitor may produce infinite units of a product, but the competitor will not drive price down by even a penny if it can bring none of the units to market.

79. This is an attempt to generalize the observation in the "raising rivals' costs" literature that "restricting the supply available to rivals of a key input without similarly restricting the amount available to satisfy the [monopolizing entity's] demand" can lead to monopoly inefficiency. Thomas G. Krattenmaker & Steven C. Salop, *Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power over Price*, 96 YALE L.J. 209, 230 (1986). This literature has argued successfully

that input control due to vertical contracting or integration is a source of monopoly inefficiency. *Cf. id.* at 215 (limiting discussion to the "horizontal effects of vertical contracts") (internal quotation marks omitted). My point is that input control, whether maintained by contract, direct ownership, or any other means, is a necessary condition for all monopoly inefficiency, both that traditionally considered "vertical" and that traditionally considered "horizontal." The foundation of the power of any monopoly, cartel, or oligopoly is input control. The raising rivals' costs literature has not gone this far, in part because it has not fully grasped the identity of collusion and exclusion. *See infra* note 103.

80. For simplicity of exposition, the discussion of input control that follows will tend to avoid talk about costs. Costs are, however, key to the ability of input control to serve as a source of monopoly profit. If the price at which you buy control is too high, there is no profit. *Cf.* Krattenmaker & Salop, *supra* note 79, at 238 (noting that when a firm tries to buy up inputs in a competitive market, it drives up price, and "[t]his higher price is paid by both the purchaser . . . and its rivals"). It should be assumed throughout the text that input control is always purchased at a cost that makes monopolization profitable. Similarly, the discussion will tend to say things like "X can charge P because X controls all the inputs." This is a heuristic for "X can charge P because X acquired all the (sufficiently) low cost inputs."

Both input control and the lower cost requirement are necessary conditions for monopoly inefficiency. The first guarantees exclusionary power. The second creates an incentive to obtain and exercise it. If a firm naively attempts a supply squeeze by buying up so much of a competitively supplied input that all other competitors are priced out of the input market, it will have succeeded at using input control to exclude competitors. That it has done so at a ruinous price that will not allow it to recoup its costs at the monopoly or any other price is quite irrelevant to the question whether it has obtained exclusionary power through control over inputs. But a sophisticated firm will not attempt such a squeeze unless it can obtain input control at a profitable price. So, for example, it might do it only if it can enter the input market in advance of the competition and buy up all the lower cost inputs. *Cf. id.* at 234–35 (the raising rivals' costs "bottleneck" model).

The role of costs as another necessary condition for monopoly power highlights a problem with input analysis generally. What allows one firm to obtain inputs at lower cost than another is inevitably that the successful firm has some other input to which the unsuccessful firm does not have access. For example, the firm that is able to enter the input market first and thereby obtain a lower price is only able to do so because it has the "first mover" input and other firms do not, which is just another way of saying that it has the first mover input at lower cost than the competition. But the guarantee of that must be some other input to which it again has access at lower cost. Turtles all the way down is not an acceptable explanation, and this exposes the limitations of the practice of explanation itself. At some point we must simply stipulate that the successful competitor has some input at a lower cost than the rest without trying to go behind it and find its foundation.

In other words, in order for monopoly inefficiency to exist you have to make advantage an axiom of your model. You can do this in the final goods model simply by stipulating that a seller has monopoly power. For the purposes of antitrust, this should be enough. Where there is power there is inefficiency. But antitrust has not taken this route, and instead has sought to step at least one level back into the input market in distinguishing between favored and unfavored forms of control over inputs. It has therefore been necessary for this article to chase antitrust down the rabbit hole and locate the source of power in the input market. But it must be recognized that at some point the existence of monopoly inefficiency must be either admitted or denied. It cannot be explained recursively forever.

The Chicago School has been quite aware of the need to stipulate advantage and was eager to give it a moral valence by calling it superiority or talent. *See supra* note 74. One firm may have the talent or superiority input. It comes from God at no cost to its owner and at an infinite price to everyone else. This seems to suggest that monopoly profit is (1) always justly deserved and (2) necessary for superiority to exist. But it really just reduces to the truism that monopoly profit is possible only when advantage is possible.

We can identify three ways in which an entity may be able to exclude from potential inputs. First, it may own all the potential inputs outright and exclude competitors by exercising its property rights directly.[81] Second, it may use contract, friendliness, fear, trickery, or some other means to coerce the owner or owners of the potential inputs to exclude competitors.[82] Third, it may exert informal control directly over the potential inputs, bypassing their true owner or owners, as, for example, when gangsters physically seize and dispose of other peoples' property.

Saying that one of these three means of excluding from potential inputs is necessary for charging P does not mean that firms always get to P by expanding their control over potential inputs in a given market. Firms often get to P not by taking control of all the potential inputs in a given market but by transforming the product in a given market into one for which they already control or may be able to achieve control over all the potential inputs.

Innovation is an example of this. A firm might create a superior product that causes consumers to abandon older offerings in a particular market or that creates a new market. The firm might control all the engineers who know how to make the superior product or all the reserves of a particular mineral essential to its production. By creating the superior product, the firm transforms the product in the market into one for which (or creates a new market for which) it already controls all the potential inputs. In this way, it achieves the ability to charge P.

Tying is another example. When the hotel ties rooms to meals, it changes the product in the market from meals to rooms-and-meals. Because the hotel controls all the rooms in the market, this transformation allows it to charge P. Similarly, when a firm engages in predatory pricing, it makes it impossible for any incumbent firm to survive in the long run without controlling all the potential inputs in the market (because only monopoly pricing will afford recoupment). This amounts

---

With respect to the first contention, it is only necessary to say that there is no doubt that advantage is advantageous, but whether it is deserved is·a separate issue. One man's talent is another's accident. The second contention is flawed because it tries to turn a conditional into a biconditional. To conclude from the need for advantage to exist in order to get monopoly profit that monopoly profit is necessary for advantage to exist misses the whole point of antitrust, which is that eliminating monopoly inefficiency results in net gains for society precisely because lower cost implies a willingness to work for less. *See supra* notes 56 & 75. It is precisely because you have an advantage that you are willing to exercise it for less than a monopoly profit. It is because your costs are lower that we can get away with lowering your revenues.

81. This is the mode of control that raising rivals' costs largely ignores. *See supra* note 79. It is referred to in this article variously as "single firm monopoly," "direct control," "input control based on direct ownership," "ownership centralization," or "property-based exclusion."

82. With its focus on contract, the raising rivals' costs literature lives here. *See supra* note 79.

to changing the nature of the product. A predatory pricer does this if she believes that the pressure put on the market by such a product transformation will result in competitors ceding control of their inputs to her.[83,84]

Antitrust often fixates on these product transformations and forgets that regardless of how a firm has managed to achieve a position in which it happens to have control over all the potential inputs in a given market, a necessary condition for the maintenance of that control is always that it be able to continue to use one of the three methods defined above to exclude competitors from potential inputs. One of those three methods is always a necessary condition, though not necessarily the sole condition, for the charging of P.

Unless the market into which the product has been transformed is a natural monopoly market, in which case undoing the transformation may be the only way out,[85] antitrust always has the option of responding by loosening control over potential inputs, rather than undoing the transformation. Antitrust can either dissolve the tie, or give competitors access to the hotels' spare rooms. In some cases, loosening control over potential inputs is the only viable option. Antitrust cannot stop the superior product from selling at P by banning the superior product and forcing a return to the inferior one.[86] To stop the superior product from selling at P, antitrust must grant access to potential inputs.[87]

---

83. Of the types of exclusionary conduct identified in Part II.B as not being covered by the exemption for refusals to deal, all except for exclusive dealing are examples of exclusion through transformation. By contrast, exclusive dealing works not by transforming a given market but by using contract to secure control over the inputs in the market. For a full mapping of exclusionary conduct claims onto the categories defined in this Part, see *infra* Part IV.B.

84. Krattenmaker and Salop sketch a model in which a supply squeezer buys up inputs in a competitive input market, thereby driving up input prices. Krattenmaker & Salop, *supra* note 79, at 236–38. They observe that "[t]his higher price is paid by both the purchaser of the exclusionary right and its rivals. It does not follow, however, that the purchaser gains no anticompetitive advantage. Competitors' cost increases may be larger if the purchaser uses the input less intensively . . . ." *Id.* The input control over which confers monopoly power in this case is not the input purchased by the supply squeezer but the technology that allows the squeezer to use that input less intensively. If the competition had this technology, the squeezer would earn no monopoly profit. The act of buying up inputs in this example counts as market transformation. It changes the structure of the market into one for which access to the technology that uses the input less intensively is now essential.

85. In the case of predatory pricing, reversing the transformation is the only option; giving competitors access to potential inputs without stopping the low pricing will not achieve entry because the low pricing creates a natural monopoly market.

86. The money spent developing the superior product would be wasted.

87. The necessity of input control applies equally to unnatural and natural monopoly. Antitrust defines a natural monopoly market as one for which the cost of producing the efficient level of output is lower when the entire output is handled by a single firm than when it is divided in any way between two or more firms. *See, e.g.,* BRAEUTIGAM, *supra* note 32, at 1294–96. Antitrust calls this a subadditive cost industry. *Id.* When antitrust thinks about subadditive costs, it

usually thinks about the classic example of increasing returns to scale technology. When technology exhibits increasing returns, it is always cheaper to produce more, so a single firm serving the entire market will always have lower costs than two or more firms trying to serve it.

Subadditive costs do not imply monopoly pricing, only that the market will be efficient only when occupied by a single firm. Arguments originating in the Chicago School suggest that a single firm monopolizing a subadditive cost market cannot charge a monopoly price if entry is quick. *Cf. id.* at 1301–05 (discussing "Demsetz competition" and contestible markets). As soon as price rises above cost, a single competitor waiting in the wings may enter and supply the entire market at a lower price. So price stays low.

A single firm monopolizing a subadditive cost industry can only charge a monopoly price if it can deny potential entrants access to inputs. *Cf. id.* at 1302 (entry into regulated natural monopoly market possible only if all inputs "available to all bidders in open markets at competitively determined prices"). It may do this in any way in which any monopoly may do it, whether by colluding with owners of potential inputs or owning the inputs outright.

Antitrust's tendency to associate natural monopoly with subadditivity fails to capture the full scope of the notion of monopoly that arises naturally. Calling a monopoly "natural" suggests that it has pricing power that arises without, or perhaps in spite of, human intervention. A natural monopoly is a monopoly that we cannot efficiently expose to the discipline of competition, no matter how we may try. Because monopolies in subadditive cost markets may be subject to pricing discipline through competition with competitors waiting in the wings, subadditivity does not always imply natural monopoly. (For the rest of this note, "natural monopoly" will have the broader meaning of monopoly that cannot be exposed to competition. Natural monopoly in the traditional sense of monopoly in subadditive cost markets will be referred to simply as monopoly in a subadditive cost market. Other than in this note, however, this article will continue to use the term natural monopoly in the narrower subadditive cost sense.)

What determines whether a monopoly is really natural in the sense that it cannot be exposed to competition is whether there exist potential inputs that competitors may use to enter the market in the event that the monopoly charges a monopoly price. If there are no such inputs, then competition is impossible. This situation is distinct from that of "unnatural" monopoly because in that case potential inputs exist, but because they are under the control of the monopolist, competitors have no access to them. Monopoly in that case is unnatural because there is something that the monopolist can do to promote competition: it can release control over potential inputs. In the case of natural monopoly, what keeps potential inputs away from competitors is not that the inputs are under the control of the monopolist but that they simply do not exist. The monopolist could not give competitors access to potential inputs even if it wanted to do so.

From this perspective, monopolies in subadditive cost markets are only natural monopolies to the extent that competitors cannot wait in the wings to discipline the incumbent because there are no potential inputs for competitors to use to do so. For example, speed can be a missing input that prevents competitors from entering to discipline a monopolist in a subadditive cost market. When there is nothing the incumbent in such a market can do to speed entry in response to its high prices, it is a natural monopoly.

Antitrust encounters natural monopoly in two contexts apart from that of subadditive cost markets, but uses other names for, and does not always seem to recognize the connection between, them. One is Ricardian rents. Antitrust speaks of Ricardian rents whenever it is impossible for industry output to reach the efficient level of production. What makes it impossible for output to reach the efficient level must be an absence of inputs. *Cf.* VARIAN, *supra* note 10, at 410–15 (associating rent with the existence of inputs in "fixed supply"). The results are the same in this case as in any case in which input control is used to drive up price. Competitors have no access to inputs, cannot enter the market, and cannot drive price back down.

The major difference is that there are no potential inputs available to anyone, even incumbent firms. We might say in this case that God is the monopolist because he denies access to potential inputs to the world. For example, by making it impossible for us to produce, or even imagine, a situation in which there might be enough supply of land in a given place to satisfy demand, God ensures that whoever happens to own the land will be able to earn a monopoly profit. But a

B.   *The Old Types, the New, and the Law: A Patchwork*

We have seen that a necessary condition for charging P is centralized control over all the actual and potential inputs in a market, whether it is brought about through (1) direct ownership, (2) indirect influence over the owners of inputs, or (3) informal control over the property of others. We have also seen that this centralized control may be achieved either by acquiring centralized control for a given product or transforming the product into one for which the entity already has or can more easily achieve control.

The patchwork coverage of the antitrust laws is reflected in the mapping of traditional categories onto this metaphysics of input control. The refusal to deal doctrine exempts property-based exclusion, which means that, roughly speaking, it exempts (1) per se.[88] Hard-core cartelization (per se illegal),[89] oligopolization (per se legal),[90] and exclusive dealing (case-by-case legal)[91] all count as (2). (3) is not regulated by antitrust, but by other legal regimes (e.g., there are general criminal and civil penalties for threats).[92] Antitrust exempts some transformations per se (e.g., product improvements)[93] and condemns others per se (e.g., predatory pricing[94] and tying[95]). The first three columns of Chart 2 on

---

situation in which potential inputs needed by competitors to enter the market are lacking is precisely the problem of natural monopoly defined above.

The other context in which antitrust encounters natural monopoly is that of unshareable essential inputs. The Chicago School was much preoccupied with these, particularly in the form of management talent or genius. *See supra* note 74. Consider a management team superior to that of any competitor and that only works well when working for the incumbent. For a competitor to enter, it must have that team, but the team cannot be transferred or shared without losing its superiority. The competition cannot enter because it lacks the "superior team" input. But a situation in which potential inputs needed by competitors to enter the market are lacking is precisely the problem of natural monopoly defined above. The problem of unshareable essential inputs is the problem of natural monopoly.

Recall that the conduct requirement exempts those firms from the antitrust laws that cannot do anything to admit competitors into their markets without driving up their own costs. *See supra* Part II.B and note 29. This means that the conduct requirement amounts to an exemption for natural monopoly broadly defined.

(The foregoing applies with equal force to natural oligopoly.)

88. *See supra* Parts II.B and II.C.

89. *See supra* note 15.

90. *See id.*

91. *See generally* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 484–88.

92. *See supra* note 15.

93. *See infra* Part IV.E.

94. Predatory pricing is not usually recognized as per se illegal, largely because the controversy over how to test for the existence of predatory pricing devours antitrust's attention in this area. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at pp. 372–80. But in the rare case in which a plaintiff proves predatory pricing, courts do not proceed to balance efficiencies that might result from the practice (e.g., investment of monopoly profits in product improvement) against costs. *See id.* at 371 n.6 (cases cited therein). It is for this reason that I categorize the practice as per se illegal.

page 384 summarize the mapping of the law onto these categories.

There are three ways in which antitrust may drive price down to C: (I) decentralize control over potential inputs in the market; (II) transform the product into one for which control over potential inputs is already decentralized; or (III) directly force whoever controls the potential inputs to charge C (i.e., government price-setting by fiat).

The great benefit of doing (III) is that it may be applied to resolve any unnatural monopoly inefficiency. If antitrust were to do (III), then it could respond to cartelization not by dissolving the relationship between members but simply by enjoining the cartel to charge C. Similarly, antitrust could respond to monopolization not by forcing an end to exclusionary conduct or breaking firms apart, but by enjoining the monopoly to charge C. Antitrust refuses to do (III), however and professes to leave that to the legislature, which can mandate rate regulation.[96] Instead, antitrust sometimes does (I) (e.g., in cartelization and exclusive dealing cases, and in transformation cases in which breakup is applied) and sometimes (II) (e.g., in predatory pricing and tying cases).

C. *Collusion Excludes, so Exclusion Is Not More Fundamental Than Collusion*

Antitrust is sometimes troubled that in the collusion story that it tells itself it must assume that there are barriers to entry that are distinct from the collusion itself. As antitrust sees it, a cartel cannot charge P unless there is some barrier in the market to prevent competitors from entering in response and driving price down. This means that there must be some exclusion present to create the barrier. This has led some scholars to declare that exclusion is more fundamental than collusion.[97] That is a step in the right direction, but it still misses the point that collusion is itself exclusion. Cartelization itself is the barrier that makes cartelization exclusionary.

A huge stumbling block in understanding cartelization has been a fixation on the terms of the cartel agreement. Cartels agree on price, which is why cartelization is often called price fixing. But the agreement on price is a symptom, and not the cause, of the ability of the cartel to charge P in quite the same way that the raising of price to monopoly levels by a single firm monopoly is a symptom, and not the cause, of the ability of the single firm to charge P. The cause is the centralization of control over all the potential inputs in the market under the authority of a

---

95. *Id.* at 435–36.

96. *See supra* note 32.

97. Thomas G. Krattenmaker et al., *Monopoly Power and Market Power in Antitrust Law*, 76 GEO. L.J. 241, 249 n.47 (1987).

single entity.[98,99] This centralization is what lets cartels and monopolies exclude.[100]

Cartelization is the pooling of control but not title over inputs under the authority of a single entity effectuated through an agreement between owners or those who control owners. Cartelization allows firms to charge P because it excludes all by itself. It takes potential inputs out of the hands of competitors and puts them under the control of a single entity.[101]

Consider all the potential inputs in a market that may be deployed in response to P to drive price back down to C. Assume the ownership of these inputs is distributed among a group of firms. Without cartelization, if any one firm cuts output to raise price, the others draw on their potential inputs to drive price back down. The cartelization of these firms puts these potential inputs under the control of an authority that is not interested in deploying them to drive price down in the event that P is achieved. Because these are all the potential inputs in the market, competition is excluded. Similarly, a monopoly excludes because it owns all the potential inputs in the industry. The inputs are off limits to competitors, so the monopoly can charge P.

Of course, if only a subset of owners of the potential inputs

---

98. In the case of cartelization, this entity is "the cartel"; in the case of monopolization this entity is the single firm.

99. The grid below matches cause and effect in cartel and monopoly cases:

	Cartel	Monopoly
Cause	Centralizing control over inputs via contract	Centralizing control over inputs via property
Effect	Charging P	Charging P

Because the law makes price fixing per se illegal, it seems to suggest that the charging of high prices is the source of a cartel's monopoly power. The law therefore suggests that the effect is the cause. The equivalent in the single firm case would be for the law to make it illegal for monopolies to charge high prices. Consistency would require that antitrust ban both price fixing by cartels and price raising by monopolies or both input control by cartels and by monopolies.

100. Thinking that only single firms exercise monopoly power, the Supreme Court has announced that "[t]he mere possession of monopoly power, and the concomitant charging of monopoly prices, is not only not unlawful; it is an important element of the free-market system." Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 407 (2004). But the power of a cartel is indistinguishable from the power of a single firm; both are built on input centralization. So the Supreme Court cannot be right here. A cartel is a contract-based monopoly; but it is unlawful per se.

101. The same is true for oligopolization. The fact that some types of oligopolies are able to charge only something less than P is irrelevant. Some cartels, too, may be imperfectly organized or enforced, with the result that they are unable to raise price all the way to P. Similarly, monopolies may have organizational problems that prevent the charging of P. We can think of such limits on the ability to reach P as corresponding to limits on the entity's control over inputs. There is a spectrum of control from zero to full. Any given cartel, oligopoly, or monopoly falls somewhere in this spectrum.

cartelizes, then to charge P they will need to use some other method to gain control over the remaining potential inputs in the industry. But this is not because in cartelizing they have not already excluded. So long as they pool control over potential inputs, they exclude competitors from that common pool.[102] That they have to do more arises from the fact that they were not able to use the particular exclusionary practice called cartelization to exclude from all the potential inputs in the market. To go the extra mile they might, for example, purchase the remaining potential inputs outright, threaten their owners with violence, or engage in exclusive dealing with them. Cartelization is not the only way in which cartels may exclude, but it is a way in which cartels necessarily exclude.[103,104]

---

102. For simplicity, this article sometimes suggests that controlling all the potential inputs in a market is a necessary condition for any amount of monopoly power. In practice, controlling some but not all inputs can often lead to some amount of monopoly power. *Cf.* CARLTON & PERLOFF, *supra* note 10, at 110–19 (discussing dominant firm/competitive fringe markets). Also, some potential inputs may be essential, meaning that production cannot take place without them. If you control all the essential inputs, you have monopoly power, even if you do not control any other potential inputs. Similarly, if you control some essential inputs, you may end up with some, but not full, monopoly power.

103. Krattenmaker et al. distinguish "Stiglerian" and "Bainian" power. Krattenmaker et al., *supra* note 97, at 249. Stiglerian power is expressed when a "firm or group of firms . . . raise[s] or maintain[s] price above the competitive level directly by restraining its own output . . . ." *Id.* Bainian power is when a "firm or group of firms . . . raise[s] price above the competitive level . . . by raising its rivals' costs and thereby causing them to restrain their output . . . ." *Id.*

This is a study in the confusion of cause and effect. It is rather like saying that the lady lying two feet above the ground and the bed she is lying on are two distinct causes of her lying two feet above the ground. But the bed is the cause here, and her lying two feet off the ground the effect. Stiglerian power is not power at all; it is a symptom of power. We know that the ability to raise price by restraining output is conditional on having the power to prevent rivals from increasing their output in response. So when you encounter Stiglerian power in the sense that you encounter an entity that is raising price above the competitive level, the next question must always be "how is the entity able to prevent rivals from increasing their output in response?" The answer is what Krattenmaker et al. call Bainian power and what this article calls input control.

The pity of the Stiglerian/Bainian distinction is that Krattenmaker et al. are quite well aware that Stiglerian power is not really an independent type of power. They write that because "successful price fixing of significant duration depends on the existence of constraints on new entry," Bainian power is a necessary condition for Stiglerian power. *Id.* at 249 n.47. But through the qualifier "significant duration" they cling to the notion that Stiglerian power is power nonetheless. In the long run, it is really an expression of Bainian power, but in the short run, they seem to say, Stiglerian power is real and independent. *See id.*

It is not clear why they think time matters. Perhaps the idea is that competitors experience frictions in responding to an output cut by a competitor. They may not have had advance warning about the output cut, and it takes time to acquire the inputs necessary to ramp up output in response. But we can think of advance warning as an input, in which case the source of the power of the output cutter to raise price is that she has this input but others do not. So there is input control and Bainian power here after all.

But even this first mover advantage is sometimes predicated on another source of Bainian power. Imagine a firm or group of firms that cuts output by 30% to raise price. Perhaps it has done nothing to tie up potential inputs that it does not already own, and perhaps before cutting output, it owned no potential inputs itself. It would seem to have no Bainian power whatsoever other than a first mover advantage. But as soon as it cuts output, it comes to control precisely the amount of

D. *Pro Monopoly and Anti Cartel, Starkly*

Here is another way to understand the bias against cartelization and in favor of monopolization. Consider a cartel and a monopoly. Assume

---

potential input that a competitor would need to increase output in response and drive price right back down. Those inputs, of course, are the excess capacity that the entity takes offline when it reduces output. To the extent that it is possible for the entity to throw these inputs back into input markets but does not because of the exemption for refusals to deal, it exercises Bainian power simply in virtue of having reduced output.

Without this power, it could not raise price, even for an instant. Imagine that there were no exemption for refusal to deal. As soon as the cartel had reduced output, it would be required to put the right to use its excess capacity on the market at a competitive price or even for free. If a machine, perhaps it would be hard to run without experienced engineers employed already by the cartel. But those experienced engineers themselves have excess capacity now that the entity has reduced output, so their services too would be required to go on the input market and could be purchased by competitors. In short, without the exemption for refusals to deal, competitors could in effect immediately buy the decision to ramp output right back up to the competitive level. The notion that it takes time for entry to drive price back down is predicated on the assumption that firms have a right to refuse to deal. This makes it necessary for competitors to respond by expanding their own operations, and thereby gives the cartel's first mover advantage its power.

I submit that it is because of the bias in favor of property-based exclusion that this was not immediately evident to Krattenmaker et al. The entity that cuts output exercises Bainian power by using property rights over the inputs it had hitherto intended to use up in production to exclude rivals. Because we tend to assume that property is inviolable, we do not give it the relevance that it deserves for antitrust. Instead, it operates as a blindspot of Freudian magnitude over the entire discipline. It is the reason for which, when the bed is anything other than property, we see a woman lying on a bed, but when it is property, we see a woman levitating two feet off the ground. When firms seem able to raise price without excluding, it is usually because they are using property rights to exclude.

104. Baker recognizes that "exclusion and collusion are complementary methods of obtaining market power." Baker, *supra* note 14, at 557. But, like Krattenmaker et al. (see *supra* note 103), he fails to grasp the identity of the two forms. He argues that there are three exclusion problems (means, sufficiency, and profitability), and that these differ from three cartel problems, which he identifies as reaching consensus, deterring deviation, and preventing new competition (i.e., exclusion). Baker, *supra* note 14, at 572.

The argument has two flaws. First, the problems of reaching consensus and deterring deviation are not unique to cartels. They amount to the problems of how to centralize and maintain control over inputs. A monopoly faces these problems too. The difference is only that the monopoly solves them through direct ownership whereas a cartel solves them via contract. It is antitrust's failure to problematize direct ownership that makes it appear as if these problems do not exist for monopolies. But they do. *Cf.* WILLIAMSON, *supra* note 21, at ch. 6 (discussing challenges of organizing production within a single firm).

Second, although Baker recognizes that the cartel problem of preventing competition amounts to the problem of engaging in exclusion, he treats exclusion as a non-essential element of cartelization. Baker, *supra* note 14, at 558 & n.150 ("Colluding firms may need to exclude in order for their collusive arrangement to succeed."). Thus he fails to recognize that input control is a necessary condition for any kind of economically meaningful collusion. If collusion does not exclude from potential inputs, it can have no effect on price.

If we remedy these deficiencies in the argument, then it becomes clear that collusion and exclusion face in common all of the problems that Baker identifies. Exclusion must solve the problems of consensus and deviation, as well as those of means, sufficiency, and profitability that Baker identifies. And in addition to solving the problems of consensus and deviation identified by Baker, collusion must necessarily also solve the exclusion problems of means, sufficiency, and profitability.

that there are only three potential inputs in the market. They might be land, minerals, buildings, IP, or something else. The cartel controls the three inputs because each cartel member controls one of the three (either through outright ownership or control over an owner) and the cartel controls the cartel members; in the monopoly case, the monopoly owns all three. From an economic perspective, the cartel and the monopoly are able to charge P for the exact same reason: each controls all the potential inputs in the market.

When antitrust intervenes to dissolve the cartel, it interferes with control over inputs just as surely as it would if it were to intervene to break up the monopoly. In both cases, control over inputs confers value on the holder of that control. And yet as an intuitive matter the interference seems much more problematic in the case of the monopoly than in the case of the cartel. One is inclined to see the breakup of the monopoly as theft or a taking, but not the dissolution of the cartel.[105] The reason the breakup of the monopoly seems more problematic is that the inputs are owned outright by the monopoly, while they are not owned at all by the cartel. This is the heart of the intuitive bias in favor of outright ownership as a mode of control over inputs. The result is that if you merely control the inputs you do not get protection against antitrust law, but if you own them, then you do.

E.   *The Refusal to Deal Exemption as Modified by Merger and Transformation Law*

We have seen that centralized control over all the actual and potential inputs in a market is exclusionary, whether it is brought about through direct ownership, indirect influence over the owners of inputs, or informal control over the property of others. The exclusionary character of the centralized control holds for the duration of that control, regardless whether one is at the start, middle, or end of it. Putting inputs off limits to competitors excludes competitors equally and for the same reason at the time the inputs are placed off limits as while they are subsequently kept off limits.

If we ignore product transformation, then the starting points for each of the three types of centralization of control are acquisition (merger), collusion,[106] and racketization,[107] respectively. A cartel

---

105. For an exploration of this intuition, see *infra* Parts V.C and V.D.

106. For purposes of the rest of this article, collusion will mean pooling of control effectuated either through agreement (which would be cartelization) or by some other means. Collusion is therefore defined to include "tacit collusion," meaning that it includes oligopolization. *Cf. supra* note 12.

107. I mean racketization here in the sense of the commencement of exertion of direct control over inputs by an entity that does not own the inputs. This includes the traditional protection

excludes when it forms in precisely the same way as it excludes over its subsequent life. Similarly, centralized direct ownership, which we call monopolization, excludes when it forms in precisely the same way as it excludes over its subsequent life. This means that when a firm acquires ownership of inputs, which we call the process of merger, it excludes in precisely the same way as the merged firm subsequently excludes for the duration of its existence. There is not one type of exclusion when the merger takes place and another while the firms remain merged.

Excluding the possibility of changed circumstances, regulating the start of exclusion is the same as regulating it later on. If you are vigilant about prohibiting the start of cartelization, then you need not prohibit cartels at all. You will have no cartels because you will have nipped them all in the bud. Similarly, if you are vigilant about prohibiting the acquisition of centralized ownership over inputs, then you need not prohibit the status of centralized ownership at all. You can be sure that all instances of centralized ownership will be efficient because you will have culled all transitions to that status that would have produced inefficient instances of centralized ownership.

There are two ways in which a firm may come to have centralized direct ownership over all the inputs in a market. It may acquire such inputs or it may transform a market into one in which it has already acquired such inputs. Refusal to deal exempts continuing exclusion via centralization of direct ownership over inputs, but has nothing to say about how a firm comes to have centralized direct ownership. Merger and transformation law do that. Merger law regulates the achievement of centralized ownership through the process of acquiring inputs in a given market. The antitrust law of market transformation regulates the achievement of centralized ownership through the transformation of a market into one in which a firm already has centralized direct ownership.

If we accept for the moment the notion that the regulation of the beginning of exclusion suffices to regulate continuing exclusion as well (i.e., if we accept no changed circumstances), then we cannot think of refusal to deal as defining a real exemption to the antitrust laws; the exemption it provides to centralized direct ownership serves only to preserve centralizations that have already been approved by merger law or transformation law. To get a proper picture of what centralizations of direct ownership are exempt from antitrust, we must look at how merger

---

racket. Security is of course an input in every business venture. The protection racket takes direct control of that input though it is not legal for it to do so. Because that input is essential, by controlling it, the racket is able to force the business to charge P in the market and turn the monopoly profit over to the racket.

and transformation law operate to determine which centralizations of direct ownership can come into existence.

How, then, do merger and transformation law combine to treat centralizations of direct ownership? Merger law directs most of its scrutiny at acquisitions of inputs directly from competing firms.[108] Even then, however, it tends to ignore smaller acquisitions.[109] It directs much less scrutiny at acquisitions of inputs from non-competitors, which is to say that when competitors are not a firm's source of inputs, merger law tends to turn a blind eye to input acquisition.[110] Transformation law prohibits market transformations based on tying and predatory pricing,[111] but does not prohibit market transformations based on the creation of a superior product.[112] Putting these two regimes together, we see that the shape of the exemption for centralized direct ownership is somewhat different from the one we identified when we associated it with refusal to deal. Filtered through the lenses of merger and transformation law, the refusal to deal doctrine exempts only centralized direct ownership of inputs achieved through small incremental acquisitions from competitors, acquisition of inputs from non-competitors, and those achieved through transformations based on product superiority.

The foregoing applies only under the assumption of no changed circumstances. But it seems reasonable to assume that markets can be

---

108. *See, e.g.*, GAVIL ET AL., *supra* note 36, at 871–72 ("[T]he principal focus of agency and judicial scrutiny has remained and is likely to remain on horizontal mergers . . . ."). The level of scrutiny is case-by-case review for efficiency. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 551 (stating that courts condemn only mergers that create a threat to competition).

109. *See* U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, HORIZONTAL MERGER GUIDELINES 19 (2010), *available at* http://www.justice.gov/atr/public/guidelines/hmg-2010.pdf ("Mergers involving an increase in the [Herfindahl-Hirschman Index ("HHI")] of less than 100 points are unlikely to have adverse competitive effects and ordinarily require no further analysis."). A firm that acquires half a percent of the market at a time will always fall beneath the 100 HHI threshold. Although enforcers are not committed to ignoring such piecemeal acquisitions, the Guidelines do provide a rough sense of where their attentions fall. In this article, I treat non-scrutiny of creeping acquisitions as amounting to a de facto, per se rule of legality for them.

110. This is the realm of vertical merger enforcement. Although enforcers technically review vertical mergers for efficiency on a case-by-case basis, in practice there is little enforcement. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 430 ("Prevailing judicial opinion now seems to be that vertical mergers should be condemned only in the most extreme circumstances."); GAVIL ET AL., *supra* note 36, at 871 (describing scrutiny of vertical mergers in the 1980s and 1990s as "small" relative to scrutiny of horizontal mergers). In this article, I treat the low level of vertical merger scrutiny as a de facto, per se rule of legality.

111. *See supra* notes 94–95.

112. I use "superior product" here as a placeholder for a larger set of exemptions of similar type. *See* United States v. Grinnell Corp., 384 U.S. 563, 571 (1966) (exempting monopolization due to superior product or business acumen); United States v. Aluminum Co. of Am., 148 F.2d 416, 430 (2d Cir. 1945) (exempting monopolization due to "superior skill, foresight and industry").

transformed, not only by firms but also by other "natural" forces beyond the control of individuals or even groups of firms. Because transformation law reviews only market transformations caused by firms, it cannot intervene when a natural transformation happens to create a market for which a particular firm already owns all the inputs.[113] Merger law, of course, cannot intervene either in such a situation because the firm's acquisition of the inputs takes place before the market comes into existence or the firm enters.

In cases such as this, refusal to deal law ceases to be merely a way of guaranteeing that a determination regarding the efficiency of an ownership centralization is respected going forward. Merger and transformation law will never have had a chance to review the efficiency of the centralization. In such cases, refusal to deal doctrine has substantive meaning. As we have seen, it exempts the ownership centralization from review regardless of its efficiency.

If we admit the possibility of product transformations that are not reviewable by transformation law, then the antitrust exemption for ownership centralizations looks like this. It exempts any ownership centralization that is achieved by (1) natural transformation; (2) superior product transformation; (3) incremental acquisition from competitors; and (4) acquisition from non-competitors. In all these cases, the ownership centralization may be inefficient, but antitrust does not care.

The second row in Chart 1 on page 383 and the rows in Chart 2 on page 384 summarize antitrust's treatment of transitions to input control; Chart 2 considered as a whole shows the relationship between antitrust's treatment of starting points of input control ("changes") and continuations of it ("states").

F. *No Reason to Think That Superior Product or IP-Based
Exclusion Is Any More Likely to Result in Better
Products Than Other Forms of Exclusion*

The exemption for superior product transformations expresses a tendency of antitrust to assume that there must be a link between the source of monopoly power and its efficiency. Thus antitrust tends to assume that monopoly power derived from creating superior products or from patents is more likely to be used to create the best possible products than monopoly power due to other types of exclusionary activity, such as predatory pricing.[114]

---

113. *Cf. Grinnell Corp.*, 384 U.S. at 571 (exempting monopolization through "historic accident").

114. We have seen that input control must be present in all cases of monopoly power. *See supra* Part IV.A. So a product superiority transformation cannot exclude unless it creates a market

But there is no reason to think that this must be true. Exclusion of any sort can be used to create an anticipated pool of quasi-profit that a firm might then spend on product improvement.[115] The amount of quasi-profit that a firm will want to spend on product improvement depends on the cost of product improvement and the nature of demand for it; it does not depend on the mode through which the exclusion is achieved.[116] While it must certainly be the case that monopoly power based on a superior product must be associated with some amount of product improvement (otherwise there would be no exclusion), there is no reason to think that superiority-based monopoly power must result in a better product or more social value than monopoly power based on other types of exclusion.[117] Antitrust does not seem to have realized that it is not

---

in which the creator has such control. When antitrust thinks of product superiority as being linked to monopoly power, it is generally thinking of the case in which input control in the transformed market is due to direct ownership of the innovative idea (e.g., ownership of the secret piece of paper on which the idea is written or of a patent on it). But although product improvement transformations sometimes result in direct ownership of a scarce idea, this is not inevitable. For example, a firm might invent a superior product that rivals find it easy to invent around, even though none would have ever thought to invent around it if the firm had not invented its product in the first place. Controlling the idea either in trade secret form or via patent is useless as a mode of exclusion. Crucially, whether the product improvement results in ownable ideas or not is quite independent of the contribution of the improvement to social value. Some great ideas may not be associated with this sort of input control. *Cf.* HOVENKAMP ET AL., *supra* note 40, § 4.2 (observing that monopoly power is "[n]ot [i]nherent in [the] [i]ntellectual [p]roperty [g]rant"). To fund them, other forms of exclusion are necessary.

115. *Cf.* ROBERT P. MERGES ET AL., INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE 201 (6th ed. 2012) (citing authority to the effect that an inventor can profit from an invention by buying up assets that will become more valuable as a result of the invention and remarking that upon discovering an unpatentable type of bamboo capable of serving as a light bulb filament Thomas Edison sought to "lock up as many acres of production of the bamboo as he could").

116. Some forms of exclusion are associated with costs that do not go toward product improvement. For example, predatory pricing can be costly because the predatory pricer must price below cost before it can take control of the market. (Obtaining and defending patents, merging, or cartelizing are other costly forms of exclusion.) These costs eat away at the post-monopolization quasi-profit, reducing the amount left over for investment in product improvement. Inventing a superior product for which one happens already to own an essential input is not associated with such costs of exclusion and therefore leaves more quasi-profit for investment in product improvement.

This does not constitute an argument for privileging product superiority as an exclusionary mechanism, however, because product superiority is not the only low cost form of exclusion. Tying can be low cost as well. But even if product superiority were the only low cost form, there is still no argument for privileging product superiority on cost grounds. Because low cost exclusion is more profitable than high cost exclusion, we can expect firms to use high cost exclusion only when low cost is unavailable to begin with. Indeed, as a way to achieve product improvements when low cost exclusion is unavailable, high cost exclusion must increase social value. (Note that this discussion assumes that antitrust enforcers modulate the power of exclusion, whatever its form, to achieve efficient results.)

117. There is empirical support for the claim that "the expectation of patent protection [is] in most cases unimportant to R&D commitments." Scherer, *Conservative Economics and Antitrust*, *supra* note 74, at 38 (see also sources cited therein).

how you make your money, but what you do with it that counts for efficiency.

G.   *Pro Property and Anti Contract, Starkly*

It is an expression of the bias in favor of property-based exclusion that there is a tendency within law and economics to see property rights as the fundamental social guarantor of compensation for production. Without them, the story goes, thieves steal and producers have no incentive to produce. That, of course, is false. What matters is that producers have actual control over their input and output, not that their control be guaranteed by the state through a system of property law.[118] This means that any mechanism that guarantees control, whether property, contract, friendship, or gangsterism, guarantees compensation.

Whenever antitrust imposes a non-transformation remedy[119] in a given market, it directly adjusts the level of control over inputs of firms in the market. This is so when it dissolves exclusive dealing contracts that give a monopolist control over input suppliers. It is also so when it dissolves cartel agreements that give the cartel control over the inputs owned or otherwise controlled by its members. Other laws do the same thing when they send gangsters to jail for stealing inputs.[120] The reallocation of control over inputs is a core remedial function of antitrust.

As a legal regime oriented toward efficiency, antitrust deploys input reallocation in the service of maximizing social value. This means that it tries to compel allocations of inputs that confer enough monopoly profit to realize efficiency gains but not so much as to turn those gains

---

118. A classic example of the tendency to reduce control to property rights is this:

> Imagine a society in which all property rights have been abolished. A farmer plants corn, . . . but when the corn is ripe his neighbor reaps it . . . . The farmer has no legal remedy . . . . Unless defensive measures are feasible (and let us assume for the moment that they are not), . . . the cultivation of the land will be abandoned . . . . [L]egal protection of property rights creates incentives to exploit resources efficiently. . . . [W]ithout property rights there is no incentive to incur [production] costs because there is no reasonably assured reward for incurring them.

RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 32 (6th ed. 2003). The example suggests that property rights are necessary for compensation. In fact they are sufficient for it, but not necessary; control is necessary. The rationale for property rights is not control, but rather low cost control. When the state defends your land for you, you benefit from economies of scale. *See, e.g.,* ROBERT COOTER & THOMAS ULEN, LAW & ECONOMICS pp. 80–84 (5th ed. 2008).

119. Recall that there are two general ways to acquire monopoly power: by expanding control over potential inputs in a given market or by transforming a given market into one for which you already have control over potential inputs. *See supra* Part IV.A. Accordingly, there are two general categories of remedies for monopoly power: those that reduce an offender's control over potential inputs in the market and those that transform the market into one in which the offender has less control over potential inputs. A non-transformation remedy is one that falls in the former category.

120. *See supra* note 15.

into losses. We saw in Part III.A that antitrust will try to make the economy look like the middle graph in the middle column in Diagram 2. Depending on the market, a movement brought about by antitrust from the bottom to the middle graph might reflect, say, redistribution of control over five talented management teams or ten suppliers of essential inputs from an incumbent to competitors.

Given that antitrust has always been in the business of regulating input control in order to compel efficient levels of reward, it is very strange that antitrust tends to regard it as anathema to regulate input control conferred by property law in particular to compel efficient levels of reward. It is tempting to say that the reason is that input control due to property law is input control that has been conferred by the government, whereas input control due to gangsterism is not. But antitrust reallocates input control due to contract all the time, and the government confers the power to contract to the same extent that it confers the power to maintain title. Why is property sacred, if contract is profane?[121]

H.   *Denial and Avoidance: Cartels Refuse to Deal*

Antitrust is not used to thinking of refusal to deal in the context of cartelization, but there is no reason why, as a conceptual matter, it could not apply there. Cartels refuse to deal every bit as much as monopolies. We have observed that cartelization is in itself exclusionary because it puts the pooled inputs of the cartel members off limits to competitors.[122] It puts them off limits because if a competitor wishes to appropriate the inputs of any cartel member to drive down price, the cartel can in effect order the cartel member to refuse access to its property, which is to say, to refuse to deal.

We saw in the ownership centralization context that refusal to deal can be thought of as preserving the decisions of merger and transformation law.[123] Because cartelization is per se illegal, however, refusal to deal in the cartel context is useless because if the per se rule against

---

121. Phillip Areeda writes:

>   The trouble with both the general jury instruction ("if you find it exclusionary, you
>   can condemn") and the essential facilities notion is that they start with the
>   assumption that all business assets are subject to sharing. Do we really want to
>   assume that everything we have is up for grabs?

Phillip Areeda, *Essential Facilities: An Epithet in Need of Limiting Principles*, 58 ANTITRUST L.J. 841, 852 n.46 (1990). But what you "have" includes not only what you own, but what you control. Why put what you have by contract up for grabs but not what you have by title? If the point is that something, anything, should be sacred lest everything be taken away, is it not the point of antitrust that what is efficient is what should be sacred? Is it not enough to know that what you actually use, as opposed to what you leave idle for power, is never up for grabs?

122. *See supra* Part IV.C.

123. *See supra* Part IV.E.

cartelization does its job, there can be no cartels to preserve. For refusal to deal to have meaning in the cartel context, it would have to swallow the per se rule against cartelization. And that is precisely why antitrust never speaks of the exemption for refusals to deal in the cartel context.[124]

V.   AGAINST REFUSAL TO DEAL

A.   *The Cost of Breakup Is Not a Good Argument for the Refusal to Deal Exemption*

Antitrust's first reason for exempting property-based exclusion is that it thinks exclusion in general might be efficient. That argument is addressed in Parts II.A and III. Antitrust's second reason for exempting property-based exclusion is that it thinks that remedying property-based exclusion is too hard.

There are two remedies for property-based exclusion: (1) transforming the market into one in which inputs are not owned by any one entity and (2) forcing the entity engaging in exclusion to give up ownership of some of its inputs.[125] Antitrust hardly seems to be aware of (1); this is discussed in Part VII.D. Antitrust believes that remedying property-based exclusion is hard because it thinks that the only option is (2) and it believes that (2) is hard to do.

Antitrust gives three reasons for which (2) is hard to do. First, redistributing potential inputs from an incumbent to rivals might raise the cost of production, either because of economies of scale[126] or transaction costs.[127] Second, while in theory there might be a way to redistribute inputs that does not drive up production costs, it might be

---

124. Antitrust does have a law of concerted refusal to deal, but this applies only where there is no price fixing involved. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 239. And, in any case, this law does not create a per se exemption. *Id.*

125. This is referred to as "redistribution of input ownership" or "breakup" in this article. It is to be emphasized that redistribution of input ownership is a remedy for excessive monopoly power only when potential inputs are redistributed. If only actual inputs are redistributed, then there can be no effect on monopoly power. *Cf.* Areeda, *supra* note 121, at 844 ("We have to be very wary . . . , particularly when anything one has that another wants may be called an 'essential facility.'").

126. This and the following concern are generally raised in discussions of "deconcentration" or "divestiture." *See, e.g.*, POSNER, ANTITRUST LAW, *supra* note 8, at 112. Deconcentration and divestiture amount to a redistribution of inputs between competitors.

127. *See* WILLIAMSON, *supra* note 21, at ch. 6 (discussing managerial efficiencies associated with managing production within a firm as opposed to through the market). Posner identifies managerial skill as a "source[ ] of efficiency besides scale." POSNER, ANTITRUST LAW, *supra* note 8, at 113. But his argument is that if a firm with a 100% share of the market has low costs because of a talented manager, then breakup into two firms with equal shares, for example, will mean that only 50% of the market can enjoy such low costs going forward because the manager can only go to one of the two firms. This is, of course, an economies of scale argument.

expensive to identify that method of redistribution.[128] Third, determining
how much compensation to give the incumbent for the inputs taken from
it might be expensive or otherwise inappropriate.[129]

None of these arguments makes a case for a blanket antitrust
exemption for refusals to deal. The first argument fails because it is
actually concerned with natural monopolies, which are exempt under the
conduct requirement, and not the unnatural monopolies engaged in prop-
erty-based exclusion that are the concern of refusal to deal law.[130] Firms
that cannot be broken up without driving up costs are firms for which
rate regulation is the traditional remedy.[131]

The second argument fails for similar reasons. In the case of the
second argument, the offending firm has locked in the monopoly charac-
ter of the industry by creating a large organization that cannot be disen-
tangled for a price less than the savings that would be generated by
reducing monopoly inefficiency. Although ex ante the market is not a
natural monopoly, it is so ex post. And that is all that matters. The issue
here again is natural monopoly; refusal to deal doctrine is not
implicated.

The two foregoing arguments have power only if they can make the
case that single firm property-based exclusion tends to take place only in
a natural monopoly setting.[132] If all single firm monopolies are unbreak-

---

128. Here is a typical discussion:

> Divestiture measures . . . can be difficult to administer. In the most simple type of
> divestiture, a court can order existing organizational units within the firm to be spun
> off as separate entities. Such a move ordinarily will require some difficult
> judgments about how to allocate personnel and assets that serve the company as a
> whole, but there generally will be no need to sever existing design or production
> teams, and perhaps no need to separate physical facilities. In the harder case, the
> firm's operations are carried out in fully integrated teams. If a restructuring program
> is to be carried out, the court will have to decide how personnel who serve in the
> unitary teams will be allocated to the new enterprise, and how equipment and
> physical facilities will be divided.

GAVIL ET AL., *supra* note 36, at 754.

129. These concerns are generally voiced in the essential facilities context; but they apply to
breakup generally. *See* Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S.
398, 408 (2004) ("Enforced sharing . . . requires antitrust courts to act as central planners,
identifying the proper price, quantity, and other terms of dealing—a role for which they are ill
suited."); HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 339–40; ROGER D. BLAIR
& DAVID L. KASERMAN, ANTITRUST ECONOMICS 168 (2d ed. 2009).

130. *See supra* Parts II.B and II.C.

131. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 33.

132. Posner seems to recognize this. Accordingly, he argues against breakup on the ground that
natural monopoly is the only explanation for *"persistently* high concentration." POSNER,
ANTITRUST LAW, *supra* note 8, at 113. That this seems obvious to Posner is an excellent example
of the blindspot in antitrust with respect to property-based exclusion.

Posner writes: "Ask yourself how it is that an industry becomes, and remains, highly
concentrated, notwithstanding that it is presumed to be charging supracompetitive prices . . . ." *Id.*

able either because of economies of scale, transaction costs, or breakup costs, then refusal to deal does nothing more than the conduct requirement. It exempts firms that cannot be broken up anyway.

There is no empirical research to support this view.[133] And as an

---

at 113–14. According to Posner, there are two real possibilities: exclusionary practices and a patent "or some other governmental grant of monopoly power that prevents new competition." *Id.* at 114. Posner argues that if exclusionary practices are to blame, then conduct remedies aimed at ending the practice are enough to eliminate monopoly; breakup is not required. *Id.* If the cause is a government grant of power, then antitrust must defer to the government grant. *Id.*

The problem with this argument is that Posner does not understand that there is such a thing as property-based exclusion, unless he means to put all property rights in the category of "some other governmental grant of monopoly power." Antitrust does not prohibit property-based exclusion, so there is no reason to think that single firm monopolies based on it cannot persist indefinitely. And even if antitrust did already address such exclusion, there is no guarantee that a conduct remedy might be available to eliminate it. Breakup might be the only or best option. *See supra* text accompanying note 86; *infra* text accompanying note 149. When one asks oneself how an industry becomes and remains highly concentrated, one thinks immediately of the use of direct ownership over essential inputs to exclude. The natural remedy for property-based exclusion is breakup; if one assumes away the existence of property-based exclusion, then one will naturally see no reason to use it.

In his defense, Posner does recognize that "[c]ontrol of an essential input could be used to block entry," but instead of recognizing the fundamental nature of input control, he dismisses it as "not an important case." POSNER, ANTITRUST LAW, *supra* note 8, at 115. He then argues that input control "can be dealt with as an abuse of the input monopoly." *Id.* This is an attempt to kick the problem of monopoly power up another level in the supply chain.

But it solves nothing. The input monopoly itself could be due to control of an essential input to the input. There is no reason to think that this could not always be the case, however far up the chain you might go. Indeed, as I argue in note 80, that is the only way to account for the existence of monopoly. And Posner implicitly relies on it when he admits that patents could be a source of exclusionary power. Patents are input control. What lets you get them at a lower price than everyone else are brains, or luck, or some other talent that in turn you must have been able to acquire at lower cost than everyone else due to some other factor that, in turn, you must have been able to acquire at lower cost than everyone else. Otherwise the patent brings you no profit. Ultimately, you cannot escape the problem of monopoly without redistributing inputs at some point in the chain.

133. Demsetz argued that any relationship between concentration and profits could be explained by cost reductions, rather than "entry barriers." Demsetz, *supra* note 75, at 3. Subsequent work suggests that there is a relationship between concentration and cost reduction. Salinger, *supra* note 27, at 310. In order for this to support the notion that concentration is associated with natural monopoly, however, there must be evidence either that the source of the cost reductions tends to be unshareable or that the minimum efficient scale of firms with these cost reductions is large. The scale literature has tended to suggest that minimum efficient scale is considerably smaller than the market in most industries. F.M. Scherer, *Economies of Scale and Industrial Concentration, in* INDUSTRIAL CONCENTRATION: THE NEW LEARNING 16, 51–52 (Harvey J. Goldschmid et al., 1974) (Scale economies "are exhausted at relatively modest concentration levels in many manufacturing industries . . . ."); Peltzman, *supra* note 74, at 231 ("A common finding . . . is that of long-run constant costs at the firm level over a wide range of output, wide enough to encompass many existing-firm sizes and a large fraction of industry output."). Unshareability has not been studied empirically and remains pure hypothesis. For a brief discussion of unshareability, see *supra* note 74.

(Peltzman uses the term "natural" for market structures in a nonstandard way. *See* Peltzman, *supra* note 74 at 231. He would apply it to a firm that achieves a cost reduction through innovation and is thereby able to generate economic rents so long as knowledge about how to replicate the

intuitive matter it seems to be false. When antitrust thinks about breakup, it tends to envision the disentangling of tightly integrated operations;[134] the court or its agent must figure out which managers go where, which computers go where, who gets which office furniture, and so on. But this is not necessarily the most common breakup paradigm. It seems reasonable to assume that the source of advantage for many firms is limited to a few key inputs.[135] Perhaps it is control over a particular raw material or ownership of a patent. In such cases, breakup only requires bureaucratic redistribution of the key input, not all the inputs in the firm.[136] Simply dividing ownership of the raw material among independent dealers, or forcing the firm to waive its patent rights,[137] would be enough to allow competitors to access the input and enter the market. The other inputs required for operations (managers, computers, office furniture) are sold on competitive markets and can be assembled separately by entering firms. No bureaucratic disentangling of these non-essential assets would be required for breakup; the market would handle that.[138]

---

innovation "diffuses slowly enough" to other firms. *See id.* This is natural monopoly in the sense that it may arise naturally. But it is not natural monopoly in the sense that the economic rents are a necessary condition of the lower costs. The fact that the ability to replicate the invention is "diffuse[able]" suggests that competition and zero economic rents are possible in the market.)

134. *See supra* note 128.

135. The limiting case in which breakup involves dividing all inputs is almost impossible to imagine. Even the most tightly integrated operation must source some inputs from competitive markets. It may buy non-customized pens and pencils, or computers, or put standardized glass in its windows, or use water from the local water utility, or employ nonskilled workers to clean its floors.

136. Of course, simply identifying the input that is essential may be costly. McGee, *Efficiency and Economies of Size*, *supra* note 74, at 63 ("[T]here are substantial costs of discovering cases of superiority, [and] discerning and evaluating their cause . . . ."). But this is done at the liability stage of a lawsuit; if a plaintiff cannot point to the inputs control over which confers monopoly power, then she cannot make a case for the existence of exclusionary conduct.

137. The rewards-for-innovation problem associated with waiver is discussed later in this Part. *See infra* text following paragraph containing note 139.

138. The courts seem largely oblivious to the relationship between the source of monopoly power and the cost of breakup. For example, the appeals court in *United States v. Microsoft* suggested that, as a rule, only firms formed through mergers and acquisitions, as opposed to organic internal growth, can feasibly be broken up. The reason it gave was that "a corporation, designed to operate effectively as a single entity, cannot readily be dismembered of parts of its various operations without a marked loss of efficiency." United States v. Microsoft Corp., 253 F.3d 34, 106 (D.C. Cir. 2001) (quoting United States v. Aluminum Co. of Am., 91 F. Supp. 333, 416 (S.D.N.Y. 1950)). As an example, it turned to *United States v. United Shoe*: "United conducts all machine manufacture at one plant in Beverly, with one set of jigs and tools, one foundry, one laboratory for machinery problems, one managerial staff, and one labor force. It takes no Solomon to see that this organism cannot be cut into three equal and viable parts." *Id.* (quoting United States v. United Shoe Mach. Corp., 110 F. Supp. 295, 348 (D. Mass. 1953)).

The problem with this approach is that it assumes that the source of a firm's monopoly power is all its inputs regarded as a whole. This may sometimes be the case, but not necessarily most or all of the time. An important contention in *United Shoe* was that its monopoly power was due to

This does not mean that there would be no costs associated with assembly of nonessential inputs by entering firms. There would be setup costs. It might also be the case that, in response to the more competitive environment, the incumbent itself would be forced to shed many of these inputs from its own operations, sending them back into input markets, and those inputs might even end up being acquired by competitors. This process of market-based reallocation is costly.

But this is the same cost involved in any naturally occurring entry of competitors into a formerly monopolized market. To the extent that the argument is that such market-based entry is as a rule prohibitively costly,[139] the argument proves too much. It would tend to undermine all remedies for monopolization, which is to say conduct remedies for recognized forms of exclusionary conduct such as tying, and not just breakup. In fact, it would go further. It would tend to justify protecting any monopoly from entry, since it amounts to saying that entry is always inefficient.

The third concern (relating to compensation to the incumbent for redistributed inputs), is real, but not unique to the problem of remedying property-based exclusion. It arises whenever antitrust imposes a remedy. The reason is that the level of compensation to a wrongdoing firm for redistributed inputs determines the penalty to the wrongdoer for her wrongdoing. If the amount entering competitors are required to compensate a wrongdoing firm for inputs redistributed from it to them is high

---

its control over access to buyers resulting from its long-term machinery leases. *See United Shoe,* 110 F. Supp. at 323–25 (D. Mass. 1953). Imagine that instead of controlling buyers through lease contracts, it owned them directly, thereby allowing it to avail itself of the exemption for refusals to deal. If we were to do away with the exemption and impose a remedy, all that would be required would be to expropriate the buyers and redistribute ownership of them to others. There would be no need to reallocate control over the plant in Beverly, the foundry, or the management staff. If the only source of monopoly power were truly United Shoe's control over buyers, these other inputs could be acquired independently by entering competitors.

Similarly, in *Microsoft,* breakup required only that the court redistribute to others the input that was the source of Microsoft's monopoly power. That input was the source code for its operating systems. Competitors did not need access to Microsoft's programmers and managers in order to compete (these inputs are abundant); with access to the code, they could have assembled their own teams and rushed into the market. Instead, the district court ordered that Microsoft be split into an operating systems business and an applications business, which the appeals court took to mean that its unitary marketing, research, operations, and other teams would have to be split apart and reallocated between the two new companies. *Microsoft Corp.,* 253 F. 3d at 99–100, 106. Not only would that have involved unnecessary difficulty, but leaving the new operating system company in full control of its source code would have done nothing to solve the problem of input control that gave rise to monopolization in the first place. "Breakup" does not have to mean slashing through all the operations of a firm, however much the courts seem to want to make it so.

139. Costly in the sense of failing to maximize social value, not costly in the sense of too expensive for firms to be willing to enter. The situation in which firms are unwilling to enter as part of breakup is effectively the natural monopoly case described earlier in this Part. *See supra* text accompanying note 130.

enough, then the competitors will be forced to charge a monopoly price on entry and pay any monopoly profits back to the wrongdoing firm in the form of compensation.[140] In this case, the wrongdoer experiences no punishment at all. As compensation falls, entering competitors will be able to compete prices lower, and the wrongdoer's profit falls. The problem of finding the right level of compensation is therefore the problem of determining the optimal level of penalty to impose on a violator of the antitrust laws. But the problem of imposing the optimal penalty is not unique to property-based exclusion. Antitrust faces this problem whenever it imposes a remedy for any sort of antitrust violation.

The proper remedy ensures that the wrongdoer loses any extra profit it would have earned relative to doing the right thing by taking power to the efficient level. This ensures that doing the wrong thing is not profitable. The proper remedy also deprives the wrongdoer of an amount equal to the loss in social value associated with its actions in taking power to the wrong level. This ensures that wrongdoers will be willing to spend up to but not beyond the actual social cost of wrongdoing on precautions to avoid wrongdoing.[141] Taken together, these two parts of the proper remedy ensure that a wrongdoer's profit is always less than the profit it would have earned had it done the right thing by an amount equal to the loss in social value associated with her doing of the wrong thing.[142,143]

Suppose a firm or group of firms uses prohibited conduct (e.g., tying or cartelization) to take the market to the one depicted in the bottom middle graph in Diagram 2. We know from Part III.A that the middle middle graph is efficient. The proper remedy will do two things. First, it will impose on the wrongdoer a loss equal to the difference in the size of the profit box in the bottom middle graph and that in the

---

140. *Cf. infra* note 176.

141. *Cf.* STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 236–37 (2004) ("If damages tend to fall short of harm, so that expected payments are below expected harm, incentives to reduce risk will be inadequate, and if damages exceed harm, so that expected damages exceed expected harm, incentives to reduce risk will be too high.").

142. Antitrust typically refers to the part of the punishment designed to eliminate excess profits relative to profit at the optimal level of power as the "overcharge." It calls the part of the punishment designed to internalize the social value loss the "deadweight loss." *Cf.* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, 712–18 (arguing that the optimal measure of damages is "the amount of the overcharge plus the deadweight loss").

143. This assumes that the social value loss is less than profit at the optimal level of power. It may not be efficient to impose on firms any social value loss greater than profit at the optimal level of power. For example, if the loss is greater and precautions approaching the value of the loss are required to avoid it, then the cost of precautions might wipe out any profit a firm might earn at the optimal level of power. This would discourage firms from entering the market at all; no market means zero social value, which is worse than having constricted social value and excessive monopoly power. So there is a ceiling on the amount of the social value loss that optimal punishment strategies will want to impose on wrongdoers.

middle middle graph. This ensures that the wrongdoer makes no more profit than it would have made doing the right thing by taking the market to the middle middle graph to begin with. Second, the proper remedy will also impose on the wrongdoer an additional loss equal to the difference in the size of the social value shapes in the two graphs. This ensures that the wrongdoer internalizes the loss to social value caused by her wrongdoing.

In practice, antitrust never strives to impose the optimal level of punishment regardless what remedy or underlying antitrust violation is at issue. When antitrust dissolves cartels or exclusive dealing arrangements, or when it reverses prohibited transformations such as tying, it never seeks to target a particular post-remedy profit level for the wrongdoer. But without such targeting, it is impossible for antitrust to ensure that post-remedy profits are below profits at the optimal level by an amount equal to the loss in social value.[144]

For example, when antitrust dissolves a cartel, it dissolves the entire cartel, without paying attention to ensuring that industry price falls to a level that will bring about the optimal level of penalty. If price falls too far, then the penalty is too high and firms may spend too much on precautions to avoid running afoul of the law. If price does not fall far enough, perhaps because a history of cartelization facilitates subsequent oligopolization,[145] then dissolution of the cartel will not be a sufficient deterrent. Similarly, in the case of tying, antitrust simply dissolves the tie without paying attention to the level of profit generated by the tying firm in the untied markets.[146]

---

144. Antitrust imposes two sorts of remedies: money damages and injunctive relief. Antitrust does a better job at approximating optimal punishment through money damages than it does through injunctions. Money damages are often related at least loosely to the extra profits earned by wrongdoers relative to some baseline (usually pre-violation profits) that at least in some cases may be the level of profit at the optimal level of power. *Cf.* GAVIL ET AL., *supra* note 36, at 1132–33 ("The structure of the efficient sanction bears a family resemblance to the private damages remedy in force today.").

145. *Cf.* WHINSTON, *supra* note 17, at 32 ("[I]t may be that firms who [sic] have been engaged in price fixing are able to maintain high prices for a period of time even after they are no longer talking.").

146. In theory, antitrust should not be able to impose optimal penalties for any violation of a rule of per se illegality. Doing so would undermine the per se rule, because it would require that antitrust impose no penalty in cases in which there was in fact no monopoly inefficiency. But the point of a per se rule is that it is cheaper to treat all cases falling under the rule alike than to inquire into efficiency on a case-by-case basis. (Strangely enough, however, antitrust does impose something like an optimal penalty in per se cases such as cartelization or tying when it imposes money damages, which tend to be related to excess profits. *See supra* note 144. Thus it is possible to be guilty of price fixing but to pay no damages.)

The rationale for indiscriminate treatment does not carry over to violations established based on case-by-case review for efficiency (e.g., exclusive dealings). When there is case-by-case review, the liability portion of the case establishes the existence of monopoly inefficiency. It

Antitrust's worry about having to determine compensation for redistributed inputs in property-based exclusion cases is that doing so "requires antitrust courts to act as central planners, identifying the proper price, quantity, and other terms of dealing—a role for which they are ill-suited."[147] But this is precisely what any determination and imposition of optimal punishments requires. For example, the optimal punishment for a cartel might be achieved only at a price substantially above marginal cost. To drive price to that level, antitrust may need to create an oligopoly market, rather than a competitive market, by breaking the cartel into two or three competing cartels, rather than dissolving it entirely.

The process of deciding how many cartels to create, which firms should belong to which, and so on, is likely to be of the same administrative complexity as the problem of deciding how much of an input to redistribute to competitors and on what terms. In both cases the problem is how to restructure the market to ensure that profits are reduced by the optimal level of punishment. Similarly, the optimal punishment for a tie might require not the dissolution of the tie but an amendment of its terms to transform the market into one for which profits will be appropriately reduced. This would, again, involve the same type of planning that courts seem to fear in the context of remedies for property-based exclusion.

Furthermore, it seems reasonable to assume that in at least some cases it might be impossible to hit the right price target without resorting to redistribution of input ownership, even where non-property-based exclusions are involved. For example, the number of cartel members, their costs and their assets might be such that no reallotment of firms into new cartels would be likely to result in competition that drives price to the right level. A finer splitting of input control might be required,

---

seems natural that the court would go on to impose a remedy tailored to eliminate that inefficiency. Antitrust does no such thing, however. It takes the same indiscriminate approach to remedying violations established through case-by-case review for efficiency that it uses to remedy violations established through per se rules (outside the money damages context).

   Thus the net effect of case-by-case liability rules is only to save cases of no monopoly inefficiency from the indiscriminate remedial axe. Antitrust imposes the same blanket remedy once the existence of inefficiency, no matter how small, is established. If you are guilty of inefficient exclusive dealing, your contracts are dissolved, regardless whether the punishment inflicted thereby bears any relationship to the extent of the inefficiency created by the exclusive dealing.

   147. Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 408 (2004).

and only redistribution of ownership could achieve that.[148,149] Similarly, no rewriting of the terms of a tie might result in the right level of profits for the tying firm; redistribution of inputs to competitors, either in the tied market or the untied markets, might be required.

It is an expression of the bias in favor of property-based exclusion that although antitrust does not in fact ever engage in planning to achieve the optimal punishment for any antitrust violation, it insists that such planning would be required if antitrust were to impose a remedy for property-based exclusion. Antitrust sees the compensation problem as making redistribution of input ownership relatively difficult only because it insists that redistribution of input ownership alone get punishment right.

Consistency requires that antitrust strive for optimal punishment in imposing any remedy, in which case any remedy from cartel to tie dissolution must resemble central planning, and there is no case to be made that the problem of compensating owners for redistributed inputs justifies a special exemption for property-based exclusion from the antitrust laws. Or consistency requires that antitrust follow its current practice for all other types of remedies and pay no attention to the optimal level of punishment in deciding how much compensation to require for redistribution of input ownership. Redistribute assets without providing any compensation at all. With optimal punishment out of the picture, that rule does as well as any, and certainly would be as cheap to administer as any dissolution of a cartel, exclusive dealing arrangement or tie imposed by antitrust today. So, again, the compensation problem presents no special costs.

B. *An Example of Unclear Thinking About Breakup*

The counterpart of breakup for property-based exclusion is agreement dissolution for contract-based exclusion. Both involve redistribution of inputs from a wrongdoer to competitors.[150] Because antitrust dissolves cartel agreements all the time but rarely breaks up monopolies,[151] it is at pains to explain why breaking up cartels is easy and

---

148. This may explain why antitrust sometimes breaks up large firms even though property-based exclusion is exempt from the antitrust laws. *See infra* the two full paragraphs immediately following the text containing note 215.

149. *Contra* POSNER, ANTITRUST LAW, *supra* note 8, at 114 (stating that deconcentration is "unlikely to be necessary or suitable" when exclusionary conduct may be "enjoined or punished").

150. Recall that the assets of a cartel are the assets of its members, and when a cartel is dissolved, its members become competitors. *See supra* Parts II.D and IV.C. So we can think about cartel dissolution as amounting to the redistribution of assets from the cartel to the newly independent and competitive members.

151. *See supra* note 15.

breaking up monopolies is hard. But the arguments it fields tend to amount to no more than elaborate appeals to the same prejudice against the breakup of monopolies that they are supposed to explain.

The following selections from the writing of Herbert Hovenkamp are typical. In striving to explain why collusion receives more scrutiny than monopolization, Hovenkamp gives the following examples:

> American Express is a single firm that issues a "unitary" general purpose credit card. By contrast, Visa, Inc., which issues a competing general purpose credit card, is a joint venture of some 6000 banks and other financial institutions. Suppose that in different courts these two organizations are accused of two different things and found guilty: (1) fixing supracompetitive consumer interest rates and (2) refusing to share their card processing facilities with a rival card issuer.[152]

Hovenkamp argues that it would be easier for the court to remedy the interest rate claim in the Visa case because to drive the rate down all it would have to do is issue "an order prohibiting the Visa members from fixing an interest rate. After that, competition among the 6000 Visa issuing banks would probably produce the appropriate rate."[153] By contrast, to remedy the interest rate claim in the AmEx case, the court would be forced to "order AmEx to set a 'competitive' interest rate," which would require the court to "develop some criteria by which the competitive interest rate is determined," which in turn would amount to issuing an order that "effectively regulates its prices."[154]

In a similar vein, Hovenkamp argues that it would be easier for the court to remedy the refusal to deal claim in the Visa case because:

> the refusal-to-deal claim in the Visa case arises from a joint venture rule forbidding member banks from issuing a competing card. A judicial decree need not "force" sharing at all; it need only enjoin enforcement of the rule. After that, competition among the 6000 banks, who are now free to issue a competitor's card if they wish, will determine whether and how many rivals' cards will be distributed through banks.[155]

By contrast, argues Hovenkamp, the remedy for refusal to deal in the AmEx case would be for a court to:

> order AmEx to share its card processing facilities with a rival. Whatever one thinks of the wisdom of such relief, granting it requires a significant regulatory effort. The price of sharing and all of the numerous terms will have to be set, just as they would be for a regu-

---

152. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 215.
153. *Id.*
154. *Id.*
155. *Id.* at 215–16 (citations omitted).

lated utility required to interconnect with a rival.[156]

Hovenkamp's interest rate argument fails because he does not pay attention to input control. The source of Visa's ability to charge a high interest rate must be that its members control potential inputs in the industry that are necessary for entry, thereby preventing competitors from entering to drive down the rate. The source of AmEx's ability to charge a high interest rate must also be that AmEx controls potential inputs that are necessary for entry. When a court orders Visa to stop fixing interest rates, it is not ordering Visa to charge a competitive price. It is ordering the dissolution of the Visa cartel, which, it is hoped, will result in interest rates coming down. That amounts to ordering the division of the inputs Visa controls among competitors. The analogue of this order in the AmEx case is an order to break up the inputs AmEx owns and divide them among competitors.

Hovenkamp misses this connection. Instead, he assumes that the analogue of the order against Visa's interest rate fixing is an order to AmEx to set a competitive price. This mistake allows him to leverage antitrust's traditional taboo against rate regulation to argue that antitrust can do nothing to remedy the price fixing violation in the single firm context. This completely misses the core question, however, which is why breaking up AmEx would be more expensive than breaking up Visa. On this point Hovenkamp is silent.[157]

Hovenkamp's refusal to deal argument runs into similar problems. It fails because although he wants this argument to explain why cartelization is easier to remedy than monopolization, what he ends up showing is only that entry is easier to impose in cases of unnatural monopoly than in cases of natural monopoly. Hovenkamp has nothing to say about costs associated with forcing Visa to allow member banks to give competing cards access to card processing facilities. He simply stipulates that "significant . . . effort" is required for AmEx to do the same with its own facility.[158]

The reasons for assuming such a cost differential are not obvious. Perhaps the idea is that because AmEx is a single firm, it has had the opportunity to realize economies of scale in its processing operations that are not available to the smaller individual members of Visa. But if that is the case, and this makes it inefficient to break up AmEx, then

---

156. *Id.* at 215.

157. If what Hovenkamp was really interested in was the appropriateness of rate regulation orders, then he should have asked whether it would be more appropriate to order Visa or AmEx to charge a competitive price. Presumably, the taboo against rate regulation applies with equal force whether the court addresses its order to the pricing department of a single firm or the cartel organ that determines cartel prices. So this does not justify treating cartels and monopolies differently.

158. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 215.

AmEx is a natural monopoly, whereas the Visa cartel is not. This tells us something about the difficulty of breaking up natural monopolies, but nothing about why cartels should be harder to break up than single firms. To really get at the question why cartels deserve more scrutiny than monopolies, what Hovenkamp would have to do is either (1) establish that the incidence of natural monopoly is higher in the case of single firms as opposed to cartels or (2) identify some other reason for which single firms are harder to break up than cartels, even when both are not natural monopolies. Hovenkamp is silent on these issues.

C. *Imagining a World Without the Refusal to Deal Exemption*

The sanctity of property-based exclusion is so ingrained in our intuition that it is difficult to imagine what it would really mean to do away with it. Here is a particularly counterintuitive case. A firm invests in a new plant. The plant has the capacity to satisfy twice the demand at the competitive price. Because of the nature of the production process, only plants of this size may be built. Because the construction of the plant required the use of the last remaining deposits of an essential mineral, and there are no substitutes, no other firms can construct a plant. The plant itself is, therefore, an essential input to the production of the product. Thus protected, the incumbent uses only a third of the factory, produces a third of capacity, and charges the monopoly price. Two thirds of the factory remain idle.

We have seen that there are two ways in which we may think of this idleness as inefficient. If we assume no connection between fixed cost and product quality, then we can think of this situation as inefficient in the way that defenders of antitrust traditionally think about inefficiency. If price were at average cost, meaning just high enough to cover cost, including the fixed cost associated with making the plant, more would be produced and enjoyed without any loss in product quality. So to maximize social value, it would be necessary to put some of the idle plant to work, thereby driving up output and driving price down to average cost.

If we assume a connection between fixed cost and product quality, then we can think of this situation as inefficient if the socially optimal amount of investment in fixed costs is lower than the amount actually invested. If the firm had expected not to be able to monopolize the market, it would have built a cheaper plant that produced somewhat lower quality products. In that more efficient world, the firm would have been compelled to keep less of its cheaper plant idle in the sense that it would have been compelled to produce more than the monopoly level of output that it now produces in the more expensive plant.

The right to property-based exclusion is so ingrained that competitors or enforcers would probably not even think to argue that the firm is engaging in exclusionary conduct by keeping competitors out of the idle part of the factory floor. But the power of the incumbent to exclude from the idle part of the plant is precisely the source of the inefficiency. It prevents competitors from using a potential input that is essential to entering the market and driving down the price.

Assume that there are no economies associated with unitary control over the entire plant.[159] Perhaps the production process is fully automated and the only thing management does is to turn a dial that indicates the level of output. Both types of inefficiency would be avoided by giving competitors the right to come onto the firm's property and harness some portion of the idle part of the plant to expand output. If the level of access were optimally drawn, enough people wishing to compete would be allowed to walk into the control room, turn up the output dial, and sell the extra output as their own, that price would fall to the efficient level. In the first type of inefficiency, this would be average cost. In the second type, this would be the level required to change the profit expectations of the firm and thereby compel it ex ante to make the optimal level of investment in product quality.

D. *Antitrust and Theft*

Letting outsiders come onto a factory floor and take control of the equipment sounds a lot like theft. The distinction between breakup and theft is important, not least because it goes to the heart of the justification for antitrust. We can think of law, including property law, as part of a system of input control. Antitrust corrects a perversity in this system according to which, contrary to the normal effect and intent of input control, it tends to reduce production instead of to increase it.

The point of allowing people to control inputs is to allow producers to capture enough of the benefits of production to make it worthwhile to them to engage in production. Without this, production does not take place and society is worse off. A blind respect for input control leads to trouble, however, because producers and consumers have no sure way of reaching agreement about how to divide the surplus[160] from production. Uniform pricing is a common solution, but it has the bad consequence of sometimes allowing producers to maximize the surplus they get by reducing the amount that they produce.

When production falls, society is worse off, and it must intervene to

---

159. That is, assume that we are dealing with a genuine refusal to deal, and not a natural monopoly protected by the conduct requirement. *See supra* Part V.A.

160. Value above cost, which makes production worthwhile.

drive production back up. Because producers control production, this
necessarily involves curtailment of producer control. But the curtailment
here is not theft, or it is efficient theft anyway, because it does not tend
to reduce production; on the contrary, its effect is to increase
production.[161]

E. *Essential Facilities Cannot Be Reconciled with Refusal to Deal Because It Is Sworn to Destroy It*

The only doctrine in antitrust law that squarely attacks the exemp-
tion for property-based exclusion is the essential facilities doctrine. It
creates liability if four requirements are met: (1) the defendant controls
the essential facility, (2) competitors are unable practically or reasonably
to duplicate it, (3) the defendant has denied use of the facility to compet-
itors, and (4) it is feasible to allow use by competitors.[162] If you take
"facility" to refer to property,[163] then the doctrine amounts to a general
rule against property-based exclusion.[164] It swallows the refusal to deal
doctrine.

The refusal to deal doctrine authorizes firms to use property rights
to exclude; the essential facilities doctrine prohibits them from using
property rights to exclude. Scholars are troubled by the conflict[165] and
have tried to reconcile refusal to deal and essential facilities, usually by

---

161. This Part contrasts antitrust regulation of input control, rather than property, with theft,
because what is troubling about allowing people to come onto your factory floor and use the
machinery is that it implies a loss of the power to demand compensation. The source of that
power, whether in the willingness of the police to enforce your property rights or your own private
army, is irrelevant. *See supra* Part IV.G.

162. *E.g.*, MCI Commc'ns Corp. v. Am. Tel. & Tel. Co., 708 F.2d 1081, 1132–33 (7th Cir.
1983); *see generally* Robert Pitofsky et al., *The Essential Facilities Doctrine Under U.S. Antitrust
Law*, 70 ANTITRUST L.J. 443 (2002).

163. The doctrine is invariably invoked when the refusal to deal exemption would otherwise
apply to bar an antitrust claim, which is to say, when exclusion is property-based. *See* Pitofsky et
al., *supra* note 162, at 448 ("[T]he doctrine represents a divergence from the general rule that even
a monopolist may choose with whom to deal . . . ."). Some commentators have suggested that
"facility" should refer not just to any kind of property but to particular types. *See* Abbott B.
Lipsky, Jr. & J. Gregory Sidak, *Essential Facilities*, 51 STAN. L. REV. 1187, 1220 (1999) (arguing
that "cases in which the 'facility' is not a single, indivisible unit, but a collection of potentially
independent and viable competitive units," as well as IP, should not be covered by the doctrine).
Although the courts have tended to recognize essential facilities only when the facility is
associated with a regulatory regime, they have not read a regulatory requirement into the test. *See*
HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 337. And the assortment of property
types that plaintiffs try to bring under the test is eclectic. *See* Lipsky, Jr. & Sidak, *supra*, at
1190–93 (ranging from soft drinks to tubing for exercise equipment).

164. *Cf.* Areeda, *supra* note 121, at 852 n.46 ("The trouble with . . . the essential facilities
notion is that [it] start[s] with the assumption that all business assets are subject to sharing. Do we
really want to assume that everything we have is up for grabs?").

165. *See, e.g., id.*

casting essential facilities as an exception to refusal to deal.[166] This can of course be achieved by arbitrarily varying the subject matter to which essential facilities may be applied.[167] But any attempt to do justice to both rules is doomed to failure. Refusal to deal is the rule that property is exempt from antitrust, and essential facilities is the rule that antitrust covers property.

VI. No Exemption for Refusals to Deal Means No More Per Se Rule in Favor of IP-Based Exclusion

IP is just another form of input control. It shares a common economic justification with traditional property and other forms of input control, namely, that it allows producers to control inputs and outputs and thereby to cover the costs of production.[168] Although it is fashionable to emphasize the harmony of IP and antitrust,[169] there is inherent conflict between IP and antitrust to the extent that there is inherent conflict between all forms of input control and antitrust. The point of antitrust is to limit input control when it is efficient to do so;[170] this means that when IP confers too much control, antitrust will want to limit IP.

Because antitrust has traditionally exempted property-based exclusion from scrutiny and antitrust treats IP as property,[171] antitrust rarely intervenes to prevent exclusion based on IP rights alone.[172] As a result, the IP/antitrust conflict has rarely had occasion to manifest itself in the context of property-based exclusion. But antitrust contains no blanket

---

166. *See, e.g.,* Pitofsky et al., *supra* note 162, at 451–52.

167. *See, e.g.,* Spencer Weber Waller, *Areeda, Epithets, and Essential Facilities,* 2008 Wis. L. Rev. 359, 371–85 (arguing that the doctrine should apply only to inputs classified as "infrastructure," which is taken to mean inputs "into a wide range of productive processes, often supporting many uses, applications and downstream markets"); Brett Frischmann & Spencer Weber Waller, *Revitalizing Essential Facilities,* 75 Antitrust L.J. 1, 4, 19–22 (2008) (same). A problem with the infrastructure approach in these sources is that it fails to explain why a blanket exemption from the antitrust laws should persist for refusals to deal based on ownership of specialized inputs that are of use only in one or a few markets (i.e., those that do not count as infrastructural). There is no reason to think that refusals to deal in such non-infrastructural inputs are any less likely than refusals to deal in infrastructure to result in monopoly inefficiency. And there is no reason to think that breaking up control over such inputs would be any less cost effective.

168. *Compare* Posner, Economic Analysis of Law, *supra* note 118, at 32 (without property in land, people can take your crop without paying for it) *with* Carlton & Perloff, *supra* note 10, at 532 (without patents, people can copy your invention without paying for it).

169. *See, e.g.,* Mark A. Lemley, *A New Balance Between IP and Antitrust,* 13 Sw. J.L. & Trade Am. 237, 245–49 (2007).

170. *See supra* Parts IV.G and V.D.

171. Despite the fact that IP has the word "property" in its name, it is not inevitable that antitrust should accord it the same respect it accords traditional property. The lines of "property" for antitrust purposes are arbitrarily and emotionally drawn. *See supra* Part II.C.

172. *See supra* note 40.

exemption for exclusions based only indirectly on property rights, such as those with a contractual element, or on certain types of market trans-formations.[173] As a result, the IP/antitrust conflict has played out largely in the context of transformations[174] and indirect exclusions.[175]

The argument against the exemption for property-based exclusion applies with equal force to IP as to traditional property. And its conse-quences are the same. Eliminating the blanket exemption for refusal to deal means that firms would only be permitted to exercise IP rights to exclude competitors from their innovations in cases in which doing so would actually be efficient (i.e., only when the monopoly power created thereby is no greater than necessary to induce the optimal level of inno-vation). In terms of Diagram 2, this means that antitrust could require an IP holder to allow enough competitors to use her invention to drive monopoly power down to the middle row. This implies that, in some cases, IP holders would have to license IP for less than the full monop-oly profit available in the market.[176] And an IP holder who refused to do so and proceeded to violate the antitrust laws could have her IP expro-priated without compensation and redistributed to competitors.[177]

Two sets of defenses of property-based exclusion tend to arise in the patent context. One is that, either because innovation leads to prod-uct improvement and growth[178] or because information is harder to con-

---

173. *See supra* Part IV.B.

174. *See, e.g.,* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 443–44 (discussing tying in relation to products protected by IP rights).

175. *See generally id.* at 265–72 (antitrust claims arising out of patent licensing agreements). A prominent recent example is pay-for-delay, which involves agreements pursuant to which patent holders pay competitors not to continue patent challenges. *See* FTC v. Actavis, Inc., 133 S. Ct. 2223 (2013).

176. If a firm licenses IP in exchange for the monopoly profit it could have earned by not licensing at all, then the IP has not in fact been shared in any meaningful sense. The licensees will have to charge monopoly prices in order to pay the license fee, which means that the fee will force them to behave like a monopolist. If social value is not maximized at monopoly prices, then to comply with antitrust laws, the IP owner will have to charge less than the monopoly profit, thereby allowing licensees to compete prices down. *See, e.g.,* Carl Shapiro, *Patent Licensing and R&D Rivalry*, 75 AM. ECON. REV. 25, 26 (1985).

177. *See supra* Part V.A (arguing that redistribution of input ownership should not be held to a higher standard than other antitrust remedies, none of which take the optimal level of compensation/punishment into account). Part V.A also argues that the cost of breakup can be low if a firm's monopoly power is due to control over some identifiable subset of its inputs. It is reasonable to think that where monopoly power is due to IP, breakup cost will be particularly low because knowledge is often cheap to share. Indeed, it is precisely because knowledge is easy to share that IP protection is thought to be especially important. *See, e.g.,* Lemley, *supra* note 169, at 238–41.

178. *Cf.* SCHERER & ROSS, *supra* note 10, at 682 ("[G]ood economic performance depends much more critically upon sustaining a vigorous pace of technological progress than upon plausible variations in allocative efficiency . . . .").

trol without government help,[179] patent rights deserve more protection than other forms of input control. The trouble with this argument is that it does not explain why antitrust should not intervene to prevent instances of too much protection. Part III showed that there is no reason to think that full protection is always efficient. The whole point of antitrust is that there is such a thing as too much reward, even for innovation.[180]

The other argument is preemption.[181] The antitrust and patent laws are both federal laws and neither contains any provision conferring precedence over the other, so the reasons for this view are not textual.[182] Instead, the prejudice in favor of patent seems to arise from the fact that Congress intends the patent laws to provide a reward for innovative activity and has specified a patent term (twenty years), which serves as a ceiling on the amount of reward obtainable. This seems to suggest that Congress has already made an efficiency calculation regarding rewards and does not need any additional help from antitrust.[183]

The problem with this argument is that it proves too much. The rationale not just for IP but for all forms of input control generally is that input control provides a reward for productive behavior.[184] If antitrust had to make way for any form of input control guaranteed by a legal rule intended to reward producers, then it could never interfere with any sort

---

179. *Cf.* WARD S. BOWMAN, JR., PATENT AND ANTITRUST LAW 17 (1973) (observing that the fact that one person's use of information does not prevent another person's use of the same information "raises marketability problems").

180. *See supra* Part III.

181. One example of this argument is contained in a leading treatise, which states that "it bears recognizing that patents and other intellectual property rights limit the reach of the antitrust laws." HOVENKAMP ET AL., *supra* note 40, § 1.3b (citing BOWMAN, *supra* note 179, at 8–9 ("The legal propriety of a basic patent monopoly has to be recognized.")). Curiously, it cites an economist (Bowman) as authority. William F. Baxter agrees, without providing authority. William F. Baxter, *Legal Restrictions on Exploitation of the Patent Monopoly: An Economic Analysis*, 76 YALE L.J. 267, 272 (1966) (stating that "it seems peculiarly appropriate that the answer" to the question how much reward to provide for innovation "should have been given by" Congress in setting the patent term). Posner also seems to take this view when he distinguishes between "exclusionary practices" actionable under the antitrust laws and "patent or some other governmental grant of monopoly power that prevents new competition." POSNER, ANTITRUST LAW, *supra* note 8, at 114.

182. 35 U.S.C. § 271(d)(4) (2006 & Supp. 2010) does explicitly exempt refusals to license patents from the patent misuse doctrine, but it places no explicit limit on the antitrust laws. *Cf.* HOVENKAMP ET AL., *supra* note 40, § 13.2 n.8 ("While the statute refers to patent misuse, and does not directly cover antitrust violations, the policy it expresses is still relevant."). Hovenkamp reads the exemption to extend to the antitrust laws, but does not explain why. HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 362 ("This provision makes clear that under the Patent Act a simple refusal to license a patent cannot be 'misuse' of a patent, and thus cannot violate the antitrust laws.").

183. *Cf.* Baxter, *supra* note 181, at 272 ("[T]he Congressional judgment of 'how much' was made . . . [i]n part . . . in the monopoly technique through which the subsidy is afforded [and in] . . . the period for which monopoly is conferred.").

184. *See supra* Part IV.G.

of property right (which it sometimes does, though rarely)[185] or any sort
of contract right (which it does frequently,[186] including in the IP con-
text),[187] since the underlying economic rationale for both these regimes
is to reward producers.[188] The existence of IP terms is of no conse-
quence, since the law also imposes plenty of restrictions on traditional
property and contract that limit the extent of the rewards these legal
regimes are capable of conferring.[189]

VII. MORE INCONSISTENCIES

A. *Pro Vertical and Anti Horizontal*

Input control may be achieved vertically or horizontally. Horizontal
control is when the potential inputs in a market are controlled either
directly or via producers in the market. The producers control the poten-
tial inputs, so controlling the producers controls them too. Cartels, oli-
gopolies, and horizontally integrated monopolies are examples of
horizontal control. Vertical control is when the potential inputs in a mar-
ket are controlled via firms ("suppliers") that produce the inputs but do
not themselves use them to produce in the market. Suppliers control the
potential inputs, so controlling the suppliers controls them too. Exclu-
sive dealings and vertically integrated monopolies[190] are examples of
vertical control.

To see how vertical and horizontal are just two ways of achieving
control over the same set of inputs, consider exclusive dealing and carte-

---

185. *See supra* note 43.

186. *See supra* note 15.

187. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 265–72.

188. *Compare* POSNER, ECONOMIC ANALYSIS OF LAW, *supra* note 118, at 32 (if the farmer does
not own her land, others may reap her corn, so the farmer will stop planting it) *with id.* at 93 (in
the absence of contract law, the home buyer may refuse to pay full price to the builder once the
builder has completed the house).

189. These restrictions are not always temporal, but there is no reason to think that a temporal
restriction indicates any greater intent to limit rewards than other types of restrictions. Here are
some examples. In property, the modern rule is that there is no easement for light and air. *See*
JOSEPH WILLIAM SINGER, PROPERTY LAW: RULES, POLICIES AND PRACTICES 347 (3d ed. 2002).
This limits the rewards to building a home on a given individual piece of land. MORTON J.
HORWITZ, THE TRANSFORMATION OF AMERICAN LAW, 1780–1860, at 43–47 (1977) (describing
the nineteenth century decline of the doctrine of prescription, including the doctrine of "ancient
lights," which prevented neighbors from building structures that would block sunlight, because
"the law of prescription had come to be associated with . . . monopolistic and restrictive
practices . . . ."). Consideration and unconscionability limit contract rewards. *See* E. ALLAN
FARNSWORTH, CONTRACTS §§ 4.26–28 (3d ed. 1999) (describing modern doctrine of
unconscionability, which polices contracts for the fairness of rewards); HORWITZ, *supra*, at
177–79 (describing how in the nineteenth century the doctrine of consideration lost its traditional
function of limiting rewards and took on significance purely as evidence of contract formation).

190. Vertically integrated monopolies, like all monopolies, are ownership centralizations. A
vertically integrated monopoly owns its suppliers, which means that it owns its inputs too.

lization. Every exclusive dealing relationship that allows firms to charge P is in effect a hub-and-spoke cartelization, with the downstream firm as the hub and the upstream firms as the spokes.[191] This hub-and-spoke cartelization differs from a conventional cartelization only in that the spokes do not produce in the relevant market. In the hub-and-spoke cartelization, the cartel exerts control over potential inputs by taking control of input suppliers (the spokes), while in conventional cartelization, the cartel takes control of potential inputs by taking control over other firms in the market, which in turn control the inputs. (It is because both exclusive dealing and cartelization are just two ways of controlling input owners indirectly that they both fit in the "indirect influence over the owners of inputs" category.)[192]

Antitrust tends to treat horizontal control very differently from vertical control. Horizontal agreements on price (cartelizations) are per se illegal.[193] Vertical agreements (exclusive dealings) receive case-by-case review.[194] Most merger enforcement is focused on horizontal mergers; vertical mergers are almost never challenged.[195] But there is no research that shows that vertical control tends to cause fewer black triangles than horizontal control.[196]

B.   *Inconsistent Treatment of Natural Monopolization*

When a monopoly can do nothing to allow competitors into the market without raising its own costs, it is a natural monopoly.[197] Antitrust treats the start of natural monopoly very differently from the continuation of it and discriminates heavily between different kinds of starts.

---

191. *See supra* note 35; CARLTON & PERLOFF, *supra* note 10, at 430 (discussing relationship between vertical restrictions and cartelization).

192. *See supra* Part IV.B and Chart 2.

193. *See supra* note 15.

194. *See generally* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 484–88. In some particularly stark cases, however, antitrust has been willing to treat vertical control the way it treats horizontal control. *See* Toys "R" Us, Inc. v. FTC, 221 F.3d 928, 934–37 (7th Cir. 2000) (applying per se rule against cartels to vertical agreements with manufacturer not to sell to competing distributors); *but see* Interstate Circuit, Inc. v. United States, 306 U.S. 208, 221, 230–31 (1939) (treating series of vertical agreements on price as equivalent of horizontal agreement, but not applying the per se rule against cartels).

195. *See supra* note 110.

196. There is a large literature on efficiency justifications for vertical control. *See generally* WILLIAMSON, *supra* note 21. There is also a literature on merger efficiencies. *See generally* WHINSTON, *supra* note 17, at ch. 3. But there is no literature comparing the efficiency of horizontal and vertical control or showing that vertical control tends to be more efficient than horizontal control.

197. The discussion in this part is limited to the case of natural monopolies in subadditive cost markets, meaning those monopolies in subadditive cost markets for which firms outside the market cannot enter in response to price increases. *See supra* note 87.

As we have seen, antitrust has traditionally refused to intervene to drive price down in continuations of natural monopolies on the ground that the incumbent competitor is not really culpable.[198] The incumbent could not allow competitors into the market even if it wanted to, at least not without raising costs in the market. Antitrust readily admits that rate regulation is sometimes an alternative that would drive price down without raising costs, but it has traditionally left it to the legislature to impose that remedy.[199]

The trouble with this line of thought is that, although entry cannot maximize social value for natural monopoly (only rate regulation can do that), it can sometimes increase social value. Entry will increase social value in natural monopoly cases whenever the loss associated with the higher costs of a competitive market is less than the gain associated with the elimination of deadweight loss through competitive pricing. Natural monopolies cannot let competitors in without raising costs, but raising costs is sometimes relatively efficient.
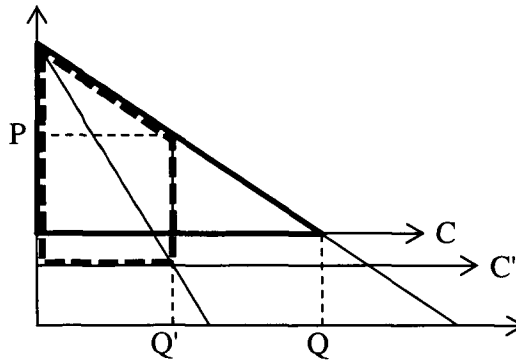
DIAGRAM 3

Diagram 3 illustrates. The market is initially a natural monopoly; price is at P, output is at Q', and costs are minimized at C'. Subsequently, antitrust intervenes to compel entry, and cost rises to C. But competition drives price down to marginal cost, which is to say, to C. The dashed bold shape shows social value before the intervention. The solid bold triangle shows it after. Social value increases in this case because the decline in social value associated with the increase in costs (the rectangle bounded by the y-axis and Q' on the left and right and C and C' on the top and bottom) is less than the part of the deadweight loss of monopoly (the triangle bounded by Q' on the left and the demand line

198. *See supra* Part II.B.
199. *See supra* note 31.

and C on the top and bottom) that has been eliminated by competitive pricing.

Antitrust could use entry to increase social value in monopoly cases, but as we have seen, the conduct requirement exempts all continuing natural monopolies from antitrust scrutiny. Because natural monopolies have done nothing to keep competitors out other than to minimize costs, antitrust refuses to find them culpable, even though it is reasonable to think that, in at least some cases, antitrust could make society better off by intervening.

The curious thing about antitrust's exemption for continuing natural monopolies is that Diagram 3 is already well known to antitrust, which regularly deploys it to prevent the formation of natural monopolies through cost-saving mergers. Antitrust calls Diagram 3 the Williamson diagram, and applies it in reverse to mergers, including mergers to natural monopoly.[200] In this merger scenario, the merger results in cost savings that drive C down to C′, but results in natural monopoly that drives price up to P.[201] The bolded triangle therefore represents pre-merger social value and the dashed shape post-merger social value. Antitrust approves the merger only if social value increases.[202] In Diagram 3, value falls and the merger would be rejected.

What the use of the Williamson diagram in merger practice tells us is that antitrust is willing to intervene to stop an inefficient natural monopolization through merger, but is not willing to intervene to break up a preexisting inefficient natural monopolization. As in the case of refusal to deal, in a world of no changed circumstances the focus on starts as opposed to continuations would be of no consequence if antitrust were to police starts to natural monopoly thoroughly. But, as in the case of refusal to deal, antitrust does no such thing. All the limitations to merger and transformation law discussed in Part IV.E apply to the application of those laws to the start of natural monopolization.[203] And here

---

200. WHINSTON, *supra* note 17, at 59; *see generally* Oliver E. Williamson, *Economies as an Antitrust Defense Revisited*, 125 U. PA. L. REV. 699 (1977).

201. *See* Williamson, *supra* note 200, at 706–09.

202. This characterization of the role of the model in the merger review process has been much simplified for purposes of exposition. Williamson himself was careful to call the model "naïve" and to point out ways in which it can be misleading. *See id.* at 710–13. Nonetheless, it stands for the recognition by merger law that there is a basic trade-off between economies of scale and monopoly inefficiency. *See, e.g.*, WHINSTON, *supra* note 17, at 58 ("The central issue in the evaluation of horizontal mergers lies in the need to balance any reductions in competition against the possibility of productivity improvements arising from a merger.").

203. Acquisitions can lead to natural monopoly in subadditive cost markets with first mover advantages. *See supra* note 87. In this context, the point of acquisitions is not to obtain more inputs with which to exclude. The structure of the market ensures that any firm that makes it to monopoly will be protected. Rather, the point of acquisitions is to get big quicker than the competition, and thereby to realize scale efficiencies that smaller competitors cannot rival. Merger

again antitrust does not regulate "natural" transformations to natural monopoly, meaning those that are not attributable to the decisions of any competitor.[204] (The far right column in Chart 2 on page 384 summarizes the legal treatment of natural monopolies.)

It would be reasonable to assume that antitrust scrutinizes some mergers and transformations to natural monopoly because it believes that these are more likely to be inefficient (in the sense of reducing social value in Diagram 3) than the mergers and transformations it does not scrutinize or than natural market changes that result in natural monopoly. But there is no research to support this difference in treatment.[205]

C. *Hard- and Soft-Core Cartels but No Hard- or Soft-Core Monopolies*

Antitrust entertains a distinction between hard-[206] and soft-core[207] cartels. By its own admission, antitrust draws the line on an ad hoc basis,[208] but there is a trend. Cartels that appear to have contemplated no product improvement before becoming cartels count as hard-core and are condemned per se, but cartels that appear to have contemplated product improvement ex ante receive case-by-case scrutiny.[209] So, for exam-

---

law does not distinguish between the different ends of acquisition in the unnatural and natural monopoly contexts, so all the same exceptions apply.

Transformations also work a bit differently in the natural monopoly context. A firm may transform a market into a natural monopoly market that it believes it can dominate. The remedy for this is either to undo the transformation or to compel entry into the natural monopoly market if that is efficient in the sense of Diagram 3. Transformation law typically does not recognize that the latter is even a remedy for illegal transformation. *See infra* Part VII.D. Even if it did, it might still not apply due to the exemption for superior product transformations.

204. *See supra* text accompanying note 113.

205. It is difficult to imagine how this difference might ever be supported. Suppose a company invents a new process that gives it a natural monopoly. Suppose instead that the company achieves natural monopoly through merger. It is hard to imagine why the efficiency of breakup would be greater in the first method than in the second, when the efficiency of breaking up the natural monopoly depends on the cost structure of the natural monopoly market and not how the natural monopolist happened to come to exploit it. (The superior products exemption to transformation law prevents antitrust from examining the first transition to natural monopoly, but merger law would scrutinize the second.)

206. The sort that fall under the per se rule against cartels.

207. The sort that fall under the "rule of reason."

208. *See* Cal. Dental Ass'n v. FTC, 526 U.S. 756, 779 (1999) ("The truth is that our categories of analysis of anticompetitive effect are less fixed than terms like 'per se,' . . . and 'rule of reason' tend to make them appear.").

209. Hovenkamp suggests that the expectation of monopoly profit cannot play any role at all in a cartel's calculus. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 279 ("[A]n ancillary agreement reduces cost or improves the product and can be profitable whether or not the firms have any market power."). If this interpretation of the law is correct, whether the cartel prepares for the profit ex ante by planning product improvements or merely spends the profit on them ex post is irrelevant. This means that cartel law, both hard- and soft-core, does not

ple, an agreement to prevent doctors from duping patients into unnecessary procedures that would probably also help raise price gets individualized attention,[210] but one that just raises price is per se illegal.[211]

Monopolies are not treated this way. Whether product improvement was contemplated when a monopoly formed is irrelevant to whether it can benefit from the refusal to deal exemption.[212] Indeed, the efficiency justification for refusal to deal has nothing to say about the intent of exclusion. As we have seen, the argument is that a monopoly ought to be perfectly free to raise price to P because the extra quasi-profit generated thereby will create an incentive for it to carry higher fixed costs and thereby make available products that otherwise it would not be viable to produce.[213] Whether the firm intends to spend more on fixed costs when it raises price is irrelevant; the increase in price will create the incentive to spend more whether the monopolist is aware that it will or not.

So greed is bad in the case of cartels, but good in the case of monopolies. And yet there is no reason to think that cartel members might be any less willing than monopolies to invest quasi-profits obtained for the wrong reason in product improvement.[214] A consistent antitrust might distinguish between hard- and soft-core monopolies, ban the former per se, and examine the latter case-by-case, instead of exempting them all from scrutiny as it does now under refusal to deal.

---

countenance any rewards theory at all; for it, there is no connection between product improvement and monopoly profit in the cartel context.

This Part assumes that courts may sometimes permit competitor agreements that contemplate generation of monopoly profits so long as the agreements also ex ante contemplate spending the monopoly profit on product improvement. This seems to explain results in the joint venture context. Courts permit joint ventures to charge membership fees to allow investors to "capture the return" on their investments. *See id.* at 248 & n.53. Such fees put a floor on market price in order to cover R&D fixed costs and are therefore based on a rewards theory.

210. *See Cal. Dental Ass'n v. FTC*, 526 U.S. at 771, 779–81.

211. *Cf.* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 279 ("An agreement is naked if it is formed with the objectively intended purpose or likely effect of increasing price or decreasing output in the short run.").

212. The discussion of cartels and monopolies in this Part is limited to states as opposed to changes. *See supra* Part IV.E and Chart 2. As discussed in Part IV.E, antitrust's treatment of transitions to monopoly is considerably more complicated than its treatment of the state of being a monopoly (i.e., than the blanket exemption created by the refusal to deal doctrine). Antitrust sometimes considers contemplation of product improvement relevant in deciding whether to permit a transition to monopoly.

213. *See supra* Part III. The ex ante story that antitrust tells itself about monopolization does presuppose an intent to spend profits on product improvement. *See supra* Part II.A. But this is just a narrow form of the general efficiency rationale for monopoly discussed in Part III. That general rationale does not require intent.

214. *See supra* notes 21, 22, and accompanying text.

D. *Inconsistent Application of Remedies: Input Control for
     Transformation but No Transformation for Input Control*

We can think of each antitrust remedy aimed at promoting entry as
reversing a particular exclusionary practice. Dissolving a tie is the rem-
edy for tying. Ordering cost pricing is the remedy for predatory pricing.
And breaking up control over inputs is the remedy for input control.
When breakup is applied to input control through cartelization, it is usu-
ally called a dissolution of the cartel agreement. When it is applied to
input control through direct ownership,[215] it retains the name breakup.
Let us call the remedy that reverses a particular form of exclusion the
"natural" remedy for that exclusion.

Antitrust does not adhere to a rule that requires it to apply only the
natural remedy to promote entry in response to a particular prohibited
exclusionary practice. This is clear from antitrust's use of the breakup
remedy. Refusal to deal exempts property-based exclusions from the
antitrust laws, which means that, roughly speaking, it exempts exclu-
sions based on direct ownership of potential inputs.[216] If antitrust were
to adhere to a rule that requires it to apply only the natural remedy to
promote entry in response to a particular prohibited exclusionary prac-
tice, then it would never apply breakup because breakup is the natural
remedy for input control through direct ownership, and input control
through direct ownership is exempt from the antitrust laws under the
refusal to deal doctrine.

But antitrust does occasionally apply breakup to input control based
on direct ownership that has been achieved through a prohibited form of
transformation, such as tying.[217] You are not guilty of monopolization if
you simply happen to own all the hotel rooms in the market. That would
be input control based on direct ownership, which is exempt under
refusal to deal. But you are guilty if you tie the hotel rooms to meals.[218]
Antitrust could respond simply by dissolving the tie, but it sometimes
responds instead by breaking up your hotel into two hotels. In effect,
antitrust sometimes applies the remedy for property-based exclusion to
remedy what it considers non-property-based exclusion.[219]

It should have been perplexing to antitrust, although the issue does
not appear ever to have been raised, that although antitrust sometimes
applies a natural input control remedy to respond to prohibited transfor-
mation, it never seems to go the other way and apply a natural transfor-

---

215. Meaning monopoly. *See supra* note 81.
216. *See supra* Parts II.B, II.C, and IV.A.
217. For a history of breakup of monopolies in antitrust, see Kovacic, *supra* note 74.
218. *See supra* note 36 and accompanying text.
219. *See supra* Part II.C (distinguishing property and non-property based exclusion).

mation remedy to respond to a prohibited form of input control, such as cartelization.

Suppose that there is not one hotel but three, collectively controlling all rooms in the market, and that the hotels independently (i.e., non collusively) all happen to tie rooms to meals, but collusively fix the price of hotel/meal combos. Two things allow the hotels to charge P: the collusion allows them to control the inputs in the hotel/meals market, and the tying allows them to operate in a hotels/meals market to begin with. Getting rid of either will allow competitors to enter and drive down price. Getting rid of the collusion will turn the three firms into price competitors (in effect allowing the firms to "enter" the market as competitors rather than colluders). Getting rid of the ties will turn the market back into a meals market in which local restaurants can compete.

Because the tying is undertaken independently, and each individual firm does not have the power to transform the market through its tie alone, courts are not likely to find the ties exclusionary on their own. But they will find that the price fixing violates the per se rule against cartelization.[220] The natural remedy for the courts to apply for the prohibited cartelization would be dissolution of the cartel, which is an input control remedy. But since the courts are already in the business of mixing remedies, they could also apply the natural remedy for tying (i.e., dissolution of the tie), which is a market transformation remedy, even though tying is not prohibited in this case. It is unheard of, however, for antitrust to employ market transformation as a remedy where the underlying violation is not market transformation.[221]

VIII. THE BALANCING TEST IS FLAWED AND THE MONOPOLY POWER REQUIREMENT IS REDUNDANT AND WORSE

When antitrust is not enforcing per se rules in favor of or against particular types of exclusion, it actually looks for the existence of inefficiency before condemning conduct. This process is called balanc-

---

220. *See supra* note 15.
221. Another example. Suppose that a cartel member is suppressing a product improvement because the technology is easy to copy, and that would allow competitors to the cartel to enter and drive down price. Today the courts would dissolve the cartel. But ordering the cartel member to bring the product improvement to market might be an equally effective remedy.

ing.[222,223] It is seriously flawed as presently applied and the flaws tend to exempt more monopolization than efficiency requires.

A court operating under the modern standard needs two things to find antitrust liability in the absence of a per se rule. First, it needs "monopoly power."[224] Second, it needs the "anticompetitive effect" of the challenged conduct to outweigh the "procompetitive benefit."[225]

Let us focus on balancing, which is the second part, first. The problem with condemning conduct for which the harm exceeds the benefit is that much of such conduct is inefficient. For example, consider a firm that engages in exclusionary conduct that allows it to charge the price in the bottom middle graph of Diagram 2 instead of the top middle graph. We know that the benefit of this conduct exceeds the harm because the social value shape is bigger in the bottom graph than in the top one. So according to the balancing test, the conduct is not a violation of antitrust law. But the conduct is inefficient. Social value is maximized in the middle graph, not the bottom one.[226]

A properly drawn balancing test does not exonerate all conduct for

---

222. This article has also referred to it as "case-by-case" review for efficiency. It turns out that when courts profess to balance, they usually find that one side of the scale bears all the weight and the other side none. *See* Andrew I. Gavil, *Moving Beyond Caricature and Characterization: The Modern Rule of Reason in Practice*, 85 S. CAL. L. REV. 733, 761 (2012) (sources cited therein). This might mean that balancing is a sham in practice, or that courts balance but prefer to pretend that they do not. This article ignores this problem and looks instead at problems with what the courts profess to do.

223. See *supra* Part IV.B for some instances in which balancing is applied. The trend over the last fifty years has been to replace per se rules of illegality with balancing. Examples of this shift include the demise of the rule that the government always wins in merger cases and the erosion of the per se rules in concerted refusal to deal and resale price maintenance cases. *See* United States v. Von's Grocery Co., 384 U.S. 270, 301 (1966) (Stewart, J., dissenting) (stating that in merger litigation "the Government always wins"); Hovenkamp, *United States Competition Policy in Crisis*, *supra* note 74, at 355 (noting that in 1963 there was a "virtual per se rule" against mergers involving large market shares); HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 239 ("It was once commonly said that concerted refusals to deal were illegal *per se*, but this rule is subject to so many exceptions that the presumption must be turned around."); Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 881–82 (2007) (overturning per se rule for resale price maintenance agreements).

224. *See, e.g.*, United States v. Microsoft Corp., 253 F.3d 34, 51 (D.C. Cir. 2001); HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 279–80 (including market power in a synthesis of approaches to identifying anticompetitive conduct outside of the per se rule). Courts sometimes refer to "market power" instead of monopoly power. This article treats the concepts as identical. *Cf.* Krattenmaker et al., *supra* note 97, at 242, 245–48 (collecting instances in which the two terms are used and arguing that courts are confused about whether they are distinct).

225. United States v. Microsoft Corp., 253 F.3d at 58–59 ("[T]he plaintiff must demonstrate that the anticompetitive harm of the conduct outweighs the procompetitive benefit."). The doctrine usually embeds balancing in a burden-shifting regime of the following approximate character. First, the plaintiff must show "anticompetitive effect." Second, the defendant must give a "procompetitive justification." Then the court balances the two. *See, e.g., id.* at 58–59; HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 279–80.

226. For simplicity of exposition, the text assumes that courts use a situation of low or no

which the benefit exceeds the harm. It exonerates only conduct for which the benefit exceeds the harm by the largest possible amount, which is to say, only conduct that takes the economy to the social value maximizing level. In Diagram 2, a properly drawn rule condemns the move from the top to the bottom graphs because although the move increases the size of the social value shape, it does not make it as large as possible. A properly drawn rule condemns all conduct that does not take the economy to the middle graph.

In this regard, a properly drawn rule differs from textbook forms of cost-benefit analysis. Cost-benefit analysis tends to consider only whether benefits exceed costs.[227] This is an efficient approach to policymaking when policymakers have the luxury of refining any given policy in the future. If they can do that, then even though a particular policy step may not maximize social value, so long as each step taken increases social value, it may be reasonable to assume that eventually policy will arrive at the efficient point.[228] Baby steps are all you need.

This reasoning does not carry over into antitrust, however, because antitrust often can only take a single step. Once an exclusionary practice has been judicially or administratively approved, it is difficult as a legal matter[229] or a political matter for the courts to revisit it. And antitrust defendants have no incentive to undertake unilateral reductions in the level of exclusion in which they engage because that would tend to reduce their profit. As a result, a court may only ever have one shot at

---

monopoly power as the baseline in balancing the benefits and harms of a particular exclusionary practice.

The choice of baseline is important. If the courts were to take the optimal level of power as the baseline, then any exclusion that did not result in the optimal level of power would fail the balancing test and the test would be efficient. In practice, courts tend to take the level of power that existed before the exclusionary practice in question took place, or that would exist without it, as the baseline. *Cf.* Gavil et al., *supra* note 36, at 1139–40 (noting that courts have traditionally calculated antitrust damages by comparing prices before and after the antitrust violation or by comparing the market in which the violation occurred to other yardstick markets). This means that if the defendant starts at a level of power below the optimal level and increases power to a point above the optimal level, the balancing test will approve the move even though the move is inefficient. But if the defendant starts at the optimal level, then the balancing test will appropriately reject any attempt to change the level of power.

227. *See, e.g.,* Posner, Economic Analysis of Law, *supra* note 118, at 11–14 (defining efficiency as value maximization but couching discussion of efficiency in terms of value increases rather than value maximization).

228. Another assumption is that there are no local maxima; if there are, then a series of improvements may get you stuck at the top of a hill at the foot of the mountain you really wanted to climb. Because you are committed to only going up, you cannot leave the hill. *Cf.* William J. Baumol, *On Taxation and the Control of Externalities,* 62 Am. Econ. Rev. 307, 313–14 (1972) (discussing the problem of local maxima in the context of Coase and externalities).

229. *See* 18 Charles Alan Wright, Arthur R. Miller & Edward H. Cooper, Federal Practice & Procedure § 4402 (2d ed. 2002 & Supp. 2013) (discussing claim and issue preclusion, which bar relitigation of claims and issues previously litigated by the parties).

evaluating exclusion in a particular market; if exclusion results in too much power and the court fails to condemn it, market players have no incentive to give the court a second chance by engaging in new types of exclusion or rolling back inefficient but profitable levels of exclusion.

Diagram 2 demonstrates these points. If an industry makes it from the top to the bottom graph, antitrust will not intervene under the current rule because social value increases. Now assume that antitrust is acting as a policymaker and can somehow subsequently take the initiative to push the market to the middle graph, using the balancing test to decide whether this is good policy. The balancing test would approve of this second move because social value expands as you go from the bottom to the middle graph. By repeating positive net benefit moves, antitrust would have arrived at the efficient result.

But antitrust does not have this kind of policymaking authority. Firms, not antitrust administrators, decide when and how to move. Once antitrust has allowed firms to push price to the bottom graph, it may never have another opportunity to evaluate their conduct because the move to the bottom graph itself is not actionable and the firms have no incentive to do anything to depart from what is for them a profit-maximizing level of exclusion.

Let us now turn to the first part of non per se analysis. The trouble with the monopoly power requirement is that if the balancing test really does balance anticompetitive harm against procompetitive benefit, then an inquiry into monopoly power should be unnecessary. The balancing test is already a test for monopoly power. This is because monopoly power is what drives changes in social value associated with exclusionary conduct. When antitrust engages in balancing, it is measuring the effects of monopoly power.

The point of balancing is to determine whether the exercise of monopoly power reduces social value.[230] When there is no power, it concludes that there is no reduction in social value. If a firm has no monopoly power, then any attempts it may make at exclusion have no effect on price.[231] As a result, they can have no effect on fixed cost, product quality, or social value. This means that the exclusion passes the balancing test. If a firm has monopoly power due to exclusionary behavior, then that power is reflected in a change in price, fixed cost, and product quality, which in turn is reflected in a change in social value.

---

230. Assuming that antitrust seeks to maximize total welfare. *See supra* note 13. The conclusions of this Part are the same if the goal is to maximize consumer welfare instead.

231. The Supreme Court has said that "power exists whenever prices can be raised above the levels that would be charged in a competitive market." Jefferson Parish Hosp. Dist. No. 2 v. Hyde, 466 U.S. 2, 27 n.46 (1984) (citations omitted).

The balancing test registers the effect on social value and determines whether it is non-negative.

The role of monopoly power in the balancing test is evident in Diagram 2. Monopoly power is reflected in the price that the industry is able to charge relative to the monopoly price. The higher the price in the industry relative to the monopoly price, the more power the industry has over price. Let us say that an industry starts in the top middle graph and attempts but fails to exclude. This failure means that the industry has been unable to increase its monopoly power. Without a change in monopoly power, the firm will not be able to charge higher prices and generate greater profits by increasing product quality and fixed costs, so demand, price, and the social value shape will remain unchanged. If someone were to sue over the failed exclusion, antitrust would compare the post-exclusion market to the pre-exclusion market, note that social value is unchanged, and exonerate the industry. No monopoly power means no offense.

An industry that starts at the top middle graph and succeeds at excluding, at least to some extent, will take the market to somewhere other than the top middle graph. The higher price it can charge relative to the monopoly price reflects the industry's monopoly power. Regardless whether the market ends up in the middle middle or bottom middle graphs, there is a change in the size of the social value shape relative to the baseline top graph. The balancing test checks the change and finds a violation if the change is negative.

Rather than leave it to the balancing test to process the efficiency effects of monopoly power, the "monopoly power" requirement imposes an additional stand-alone monopoly power inquiry that serves as a gateway to the balancing test. The stand-alone inquiry requires either proof of high market share plus the existence of barriers to entry[232] or direct evidence that the firm can raise price without losing market share.[233]

If the purpose of the monopoly power gateway were simply to confirm the relationship between exclusion and price on which the balancing test turns, then its offense would be redundancy and wastefulness. But the monopoly power gateway fails even as a redundant test for power. Because courts favor the market share approach,[234] the test usually amounts to checking shares and barriers. Because high shares and

---

232. United States v. Microsoft Corp., 253 F.3d 34, 54–55 (D.C. Cir. 2001); *see generally* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 132–39, 295–96 (showing that market share is used to establish monopoly power in many cases in which a per se rule of illegality is not applied and discussing barriers to entry in the monopolization context).

233. *Microsoft*, 253 F.3d at 56–58 (discussing direct proof); *see generally* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 146–55.

234. *See* HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 10, at 139 ("For the present

large barriers are usually required,[235] the effect of the monopoly power gateway is to exclude all but the largest or potentially largest firms from antitrust scrutiny under the balancing test.

The trouble with this is that it is quite easy to imagine social-value-reducing exclusion by relatively small firms with no potential for becoming big. Consider, for example, an industry with ten members, each of which is producing at capacity and has a market share of 10%. Price is at marginal cost. Now assume that one of the ten cuts output in half. Industry output falls by 5%. Assume, for simplicity, no relationship between fixed cost and product quality and that the nature of demand is such that the output drop results in an increase in price of 5%. Assume further that the firm that has reduced output owns an essential potential input and refuses to deal, which explains why no firm enters, or no incumbent competitor expands, to drive output back up to the original level. Price will be permanently higher by 5%. The refuser to deal is better off because a nearly 5% profit on 50% of her original output is better than 0% profit on 100% of her original output. But social value falls. The exclusion is inefficient and the balancing test would be violated, but an antitrust suit will never reach the balancing test because the excluding firm has scarcely more than 5% of the market.

Part of the problem with the monopoly power requirement is that what really matters from the perspective of textbook economics is not the fortunes or size of any given excluding firm. What matters is the effect of its actions on the fortunes of the industry as a whole, because it is industry-wide effects that determine social value. If, as a result of exclusion, price rises, then all firms in the industry profit, and all consumers are injured, even if the market shares of the firms in the industry do not change at all.

IX. Conclusion

Aaron Edlin and Joseph Farrell remark that "[a]ntitrust protects the potential beneficial trades between competitors and consumers. . . . Consumers are not protected from all high prices, but only from those that a competitor would be happy to beat but for some thwarting action[.]"[236]

---

time, at least, market share . . . is the predominant . . . mechanism by which court's [sic] measure market power.").

235. A 15% share is probably a generous lower bound. *See id.* at 137–39 (tying and exclusive dealing "gravitating" toward 30% share minimum, share below 50% in monopolization cases rare); U.S. Dep't of Justice & Fed. Trade Comm'n, *supra* note 109, at 19 (a market with HHI below 1500 (e.g., a market with six 15% firms and one 10% firm) "ordinarily" requires no scrutiny).

236. Aaron Edlin & Joseph Farrell, *Freedom to Trade and the Competitive Process* 9 (Nat'l Bureau of Econ. Research, Working Paper No. 16818, 2011).

To these authors, this explains why "[a]ntitrust does not just fail, but explicitly doesn't try, to protect total or consumer welfare against certain obvious threats, notably the exercise of legitimately acquired monopoly power by raising price" and "antitrust's refusal to try seems proud and categorical rather than regretful and pragmatic."[237] Their argument seems to be that if there are no potential beneficial trades to be had because there are no competitors, then a monopoly ought to be welcome to charge whatever it wants.

This article has made clear the blindspot in such an argument. It is indicative of antitrust's bias in favor of property-based exclusion that the argument ignores the possibility that the competitor that would be "happy to beat" a monopoly's high prices might be trapped inside the monopoly itself. A competitor is a competitor, whether it is born through the expropriation of inputs from an incumbent monopolist or the acquisition of inputs from some source outside the market. Indeed, this article has argued that in order for a monopoly to be able to charge a high price, it must control any outside sources of inputs that might be used by competitors seeking to enter from without; so competitors can only, in a sense, come from within.

This article has undertaken five related tasks. It has: (1) identified a number of unjustified inconsistencies in antitrust enforcement, the most important of which is the exemption for property-based exclusion;[238] (2) described a simple model showing that even if monopoly profit may be thought to provide a reward for innovation, there can still be such a thing as too much monopoly profit; (3) argued that input control is a necessary condition for all monopoly power and showed how much of antitrust law and economics may be categorized in terms of different forms of input control and transitions thereto; (4) argued against remedies-based justifications for exempting property-based exclusion from antitrust and

---

237. *Id.*

238. The rhetorical strategy of this article in identifying unjustified inconsistencies in antitrust enforcement has taken the following form. First, identify a difference in enforcement (e.g., a per se rule of illegality and an exemption). Second, show that as an empirical matter there is little or no work that supports the notion that the enforcement difference maps onto an efficiency difference (e.g., show that studies have not established that cartels tend to be less efficient than monopolies). Third, show that as a theoretical matter the absence of an efficiency difference cannot be ruled out (e.g., show that we can imagine ways in which cartels might invest as heavily in R&D as monopolies or that we can imagine ways in which monopolies might invest too much in R&D).

The article has taken this approach in attacking the following enforcement differences: per se rule against hard-core cartelization / refusal to deal exemption for monopolies; more horizontal enforcement / less vertical enforcement; per se rule against hard-core cartelization / no per se rule against hard-core monopolization; case-by-case review of merger to natural monopoly / conduct requirement exemption for natural monopolies.

considered the consequences for IP of eliminating the exemption; and (5) critiqued the current balancing standard.

The tasks are tied together by an overarching argument that may be summarized as follows. Antitrust contains a lot of exemptions for which there is surprisingly little empirical or theoretical support. These inconsistencies are starkest when monopoly power is understood to be based on input control. From this perspective, the point of antitrust is to regulate input control, but the current antitrust laws appear arbitrarily to condemn some forms of control but exempt others. This uneven enforcement might be because over the last thirty years arguments for the efficiency of monopoly profit have become popular. But it is easy to see that the most important such argument, that monopoly profit creates an incentive to innovate, does not actually prove that there is no such thing as too much monopoly profit and therefore no role for antitrust. But the argument against closing the exemption for property-based exclusion is not just that monopoly profit is efficient, but also that there is no workable remedy for property-based exclusion. Redistribution of inputs is just too hard. This argument fails, however, in part because it assumes that only natural monopolies engage in property-based exclusion, which amounts to presupposing its own conclusion. One possible response to unjustified exemptions in antitrust is to replace them with liability based on a balancing test. But antitrust's current balancing test itself contains important flaws that tend to bias it against enforcement; eliminating these flaws should therefore be a part of any effort to replace the exemptions with balancing.

This overarching argument is an argument for more antitrust enforcement, but it does not pretend to be a proof that more enforcement is required. It shows that there is reason to doubt the efficiency of the current set of exemptions in antitrust, but it does not prove that they are inefficient. Further empirical or theoretical work could put them on a more secure foundation. But it also could fail to do so. The article establishes that in the absence of more knowledge, those with a hunch that more antitrust enforcement would be a good thing have no reason to fear that they are on the wrong side of current economics.

Chart 1

	Exempt	Prohibited	Case by case
States	Unnatural monopoly (refusal to deal) Natural monopoly (conduct requirement) Oligopoly	Hard-core cartel Racket	Soft-core cartel Exclusive dealing arrangement
Changes	Exclusion via superior product Natural monopolization through internal cost reduction (process innovation) Acquisitions of noncompetitors or small competitors Oligopolization Any change not attributable to the acts of market participants (natural transformation)	Exclusion via tying Exclusion via predatory pricing Racketization Hard-core cartelization	Acquisitions of large competitors Natural monopolization via cost-saving merger Soft-core cartelization Exclusion via exclusive dealing

Chart 2

States

Changes (traditional terminology in parentheses, where applicable)	Direct ownership of inputs, monopoly form	Indirect influence over owners	Direct control over property of others	Direct ownership of inputs, natural monopoly form
Acquisition (merger)	Exempt (refusal to deal) / Case by case (acquisitions of large competitors) / Exempt (otherwise)			Exempt (conduct requirement) / Case by case (acquisitions of large competitors)
Collusion (collusion)		Exempt (oligopolies) / Prohibited (hard-core cartels) / Case by case (soft-core cartels, exclusive dealing)		
Racketization			Prohibited (ex antitrust)	
Transformation (exclusion / superior competitiveness)	Exempt (superior product) / Prohibited (tying, predatory pricing)	See above		Exempt (process innovation)
Natural transformation	Exempt			See above

A caption for this chart is located on page 188.

Chart 2 on page 187 shows the inconsistent treatment of transitions to input control or natural monopoly and continuations of input control or natural monopoly. Following any column down gives antitrust treatment of a mode of input control or natural monopoly in its continuing form. Following any row across gives antitrust treatment of the transition to a mode of input control or natural monopoly. Split boxes reflect different treatment of transitions and continuing states. For example, if you follow the "Direct ownership of inputs, monopoly form" column down to the first row, the split box shows that transitions to direct ownership via merger are sometimes treated on a case-by-case basis but continuing direct ownership is always exempt under the refusal to deal doctrine. If you continue along that column down to "Collusion," you find the box empty because collusion is not a mode of transitioning to direct owner-ship; collusion is a mode of transitioning to indirect control. If you continue further down to "Transformation," you encounter another split box. This reflects the different treatment of transitions to direct ownership through market transformations, such as tying or predatory pricing, which are prohibited, and continuing direct ownership, which, as we have seen, is always exempt. A parenthetical next to the name for a change category gives the traditional name for the category if one exists.